Les Cahiers du GERAD ISSN: 0711–2440

Shortfall risk models when information of loss function is incomplete

E. Delage,
S. Guo, H. Xu

G–2017–104

November 2017
Second revision: June 2020

La collection *Les Cahiers du GERAD* est constituée des travaux de recherche menés par nos membres. La plupart de ces documents de travail a été soumis à des revues avec comité de révision. Lorsqu'un document est accepté et publié, le pdf original est retiré si c'est nécessaire et un lien vers l'article publié est ajouté.

The series *Les Cahiers du GERAD* consists of working papers carried out by our members. Most of these pre-prints have been submitted to peer-reviewed journals. When accepted and published, if necessary, the original pdf is removed and a link to the published article is added.

CITATION ORIGINALE / ORIGINAL CITATION

Delage, E., Guo, S., Xu, H., Shortfall risk models when information on loss function is incomplete, Operations Research, Articles in Advance, pp.1–8, 2022, http://dx.doi.org/10.1287/opre.2021.2212.

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2020
– Bibliothèque et Archives Canada, 2020

Legal deposit – Bibliothèque et Archives nationales du Québec, 2020
– Library and Archives Canada, 2020

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Shortfall risk models when information of loss function is incomplete

Erick Delage [a]

Shaoyan Guo [b]

Huifu Xu [c]

[a] *GERAD & Department of Decision Sciences, HEC Montréal, Montréal (Québec), Canada*

[b] *School of Mathematical Sciences, Dalian University of technology, Dalian, China*

[c] *Department of Systems Engineering & Engineering Management, The Chinese University of Hong Kong, Hong Kong*

erick.delage@hec.ca
syguo@dlut.edu.cn
hfxu@se.cuhk.edu.hk

November 2017
Second revision: June 2020
Les Cahiers du GERAD
G–2017–104

Abstract: Utility-based shortfall risk measure (SR) effectively captures decision maker's risk attitude on tail losses by an increasing convex loss function. In this paper, we consider a situation where the decision maker's risk attitude towards tail losses is ambiguous and introduce a robust version of SR which mitigates the risk arising from such ambiguity. Specifically, we use some available partial information or subjective judgement to construct a set of utility-based loss functions and define a so-called preference robust SR (PRSR) through the worst loss function from the (ambiguity) set. To implement PRSR in practice, we propose three approaches for constructing the ambiguity set. We then apply the PRSR to optimal decision making problems and demonstrate how the latter can be reformulated as tractable convex programs when the underlying exogenous uncertainty is discretely distributed. In the case when the probability distribution is continuous, we propose a sample average approximation scheme and show that it converges to the true problem in terms of the optimal value and the optimal solutions as the sample size increases.

Keywords: Preference robust optimization, utility-based shortfall risk measure, preference elicitation, linear programming, tractability, sample average approximation

---

Acknowledgments: The authors thank the associate editor and two referees for their valuable comments which helped us improve the quality of this paper. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada [Grant RGPIN–2016–05208], the Canada Research Chair program [Grant 950–230057], and National Natural Science Foundation of China [Grants 71432002, 91746210, 11801057]. The authors also thank Alexander Schied for some useful discussions regarding the proof of Lemma A.5 presented in the appendices.

1 Introduction

In their highly influential work, Föllmer and Schied (2002) introduced the following utility-based shortfall risk (SR) measure

$$\mathrm{SR}^P_{l,\lambda}(Z) := \inf_{t\in\mathbb{R}}\{t : \mathbb{E}_P[l(-Z-t)] \le \lambda\}, \tag{1}$$

where $Z : \Omega \to \mathbb{R}$ is a random variable in $L_p(\Omega, \mathscr{B}, P)$, defined on probability space $(\Omega, \mathscr{B}, P)$, and representing a financial position, $l : \mathbb{R} \to \mathbb{R}$ is an increasing convex loss function which is not identically constant, and where $\lambda \in \mathbb{R}$. Föllmer and Schied showed that $\mathrm{SR}^P_l(Z)$ is a convex risk measure (satisfying monitonicity, convexity and translation invariance). It is also known as the family of risk measure that satisfies the principle of zero utility (see Buhlmann (1970)). In the case when $l(s) = s$ and $\lambda = 0$, $\mathrm{SR}^P_{l,\lambda}(Z)$ coincides with the negative expected value. Alternatively, when $l(s) = \max(\tau s, (1-\tau)s)$ with $\lambda = 0$, it is known as an expectile (see Bellini and Bignozzi (2015)), and as the entropic risk measure when $l(s) = e^{\beta s}$ with $\beta > 0$. Finally, when $l(\cdot)$ takes the form of the non-convex function $\mathbb{1}_{(0,+\infty]}(\cdot)$, i.e., $\mathbb{1}_{(0,+\infty]}(s) = 1$ if $s \in (0, +\infty]$, and 0 otherwise, $\mathrm{SR}^P_{l,\lambda}(Z)$ coincides with the popular non-convex value-at-risk measure $\mathrm{VaR}_\lambda(Z)$.

Compared to many other monetary risk measures, a notable property of SR is that it is invariant under randomization and is an elicitable risk measure, which is amenable to perform backtesting, i.e., the activity of periodically comparing the forecast risk measure with the realized value of the variable of interest in order to try to increase the accuracy of the forecasting methodology (Bellini and Bignozzi (2015)). Some studies have also argued that it is more sensitive to financial losses from extreme events (see Giesecke et al. (2008)) than value-at-risk and conditional value-at-risk, and has favorable consistency properties for dynamic measurement of risks over time (see Weber (2006)).

From the discussions above, we can see clearly that the loss function $l(\cdot)$ plays an important role in determining the distinct nature of SR. In practice, there could be considerable ambiguity in the choice of l. For example, the decision maker could hesitate about the type of loss function that best characterizes his preference, or a group of decision makers might fail to reach a consensus or the prospect space is too large and/or too complex to elicit. In other words, the loss function $l(\cdot)$ is not uniquely/explicitly given. To overcome the risk arising from model mis-specification, in these circumstances, it might be sensible to use partially available information and/or subjective judgement to construct a set of loss functions, denoted by L, and consider the worst SR based on the set to mitigate the risk arising from ambiguity/misjudgement of the true loss function:

$$\text{(PRSR)} \quad \mathrm{SR}^P_L(Z) := \inf\left\{t : \sup_{l\in L}\mathbb{E}_P[l(-Z-t)] - l(0) \le 0\right\}. \tag{2}$$

Here by setting $\lambda = l(0)$, we focus on the class of normalized (i.e. $\rho(Z) = 0$ when $Z = 0$) convex utility-based shortfall risk measures. We call $\mathrm{SR}^P_L(Z)$ Preference Robust SR (PRSR) in that it returns an estimate of the riskiness of Z which is guaranteed to bound from above the risk measured using any $\mathrm{SR}^P_l(Z)$ with $l \in L$.

A key element in the above model is the construction of the set L and our work will be strongly inspired by principles that have been proposed in the preference robust optimization (PRO) literature. In particular, Armbruster and Delage (2015) applied PRO in the context of expected utility maximization problems. They considered three classes of utility functions (namely concave, S-shaped and functions that capture the notion of prudence) and allowed to account for ad-hoc pairwise comparisons of lotteries. Moreover, they proposed tractable approaches to dealing with the resulting optimization problems when the underlying random variable follows a finite discrete distribution. Following up on this methodology, Haskell et al. (2016) extended this preference robust expected utility framework to cases where there is also ambiguity about the choice of a probability function. More recently, PRO was also applied to the case that a convex risk measure is ambiguous, that is,

$$\varrho_{\mathcal{R}}(Z) := \sup_{\rho\in\mathcal{R}} \rho(Z), \tag{3}$$

where $\mathcal{R}$ is an ambiguity set of convex risk measures. While the concept of worst-case risk appears for the first time in Föllmer and Schied (2002), Delage and Li (2018) recommends using $\mathcal{R}$ as a set of all convex/coherent/law invariant risk measures that satisfies the partial ordering expressed in an elicited list of pairwise comparisons. Once again, tractable reformulations were proposed for problems described on a finite outcome space. Unfortunately, the reformulations provided in Delage and Li (2018) do not apply when the convex risk measure is also known to be a utility-based shortfall risk measure. Another important limitation of all methods above consist in the fact that tractable reformulations are only established under the hypothesis of discrete random variables. It remains to be explored as to whether there is a tractable numerical scheme for preference robust risk minimization problems where the underlying random variables are continuously distributed.

In this paper, we consider the following preference robust risk minimization problem:

$$\text{(PRSR-Opt)} \quad \min_{x\in X} \varrho_{\mathcal{R}}(-c(x,\xi))$$

where $c(x,\xi): \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a continuous function representing costs, $\xi : \Omega \to \mathbb{R}^m$ and $\mathcal{R}$ is a set of utility-based shortfall risk measures defined using information about the certainty equivalent of a list of random variables, about whether the risk measure is coherent, and possibly also about a bound on the sensitivity of the risk measure to losses of large size. Given such information about ρ, we will derive an ambiguity set L for which

$$\text{(PRSR-Opt)} \quad \min_{x\in X} \varrho_{\mathcal{R}}(-c(x,\xi)) \quad \equiv \quad \min_{x\in X} \mathrm{SR}_L^P(-c(x,\xi)) \tag{4}$$

This will allow us to demonstrate for the first time that when $c(x,\xi)$ is a piecewise linear function and P is discrete, (PRSR-Opt) can be reformulated either as a linear programming problem of finite dimension or as a semi-infinite LP for which we design an efficient column-generation method, depending on the type of sensitivity bound. Alternatively, when P is a continuous probability measure, we propose a sample average approximation scheme

$$\text{(PRSR-Opt-N)} \quad \min_{x\in X} \mathrm{SR}_L^{P_N}(-c(x,\xi)) \tag{5}$$

and establish conditions under which (PRSR-Opt-N) converges to (PRSR-Opt) in terms of the optimal value and optimal solutions as the sample size N increases. Note that the new bound on the sensitivity to tail losses that is imposed on $\mathcal{R}$ will play a crucial role in satisfying these conditions, i.e., equicontinuity of the ambiguity set of loss functions over a compact set.

The rest of the paper is organized as follows. Section 2 introduces different sets of preference robust risk measures that will be studied in this paper and derives their representation in terms of SR_L^P for some ambiguity set of loss functions L. Section 3 presents tractable reformulations for (PRSR-Opt) with the characterizations of L derived in Section 2 when the underlying random variable is finitely distributed. Section 4 proposes a discretization scheme for (PRSR-Opt) when the underlying random variables are continuously distributed. This is followed with Section 5, which reports numerical experiments on the (PRSR-Opt) model. We finally conclude in Section 6. All proofs are relegated to the appendices of the paper.

Remark 1 *The shortfall risk measure is closely linked to the well-known optimized certainty equivalent risk measure due to Ben-Tal and Teboulle (2007). In Ben Tal & Teboulle (2007), three types of certainty equivalents are introduced:* Certainty Equivalent $C_u(Z) = u^{-1}(\mathbb{E}[u(Z)])$, *$u$-mean $M_u(Z)$ which satisfies $\mathbb{E}[u(Z - M_u(Z))] = 0$ and* Optimized Certainty Equivalent $S_u(Z) = \sup_{\eta\in\mathbb{R}}\{\eta + \mathbb{E}[u(Z-\eta)]\}$. *It shows that*

$$\mathrm{SR}_l(Z) = -M_u(Z) = -\inf_{\lambda>0} S_{\lambda u}(Z).$$

These relationships show that shortfall risk measures are closely related to certainty equivalent risk measure and indeed they coincide under some circumstances. We note that additional relationships between shortfall risk measures, regret measures, statistics, and utility theory could be further identified based on the risk quadrangle defined in Rockafellar and Uryasev (2013).

It should also be mentioned that Föllmer and Schied (2002) introduced a robust version of utility-based shortfall risk measure, which is defined as

$$\mathrm{SR}_l^{\mathcal{Q}}(Z) := \inf \left\{ t : \sup_{P\in\mathcal{Q}} \mathbb{E}_P[l(-Z-t)] - l(0) \le 0 \right\} = \sup_{P\in\mathcal{Q}} \mathrm{SR}_l^P(Z).$$

The above version focuses on the ambiguity of probability measure, which means the robustness is in the sense of exogenous uncertainty which is not associated with decision maker's risk attitude. We refer readers to Guo and Xu (2019) for a specific discussion on distributionally robust formulation of SR and its application in risk management.

2 Preference robust normalized utility-based shortfall risk measures

In this section, we exploit elicited risk preference information[1], inspired by the preference robust risk minimization paradigm introduced in Delage and Li (2018), to characterize what set of loss functions L should be employed in $\mathrm{SR}_L^P(Z)$ in order to produce a robust measurement of risk based on partial information about the decision maker's risk preferences. In particular, we will investigate how to account for information about whether the risk measure is positive homogeneous, about some "confidence" intervals for the risk of a list of random variables, and finally about how sensitive the decision maker is regarding events that occur in the tail of Z. To be more specific we introduce the following classes of risk measures.

Definition 1 *(Normalized utility-based shortfall risk measures) Let $\mathcal{R}_{ubsr}$ be the set of all normalized utility-based shortfall risk measures, i.e., all utility-based shortfall risk measures (see Föllmer and Schied (2002)) such that $\rho(0) = 0$.*

Definition 2 *(Coherent risk measures) Let $\mathcal{R}_{coh}$ be the set of coherent risk measures (Artzner et al. 1999).*

As mentioned in Delage and Li (2018), verifying whether the risk measure that captures the decision maker's risk attitude is a member of $\mathcal{R}_{coh}$ or not reduces to establishing whether the ordinal comparison of riskiness of two random variables should be affected by a uniform positive scaling of both random variables. Yet, this information generally is not very restrictive when identifying the right $\mathcal{R} \in \mathcal{R}_{ubsr}$.

In order to refine the characterization of $\mathcal{R}_{ubsr}$ or $\mathcal{R}_{ubsr} \cap \mathcal{R}_{coh}$, and implicitly L as we will soon demonstrate, one should try to exploit information about the absolute riskiness level of a set of random variables. This gives rise to the following subclass of risk measures.

Definition 3 *Let $\{W_k\}_{k=1}^K$ be a sequence of random variables with an associated set of "confidence" intervals $[w_k^-, w_k^+] \subseteq [\operatorname{essinf} W_k, \operatorname{esssup} W_k]$ for the "certainty equivalent" of each W_k. The set $\mathcal{R}_{ce}(\mathcal{W})$ denotes the set of all risk measures which evaluate the risk of each W_k to be larger than the risk of w_k^+ and lower than the risk of w_k^-, i.e.,*

$$\mathcal{R}_{ce}(\mathcal{W}) := \{\rho : \mathscr{L}_p \to \mathbb{R} \,|\, \rho(w_k^+) \le \rho(W_k) \le \rho(w_k^-)\,,\, \forall\, k \in \{1, 2, \ldots, K\}\,\}$$

with $\mathcal{W} := \{(W_k, w_k^-, w_k^+)\}_{k=1}^K$.

Note that a natural method that can be used to identify two bounds for the riskiness of a random variable W_k would take the form of questions such as:

- Lower bound w_k^-: "What is the largest amount of money that you would decline in order to be exposed to the risk of W_k?"
- Upper bound w_k^+: "What is the lowest amount of money that you would be willing to receive instead of being exposed to the risk of W_k?"

When the answers to both questions are such that $w_k^- = w_k^+ = \bar{w}_k$, this implies that we have identified the certainty equivalent of W_k, i.e. $\rho(W_k) = \bar{w}_k$, yet in practice it is more often the case that only an interval $[w_k^-, w_k^+]$ will be obtained for this value. Moreover, the assumption that $[w_k^-, w_k^+] \subseteq [\text{essinf } W_k, \text{esssup } W_k]$ is not restrictive since otherwise one can simply replace $w_k^- := \max(w_k^-, \text{essinf } W_k)$ and $w_k^+ := \min(w_k^+, \text{esssup } W_k)$ following the monotonicity property of normalized convex utility-based shortfall risk measures (see Section A1.1 of the appendices for more details).

We will finally find it useful in our later analysis to have in hand a bound on how sensitive the utility-based shortfall risk measure is to losses of large size. To do so, we consider the following class of risk measures.

Definition 4 *Let $\varepsilon : \mathbb{R}_+ \to (0, 1]$ be a non-increasing function and $\{Z_M^\varepsilon\}_{M\geq 1}$ be a set of discrete random variables supported at $-M$ and 0 with respective probabilities $\varepsilon(M)$ and $1-\varepsilon(M)$, i.e.,*

$$Z_M^\varepsilon = \begin{cases} -M & \text{w.p. } \varepsilon(M), \\ 0 & \text{w.p. } 1-\varepsilon(M). \end{cases} \tag{6}$$

Denote by $\mathcal{R}_{bnd}(\varepsilon)$ the set of risk measures that assigns to each random variable in the set $\{Z_M^\varepsilon\}_{M\geq 1}$ a risk that is lower than the risk of a certain loss of one, namely

$$\mathcal{R}_{bnd}(\varepsilon) := \{\rho : \mathscr{L}_p \to \mathbb{R} \,|\, \rho(Z_M^\varepsilon) \leq \rho(-1), \forall\, M \geq 1\}. \tag{7}$$

Theoretically speaking, if a decision maker agrees to be using a normalized convex utility-based shortfall risk measure, then he should agree that such a function $\varepsilon(\cdot)$ necessarily exist. This is due to the fact for any increasing convex loss function l, the risk measure $\text{SR}_l^P \in \mathcal{R}_{bnd}(\varepsilon^*)$ with $\varepsilon^*(M) := (l(0)-l(-1))/(l(M-1)-l(-1))$ based on the argument that:

$$\begin{aligned}&\varepsilon^*(M)l(M-1)+(1-\varepsilon^*(M))l(-1) = l(0)\\ &\qquad\Rightarrow\ \text{SR}_l^P(Z_M^{\varepsilon^*}) = \inf\{t : \varepsilon^*(M)l(M-t) + (1-\varepsilon^*(M))l(-t) \leq l(0)\} \leq 1 = \text{SR}_l^P(-1).\end{aligned}$$

Practically speaking, identifying $\varepsilon^*(M)$ can be as difficult as identifying $l(\cdot)$ itself. However, one could for instance establish that for a random loss Z_λ which follows an exponential distribution with mean λ, the risk of $\rho(-Z_\lambda)$ is considered lower than a guaranteed loss of one for some $\bar{\lambda}$. This information can be exploited to conclude that $\rho \in \mathcal{R}_{bnd}(\bar{\varepsilon})$ with $\bar{\varepsilon}(M) := \exp(-\bar{\lambda}M)$ since the fact that $Z_M^{\bar{\varepsilon}}$ stochastically dominates $-Z_{\bar{\lambda}}$ in the first order, for all $M \geq 1$, implies that $\rho(Z_M^{\bar{\varepsilon}}) \leq \rho(-Z_{\bar{\lambda}}) \leq \rho(-1)$.

We hence are equipped to present the key results of this section which consists in demonstrating how to transform information about risk preferences, which takes the form of some $\mathcal{R} \in \{\mathcal{R}_{ubsr} \cap \mathcal{R}_{ce}(\mathcal{W}), \mathcal{R}_{ubsr} \cap \mathcal{R}_{coh} \cap \mathcal{R}_{ce}(\mathcal{W}), \mathcal{R}_{ubsr} \cap \mathcal{R}_{ce}(\mathcal{W}) \cap \mathcal{R}_{bnd}(\varepsilon)\}$, into a set of ambiguous loss functions L that make $\varrho_{\mathcal{R}}(Z) = \text{SR}_L^P(Z)$ for any random variable Z.

Proposition 1 (Characterization of L for $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})}(Z)$) *Let $\mathcal{R}_{ubsr}$ and $\mathcal{R}_{ce}(\mathcal{W})$ be defined as in Definitions 1 and 3. Then the preference robust risk measure $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})}(Z) = \text{SR}^P_{L_{ubsr}\cap L_{ce}(\mathcal{W})}(Z)$ for all $Z \in \mathscr{L}_p$, where L_{ubsr} is the set of all convex non-decreasing functions $l : \mathbb{R} \to \mathbb{R}$ that are strictly increasing over $[z_0, \infty)$ for some $z_0 < 0$, and where*

$$L_{ce}(\mathcal{W}) := \left\{ l \,\middle|\, \begin{array}{l} \mathbb{E}_P[l(-W_k + w_k^-)] \leq l(0) \\ \mathbb{E}_P[l(-W_k + w_k^+)] \geq l(0) \end{array},\ \forall\, k \in \{1, 2, \ldots, K\} \right\}. \tag{8}$$

Representing $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})}(Z)$ as $\text{SR}^P_{L_{ubsr}\cap L_{ce}(\mathcal{W})}(Z)$ is useful for solving (PRSR-Opt) since it reduces the evaluation of the risk measure to finding the optimal value of a stochastic programming problem with semi-infinite stochastic constraints:

$$\min_{x\in X}\ \varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})}(-c(x,\xi)) \quad \equiv \quad \begin{array}{ll} \displaystyle\min_{x\in X, t} & t \\ \text{s.t.} & \mathbb{E}_P[l(c(x,\xi)-t)] - l(0) \leq 0,\ \forall\, l \in L_{ubsr} \cap L_{ce}(\mathcal{W}). \end{array} \tag{9}$$

In Section 3, we will address the computational challenges arising from this reformulation by using an analysis that is similar in spirit to the one used in Armbruster and Delage (2015) for the case where ξ has a finite discrete distribution. Later, in Section 4, we will derive the theory that can be used to justify a discrete approximation of this optimization model when ξ is continuously distributed. It is worth noting that the preference robust optimization model employed in Armbruster and Delage (2015) can handle comparison of arbitrary pairs of random variables which appears to be more difficult to integrate in this preference robust risk measure framework.

We now turn to imbedding the property of positive homogeneity in the characterization.

Proposition 2 (Characterization of L for $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{coh}\cap\mathcal{R}_{ce}(\mathcal{W})}(Z)$) *Given that $\mathcal{R}_{ubsr}\cap\mathcal{R}_{coh}\cap\mathcal{R}_{ce}(\mathcal{W}) \neq \emptyset$, the preference robust risk measure $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{coh}\cap\mathcal{R}_{ce}(\mathcal{W})}(Z)$ is equivalent to $\mathrm{SR}^P_{l_\tau}(Z)$ for all $Z \in \mathscr{L}_p$, where*

$$l_\tau(s) = \max(\tau s,\, (1-\tau)s) \;\text{ for }\; \tau = \min_k \frac{\mathbb{E}[(W_k - w_k^-)^+]}{\mathbb{E}[|W_k - w_k^-|]}, \tag{10}$$

with $(s)^+ := \max(0, s)$. Moreover, $\mathcal{R}_{ubsr}\cap\mathcal{R}_{coh}\cap\mathcal{R}_{ce}(\mathcal{W})$ is non-empty if and only if

$$\min_k \frac{\mathbb{E}[(W_k - w_k^-)^+]}{\mathbb{E}[|W_k - w_k^-|]} \geq \max\left(\frac{1}{2},\, \max_k \frac{\mathbb{E}[(W_k - w_k^+)^+]}{\mathbb{E}[|W_k - w_k^+|]}\right).$$

This result leads straightforwardly to a convenient finite dimensional reformulation for the (PRSR-Opt) problem and the problem of minimizing the worst-case risk achieved by any $x \in X$. In particular,

$$\min_{x\in X} \;\; \varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{coh}\cap\mathcal{R}_{ce}(\mathcal{W})}(-c(x,\xi)) \qquad \equiv \qquad \begin{array}{ll} \displaystyle\min_{x\in X,t} & t \\ \text{s.t.} & \mathbb{E}_P[l_\tau(c(x,\xi)-t)] \leq 0\,, \end{array} \tag{11}$$

with τ defined according to Proposition 2. In comparison with the case where the risk measure is not known to be positive homogeneous, formulation (11) is more attractive from computational perspective because the latter is a convex minimization problem with a single stochastic inequality constraint. Indeed, it is an ordinary nonlinear programming problem for which existing NLP codes can be readily applied to solve. We conclude this section with a characterization of L when one does not impose positive homogeneity but instead constrains on the sensitivity of the utility-based shortfall risk measure to losses of large size.

Proposition 3 (Characterization of L for $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})\cap\mathcal{R}_{bnd}(\varepsilon)}$) *Let $\mathcal{R}_{ubsr}$, $\mathcal{R}_{ce}(\mathcal{W})$, and $\mathcal{R}_{bnd}(\varepsilon)$ be defined as in Definitions 1, 3, and 4. Then the preference robust risk measure $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})\cap\mathcal{R}_{bnd}(\varepsilon)}(Z)$ is equivalent to $\mathrm{SR}^P_{L_{ubsr}\cap L_{ce}(\mathcal{W})\cap L_{bnd}}(Z)$ for all $Z \in \mathscr{L}_p$, where*

$$L_{bnd} := \{l \,|\, \varepsilon(z)l(z-1) + (1-\varepsilon(z))l(-1) \leq l(0)\,,\, \forall\, z \geq 1\}\,. \tag{12}$$

Moreover, each loss function $l \in L_{ubsr}\cap L_{ce}(\mathcal{W})\cap L_{bnd}$ satisfies

$$l(z_1) - l(0) \leq \frac{1-\varepsilon(z_1+1)}{\varepsilon(z_1+1)}(l(0)-l(-1))\,,\, \forall\, z_1 \geq 0 \tag{13}$$

and

$$l'_+(z_1) \leq \phi(z_1)(l(0)-l(-1)) \,\text{for}\, 0 \leq z_1 < z_2, \tag{14}$$

where $l'_+(z_1)$ denotes the right derivative of $l(\cdot)$ at z_1, and

$$\phi(z_1) = \inf\left\{\left(\frac{1-\varepsilon(z+1)}{\varepsilon(z+1)} - z_1\right)\frac{1}{z-z_1} \;:\; z > z_1\right\}. \tag{15}$$

Similarly, to the case of Proposition 1, this result leads to a semi-infinite reformulation for the (PRSR-Opt) problem. Namely,

$$\min_{x\in X} \varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})\cap\mathcal{R}_{bnd}(\varepsilon)}(-c(x,\xi)) \equiv$$

$$\begin{array}{ll} \min_{x\in X,t} & t \\ \text{s.t.} & \mathbb{E}_P[l(c(x,\xi)-t)]-l(0)\le 0\,,\ \forall\, l\in L_{ubsr}\cap L_{ce}(\mathcal{W})\cap L_{bnd}\,. \end{array} \tag{16}$$

Remark 2 *We may draw a couple of useful conclusions from the proposition.*

(i) *$\phi(z)$ is non-decreasing over $[-M,M]$. To see this, we note that the convexity of $l(\cdot)$ ensures $l(z)-l(0)\ge z(l(0)-l(-1)), \forall z>0$ and through (13), we have $\frac{1-\varepsilon(z+1)}{\varepsilon(z+1)}-z\ge 0, \forall z>z_1$. This implies the objective function of the minimization problem in (15) is strictly increasing in z_1 for fixed $z>z_1$. Together with the fact that the feasible set is getting smaller as z_1 increases, this implies $\phi(z_1)$ is non-decreasing for $z_1\in[-M,M]$.*

(ii) *The monotonicity of $\phi(\cdot)$ implies that $\phi(z)\le\phi(M), \forall z\in[-M,M]$. On the other hand, since $l(\cdot)$ is convex and non-decreasing, then the inequality above and inequality (14) imply that $l'_-(z)\le l'_+(z)\le\phi(M)(l(0)-l(-1)), \forall z\in[-M,M]$, where $l'_-(z)$ denotes the left derivative of $l(\cdot)$ at z. The latter ensures that every loss function in $L_{ubsr}\cap L_{ce}(\mathcal{W})\cap L_{bnd}$ is equicontinuous over any interval $[-M,\,M]$, i.e. $\frac{l(z_2)-l(z_1)}{z_2-z_1}\le\phi(M)(l(0)-l(-1))\,,\ \forall\, z_1,z_2\in[-M,M]\,, \forall\, l\in L_{ubsr}\cap L_{ce}(\mathcal{W})\cap L_{bnd}$, which means the class of loss functions in $L_{ubsr}\cap L_{ce}(\mathcal{W})\cap L_{bnd}$ are equicontinuous over $[-M,\,M]$ for any $M>0$.*

3 Tractable formulation of (PRSR-Opt)

The representations that we have developed for $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})}(Z)$ and $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})\cap\mathcal{R}_{bnd}(\varepsilon)}(Z)$ in (9) and (16) are not numerically tractable as they are semi-infinite programming problems. In this section, we demonstrate that when P follows a finite discrete distribution, both of them can be reformulated as finite dimensional convex programming problems. We do so by replacing Z with $-c(x,\xi)$ so that the results can be directly plugged into (PRSR-Opt).

To this end, let us write Ξ as $\Xi:=\{\xi_1,\ldots,\xi_N\}$ with associated probabilities $p_i:=P(\xi=\xi_i)$.

Proposition 4 *The (PRSR-Opt) problem with $L_{ubsr}\cap L_{ce}(\mathcal{W})$ can be reformulated as the following problem:*

$$\min_{x\in X,t,u,\gamma,\eta^{(1)},\eta^{(2)}} \quad t \tag{17a}$$

$$\text{s.t.} \quad p_i(c(x,\xi_i)-t)-\sum_{j=1}^{M}u_{ij}y_j\le 0,\ \forall\, i=1,\ldots,N, \tag{17b}$$

$$p_i-\sum_{j=1}^{M}u_{ij}=0,\ \forall\, i=1,\ldots,N, \tag{17c}$$

$$\sum_{i=1}^{N}u_{ij}+\sum_{m=1}^{M}\gamma_{jm}-\sum_{m=1}^{M}\gamma_{mj}-\sum_{k=1}^{K}\eta_k^{(1)}P(-W_k+w_k^-=y_j) + \sum_{k=1}^{K}\eta_k^{(2)}P(-W_k+w_k^+=y_j)=0,\ j=\{1,\ldots,M\}\setminus\{j_0,j_-\}, \tag{17d}$$

$$\sum_{i=1}^{N}u_{ij_-}+\sum_{m=1}^{M}\gamma_{j_-m}-\sum_{m=1}^{M}\gamma_{mj_-}-\sum_{k=1}^{K}\eta_k^{(1)}P(-W_k+w_k^-=y_{j_-}) + \sum_{k=1}^{K}\eta_k^{(2)}P(-W_k+w_k^+=y_{j_-})\ge 0, \tag{17e}$$

$$\sum_{m=1}^{M}\gamma_{mj}(y_m-y_j)\geq 0\,,\,\forall\, j=1,\ldots,M, \tag{17f}$$

$$u_{ij}\geq 0,\gamma_{mj}\geq 0,\eta^{(1)}\geq 0,\eta^{(2)}\geq 0,\; i=1,\ldots,N,m,j=1,\ldots,M\,, \tag{17g}$$

where $\{y_j\}_{j=1}^{M}$ *is an indexed list of the elements of* $\mathcal{Y}:=\bigcup_{k=1}^{K}\mathrm{supp}(-W_k+w_k^-)\cup\mathrm{supp}(-W_k+w_k^+)\cup\{0\}\cup\{-1\}$*, while* j_0 *and* j_- *are the indexes such that* $y_{j_0}=0$ *and* $y_{j_-}=-1$*. In particular, if* $c(x,\xi)$ *is a convex piecewise linear function of* x*, then problem (17) can be reformulated as a finite dimensional linear programming problem.*

We now move on to discuss a finite dimensional reformulation of (PRSR-Opt) with $L_{ubsr}\cap L_{ce}(\mathcal{W})\cap L_{bnd}$ in the case where $\varepsilon(y)^{-1}:=1/\varepsilon(y)$ is a piecewise-linear function on $\mathbb{R}$ and X is a bounded set.

Proposition 5 *Let* X *be bounded and* $\varepsilon(\cdot)^{-1}$ *be a piecewise-linear function with* $\mathcal{Y}_\varepsilon$ *as an unbounded set containing all of its non-differentiable points. Then the (PRSR-Opt) problem with* $L_{ubsr}\cap L_{ce}(\mathcal{W})\cap L_{bnd}$ *is equivalent to*

$$\min_{x\in X,t,u,\sigma,\gamma,\rho,\theta,\eta^{(1)},\eta^{(2)}}\quad t \tag{18a}$$

$$\text{s.t.}\quad t_-\leq t\leq t_+ \tag{18b}$$

$$-\left(\sum_{i=1}^{N}u_{ij_-}+\sum_{m=1}^{M}\gamma_{j_-m}-\sum_{m=1}^{M}\gamma_{mj_-}-\sum_{k=1}^{K}\eta_k^{(1)}P(-W_k+w_k^-=y_{j_-})\right.$$
$$\left.+\sum_{k=1}^{K}\eta_k^{(2)}P(-W_k+w_k^+=y_{j_-})\right)+\sum_{j\in\mathcal{I}}\theta_{j-M}(z_j-1)+\sum_{j=j_0+1}^{M}\rho_{j-j_0}(z_j-1)\leq 0, \tag{18c}$$

$$p_i(c(x,\xi_i)-t)-\sum_{j=1}^{M}u_{ij}y_j-\sigma_i\leq 0,\,\forall\, i=1,\ldots,N, \tag{18d}$$

$$p_i-\sum_{j=1}^{M}u_{ij}=0,\,\forall\, i=1,\ldots,N, \tag{18e}$$

$$\sum_{i=1}^{N}u_{ij}+\sum_{m=1}^{M}\gamma_{jm}-\sum_{m=1}^{M}\gamma_{mj}-\sum_{k=1}^{K}\eta_k^{(1)}P(-W_k+w_k^-=y_j) \tag{18f}$$
$$+\sum_{k=1}^{K}\eta_k^{(2)}P(-W_k+w_k^+=y_j)=0,\; j=\{1,\ldots,j_0-1\}\setminus\{j_-\},$$

$$\sum_{i=1}^{N}u_{ij}+\sum_{m=1}^{M}\gamma_{jm}-\sum_{m=1}^{M}\gamma_{mj}-\sum_{k=1}^{K}\eta_k^{(1)}P(-W_k+w_k^-=y_j) \tag{18g}$$
$$+\sum_{k=1}^{K}\eta_k^{(2)}P(-W_k+w_k^+=y_j)=-\rho_{j-j_0},\forall\, j=j_0+1,\ldots,M-1,$$

$$\sum_{i=1}^{N}u_{iM}+\sum_{m=1}^{M}\gamma_{Mm}-\sum_{m=1}^{M}\gamma_{mM}-\sum_{k=1}^{K}\eta_k^{(1)}P(-W_k+w_k^-=y_M) \tag{18h}$$
$$+\sum_{k=1}^{K}\eta_k^{(2)}P(-W_k+w_k^+=y_M)=-\rho_{M-j_0}-\sum_{j\in\mathcal{I}}\theta_{j-M},$$

$$\sum_{m=1}^{M}\gamma_{mj}(y_m-y_j)\geq 0\,,\,\forall\, j=1,\ldots,M-1, \tag{18i}$$

$$\sum_{m=1}^{M}\gamma_{mM}(y_m-y_M)\geq\sum_{i=1}^{N}\sigma_i-\sum_{j\in\mathcal{I}}\theta_{j-M}(y_j-y_M), \tag{18j}$$

$$u_{ij}\geq 0,\gamma_{mj}\geq 0,\forall\, i=1,\ldots,N,m,j=1,\ldots,M, \tag{18k}$$

$$\sigma\geq 0,\rho\geq 0,\theta\geq 0,\eta^{(1)}\geq 0,\eta^{(2)}\geq 0, \tag{18l}$$

where $u \in \mathbb{R}^{N\times M}$, $\sigma \in \mathbb{R}^N$, $\gamma \in \mathbb{R}^{M\times M}$, $\rho \in \mathbb{R}^{M-j_0}$, $\theta \in \mathbb{R}^{|\mathcal{I}|}$, $\eta^{(1)} \in \mathbb{R}^K$ *and* $\eta^{(2)} \in \mathbb{R}^K$. *One also has* $t_- := \min_{x\in X} \min_i c(x,\xi_i)$ *and* $t_+ := \max_{x\in X} \max_i c(x,\xi_i)$, *while* $\mathcal{Y}' := \{y_j\}_{j=1}^M$ *is an ordered list of the elements of* $\mathcal{Y} \cup \{y_*, y^*\} \cup (\mathcal{Y}'_\varepsilon \cap [y_*, y^*])$, *where* $\mathcal{Y}'_\varepsilon := \{y \in \Re \,|\, y+1 \in \mathcal{Y}_\varepsilon\}$

$$y_* := \min\{\min_{y\in\mathcal{Y}} y\,,\, t_- - t_+\}, \qquad\qquad y^* := \max\{\max_{y\in\mathcal{Y}} y\,,\, t_+ - t_-\}\,,$$

and the indexes j_- *and* j_0 *refer to* $y_{j_-} = -1$ *and* $y_{j_0} = 0$. *Finally,* $\{y_j\}_{j\in\mathcal{I}}$ *is an ordered list of the elements of* $\mathcal{Y}'_\varepsilon \cap]y^*, \infty[$ *with* $\mathcal{I} := \{M+1, M+2, \ldots, \}$ *while* $z_j := \varepsilon(y_j+1)^{-1}$ *for* $j = j_0+1, j_0+2, \ldots$.

Note that in Proposition 5 we impose that the set of kinks of $\varepsilon(\cdot)^{-1}$ extends arbitrarily far in the positives. This is somehow without loss of generality since if $\mathcal{Y}_\varepsilon$ is bounded then one can always create an infinite series of artificial kinks (with linear functions of same slope on both sides) at increasingly far locations on the real line. Alternatively, if $\mathcal{Y}_\varepsilon$ happens to be finite, a finite dimensional reformulation can be obtained using a similar analysis as below. We omit this reformulation for the sake of brevity but simply mention that it requires imposing that $f(y^*)$ be smaller than the Lipschitz constant of $\varepsilon(\cdot)^{-1}$.

Given that Proposition 5 assumes $\mathcal{Y}_\varepsilon$ to be unbounded, it must be that $\mathcal{I}$ is an infinite set thus the linear program (18) has an infinite number of decision variables which makes it impossible to solve using conventional solvers. For this reason, we propose to employ Algorithm 1 which implements a decomposition scheme that is based on the column generation approach (see Desrosiers and Lübbecke (2005) and reference therein).

Algorithm 1 Column generation for solving (PRSR-Opt) problem (18)

1: $\mathcal{I}' \leftarrow \emptyset$
2: **repeat**
3: Solve problem (18) with $\mathcal{I}$ restricted to $\mathcal{I}'$
4: $\alpha_M^* \leftarrow$ dual variables for constraint (18h) at optimum of restricted problem
5: $\beta_M^* \leftarrow$ dual variables for constraint (18j) at optimum of restricted problem
6: $j' \leftarrow \arg\min_{j\in\mathcal{I}} z_j - 1 - \alpha_M^* - \beta_M^*(y_j - y_M)$
7: $\mathcal{I}' \leftarrow \mathcal{I}' \cup \{j'\}$
8: **until** $z_{j'} - 1 - \alpha_M^* - \beta_M^*(y_{j'} - y_M) \geq 0$

While Algorithm 1 is not guaranteed to terminate in a finite number of iterations, it can be interrupted at any iteration to produce a conservative approximation for x. We also expect that in practice the stopping criterion will quickly be met especially when a large amount of confidence intervals (such that K is large in Definition 3) are used which should make constraint (45b) become redundant. We are left with the task of identifying an efficient procedure for completing step 6 of the algorithm. The proposition below addresses this.

Proposition 6 *Let* $\varepsilon(\cdot)^{-1}$ *be a piecewise-linear approximation of a convex non-decreasing function* $\bar{\varepsilon}(\cdot)^{-1}$ *such that* $\varepsilon(y)^{-1} \geq \bar{\varepsilon}(y)^{-1}$ *for all* $y \in \mathbb{R}$ *while* $\varepsilon(y)^{-1} = \bar{\varepsilon}(y)^{-1}$ *for all* $y \in \mathcal{Y}_\varepsilon$. *Then, an optimal solution to*

$$y^*_{\alpha,\beta} := \arg \min_{y\in\mathcal{Y}'_\varepsilon \cap]y^*,\,\infty[} \varepsilon(y+1)^{-1} - 1 - \alpha - \beta(y - y^*) \tag{19}$$

can be found by solving

$$\bar{y}^*_{\alpha,\beta} := \arg \min_{y\in[y^*,\,\infty[} \bar{\varepsilon}(y+1)^{-1} - 1 - \alpha - \beta(y - y^*) \tag{20}$$

and letting $y^*_{\alpha,\beta} := \lceil \bar{y}^*_{\alpha,\beta} \rceil$ *if the set* $\mathcal{Y}'_\varepsilon \cap (y^*, \bar{y}^*_{\alpha,\beta}) = \emptyset$, *otherwise using*

$$y^*_{\alpha,\beta} := \arg \min_{y\in\{\lfloor \bar{y}^*_{\alpha,\beta}\rfloor, \lceil \bar{y}^*_{\alpha,\beta}\rceil\}} \varepsilon(y+1)^{-1} - 1 - \alpha - \beta(y - y^*)$$

where

$$\lfloor y \rfloor := \sup\{y' \in \mathcal{Y}'_\varepsilon \cap (y^*,\,\infty) \;:\; y' \leq y\}, \qquad \lceil y \rceil := \inf\{y' \in \mathcal{Y}'_\varepsilon \cap (y^*,\,\infty) \;:\; y' \geq y\}\,.$$

Based on Proposition 6, we can solve (19) by solving problem (20) and making a follow-up projection of the optimal solution on $\mathcal{Y}_\epsilon \cap [y^*, \infty)$. This will effectively reduce complexity of implementing Step 6.

We conclude this section with an example involving $\bar{\varepsilon}(y) := \exp(-\bar{\lambda}y)$ as described in Section 2. In this case, it is clear that $\bar{\varepsilon}(y)^{-1} = \exp(\bar{\lambda}y)$ does not satisfy the condition imposed in Proposition 5, but for a given discretization of $\mathbb{R}^+$ such as $\mathcal{Y}_\varepsilon := \{\Delta_y^k\}_{k=0}^\infty$, one can conservatively approximate $\bar{\varepsilon}(\cdot)^{-1}$ by letting $\varepsilon(\cdot)^{-1}$ be the piecewise linear inner approximation of $\bar{\varepsilon}(\cdot)^{-1}$ with kinks at $\mathcal{Y}_\varepsilon$ where the two functions return the same values. To better illustrate this, Figure 1 presents both $\bar{\varepsilon}(y)$ and $\varepsilon(y)$ and the associated bounds are imposed on the loss function with $\bar{\lambda} = 0.6946$, $l(y) := \max(3y, y)$ and $\mathcal{Y}_\varepsilon := \{2^k\}_{k=0}^\infty$. With this particular choice of characterization for $\varepsilon(y)$, Algorithm 1 can be used to solve the (PRSR-Opt) problem (18). Following the result of Proposition 6, the new candidate j' can be obtained by first solving

$$\bar{y}^*_{\alpha,\beta} := \arg \min_{y \in [y^*, \infty[} \exp(\bar{\lambda}(y+1)) - 1 - \alpha - \beta(y - y^*) = \begin{cases} y^* & \text{if } \beta \leq \bar{\lambda}\exp(\bar{\lambda}(y^*+1)), \\ \frac{1}{\bar{\lambda}} \ln\left(\frac{\beta}{\bar{\lambda}}\right) - 1 & \text{otherwise} \end{cases}$$

and, if $\bar{y}^*_{\alpha,\beta} > y^*$, then identifying the index j' in $\mathcal{I}$ such that

$$j' = \arg \min_{k \in \{\lfloor \ln(\bar{y}^*_{\alpha,\beta})/\ln(2) \rfloor, \lceil \ln(\bar{y}^*_{\alpha,\beta})/\ln(2) \rceil\}} \exp(\bar{\lambda}(2^k+1)) - 1 - \alpha - \beta(2^k - y^*),$$

where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ refer respectively to the standard ceil and floor operators.

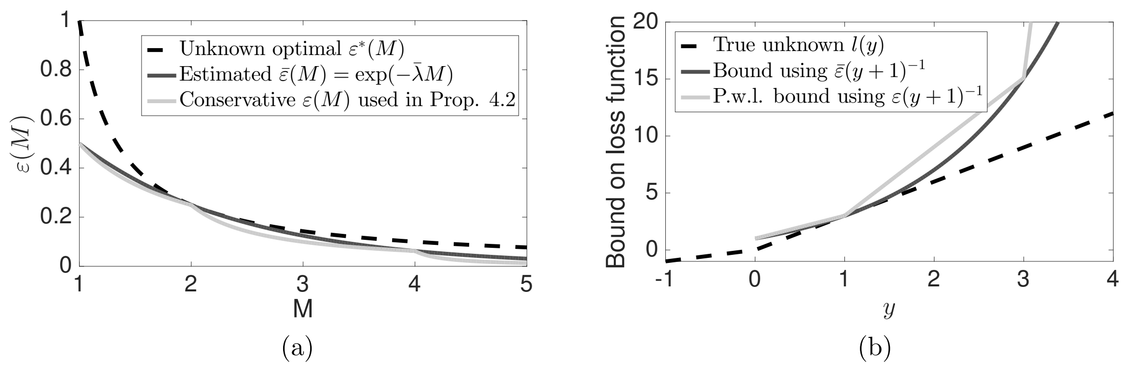

Figure 1: Illustration of the effect of manipulating $\bar{\varepsilon}$ in order to satisfy the assumptions made in propositions 5 and 6 thus making the resolution of problem (18) more tractable using Algorithm 1.

4 Discrete approximation of (PRSR-Opt) when P is continuous

A key condition for tractable reformulation of problem (4) in the previous section is that P must be discrete. In this section we concentrate on the case that P is continuous and we propose a discretization scheme for it.

4.1 Problem set-up

By the definition of the preference robust normalized risk measure, we can write problem (4) as

$$\begin{array}{ll} \min\limits_{x \in X, t \in \mathbb{R}} & t \\ \text{s.t.} & \sup\limits_{l \in L} \mathbb{E}_P[l(c(x,\xi) - t)] - l(0) \leq 0. \end{array} \tag{21}$$

Let P_N be a discrete approximation of P. We consider

$$\begin{array}{ll} \min\limits_{x \in X, t \in \mathbb{R}} & t \\ \text{s.t.} & \sup\limits_{l \in L} \mathbb{E}_{P_N}[l(c(x,\xi) - t)] - l(0) \leq 0. \end{array} \tag{22}$$

In the literature of stochastic programming, there are various approaches to construct P_N such as Monte Carlo sampling and quasi-Monte Carlo sampling method. In this paper, we will use empirical probability distribution constructed through independent and identically distributed (iid) samples.

To ease the exposition, let $v(x,t) := \sup_{l\in L} \mathbb{E}_P[l(c(x,\xi) - t)] - l(0)$ and $v_N(x,t) := \sup_{l\in L} \mathbb{E}_{P_N}[l(c(x,\xi) - t)] - l(0)$. Consequently, we can rewrite (21) and (22) as

$$\text{(PRSR-Opt)} \quad \min_{x\in X, t\in \mathbb{R}} \ t \quad \text{subject to} \quad v(x,t) \leq 0, \tag{23}$$

and

$$\text{(PRSR-Opt-N)} \quad \min_{x\in X, t\in \mathbb{R}} \ t \quad \text{subject to} \quad v_N(x,t) \leq 0. \tag{24}$$

Let $\mathcal{F}, S$ and ϑ denote the feasible set, the set of optimal solutions and the optimal value of problem (PRSR-Opt) respectively. Likewise, we define $\mathcal{F}_N, S_N$ and ϑ_N for problem (PRSR-Opt-N). Throughout this section, we make the following assumption.

Assumption 1 *We assume: (a) X is a compact set, (b) $c(\cdot,\cdot)$ is a continuous function on $X \times \Xi$, (c) (PRSR-Opt) satisfies Slater condition, i.e., there exist a positive constant number θ, and $x_0 \in X, t_0 \in \mathbb{R}$ such that*

$$\sup_{l\in L} \mathbb{E}_P[l(c(x_0,\xi) - t_0)] - l(0) \leq -\theta, \tag{25}$$

(d) let $Z := \min_{x\in X} c(x,\xi)$, $\mathbb{E}_P[Z] < +\infty$.

Under Assumption 1, the optimal value ϑ of (PRSR-Opt) is finite. To see this, we note that

$$\sup_{l\in L} \mathbb{E}_P[l(Z - t_0)] - l(0) \leq \sup_{l\in L} \mathbb{E}_P[l(c(x_0,\xi) - t_0)] - l(0) \leq -\theta,$$

which implies $\vartheta \leq t_0$. On the other hand, by Jensen's inequality, $\mathbb{E}_P[l(Z-t)] \geq l(\mathbb{E}_P[Z] - t)$, thus $\sup_{l\in L} \mathbb{E}_P[l(Z-t)] \geq \sup_{l\in L} l(\mathbb{E}_P[Z] - t) \to +\infty$ as $t \to -\infty$, we deduce that the t component of the feasible set $\mathcal{F}$ must have a lower bound and hence the optimal value $\vartheta > -\infty$.

4.2 Sample average approximation

Let $\xi^1, \ldots, \xi^N$ be iid samples of ξ and $P_N(\cdot) := \frac{1}{N}\sum_{k=1}^N \mathbb{1}_{\xi^k}(\cdot)$ the empirical probability measure, where $\mathbb{1}_{\xi^k}(\cdot)$ is an indicator function with $\mathbb{1}_{\xi^k}(\xi) = 0$ for $\xi = \xi^k$ and 0 otherwise. Instead of deriving uniform approximation specifically for $v_N(x,t)$ defined in (24), we establish a general convergence result which may be of interest beyond this paper.

Let $g : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a continuous function and $\mathcal{W}$ be a set of continuous functions of $\phi : \mathbb{R} \to \mathbb{R}$. Let $\Phi(x) := \sup_{\phi\in\mathcal{W}} \mathbb{E}_P[\phi(g(x,\xi))]$ and $\Phi_N(x) := \sup_{\phi\in\mathcal{W}} \mathbb{E}_{P_N}[\phi(g(x,\xi))]$. In what follows, we show uniform convergence of $\Phi_N(x)$ to $\Phi(x)$ under some appropriate conditions. Note that in the literature of stochastic programming, there have been a number of recent results on uniform convergence of sample average approximation of a random function, see Shapiro and Xu (2008) and references therein. Here we consider a slightly different setting where it concerns uniform convergence of the maximum of a class of sample average approximated random functions as opposed to a single sample averaged approximated function in the literature.

In order to establish the desired convergence results, we need to impose some conditions on $\mathcal{W}$ and g.

Assumption 2 *Let Ξ be the support set of ξ.*

(a) $\phi(0) = 0$ and for any $\epsilon > 0$, there exists a compact set $\Xi_\epsilon \subset \Xi$ such that

$$\sup_{N, x\in X, \phi\in\mathcal{W}} \mathbb{E}_{P_N}[|\phi(g(x,\xi)\mathbb{1}_{\Xi\setminus\Xi_\epsilon}(\xi))|] \leq \epsilon, \text{ w.p.1.} \tag{26}$$

(b) *For any $M > 0$, there exists a positive constant κ_M such that*

$$\sup_{\phi\in\mathcal{W}} |\phi(z_1) - \phi(z_2)| \leq \kappa_M |z_1 - z_2|, \forall z_1, z_2 \in [-M, M], \tag{27}$$

where κ_M increases as M increases.

(c) *There exist a measurable function $r(\xi) : \Xi \to \mathbb{R}_+$ and a constant $\nu > 0$ such that*

$$|g(x,\xi) - g(x',\xi)| \leq r(\xi)\|x - x'\|^\nu, \forall x, x' \in X, \xi \in \Xi. \tag{28}$$

Assumption 2 (a) is a kind of uniform integrability condition for all $\phi \in \mathcal{W}$. The condition is well known in probability theory, see Chapter 3 in Billingsley (1999). Condition (b) requires the class of functions in $\mathcal{W}$ to be equi-Lipsthitz continuous over $[-M, M]$ for any $M > 0$. Condition (c) requires g to be Hölder continuous in x. The condition is used in Theorem 5.2 of Shapiro and Xu (2008).

Let

$$\mathcal{W}_M := \{\phi|_{[-M,M]}(\cdot) : \forall \phi(\cdot) \in \mathcal{W}\}, \tag{29}$$

where $\phi|_{[-M,M]}(\cdot)$ stands for the restriction of function $\phi(z)$ to being defined over interval $[-M, M]$. Under Assumption 2 (a)-(c), the set $\mathcal{W}_M$ is bounded by

$$\sup_{\phi\in\mathcal{W}_M} \|\phi\|_\infty \leq \kappa_M M. \tag{30}$$

By Ascoli-Arzela Theorem (Brown (2004)), the equi-Lipschitz continuity condition (27) and the unform boundedness (30) guarantee that $\mathcal{W}_M$ is relatively compact (albeit it is not necessarily compact). Note that the relative compactness of $\mathcal{W}_M$ ensures existence of an ϵ-net of $\mathcal{W}_M$, that is, for any $\epsilon > 0$, there exists a set of finite number of functions $\{\phi_1, \ldots, \phi_K\} \subset \mathcal{W}_M$ such that

$$\mathcal{W}_M = \bigcup_{k=1}^{K} (\mathcal{W}_M)_k^\epsilon \tag{31}$$

where $(\mathcal{W}_M)_k^\epsilon := \{\phi \in \mathcal{W}_M : \|\phi - \phi_k\|_\infty \leq \epsilon\}$ for $k = 1, \ldots, K$.

Lemma 1 *Let Assumption 2 hold. Then for any $\delta > 0$, there exist positive constants $\epsilon < \delta/4$, $C(\epsilon, \delta)$ and $\beta(\epsilon, \delta)$, independent of N such that*

$$\text{Prob}\left(\sup_{x\in X} \left|\Phi_N(x) - \Phi(x)\right| \geq \delta\right) \leq C(\epsilon, \delta) e^{-N\beta(\epsilon,\delta)}. \tag{32}$$

Note that in Lemma 1, the probability measure "Prob" should be understood as the product probability measure of P over measurable space $\Xi \times \Xi \times \cdots$ with product Borel sigma algebra $\mathscr{B} \times \mathscr{B} \times \cdots$.

A crucial requirement in Lemma 1 is Assumption 2 (a) which is used to ensure the relative compactness of $\mathcal{W}_M$. In the case when Ξ is compact, this condition holds automatically.

Corollary 1 *Let Ξ be compact and Assumptions 2 (b)-(c) hold. Then for any $\delta > 0$, there exist positive constants $\epsilon < \delta/2$, $C(\epsilon, \delta)$ and $\beta(\epsilon, \delta)$ independent of N such that*

$$\text{Prob}\left(\sup_{x\in X} \left|\Phi_N(x) - \Phi(x)\right| \geq \delta\right) \leq C(\epsilon, \delta) e^{-N\beta(\epsilon,\delta)}.$$

Note that Haskell et al. (2017) considered similar discretization approaches for approximating integral stochastic dominance constraints whereby they used piecewise linear increasing convex functions to form an ϵ-net the associated utility functions. They established exponential rate of convergence under the condition that the utility functions are Lipschitz continuous and defined on a compact set. Here we relax the compactness condition by replacing it with uniform integrability condition, this will effectively allow us to apply the convergence results to (PRSR-Opt) where the utility loss functions are defined on $\mathbb{R}$ rather than a compact set. Note also that the discretization scheme should be distinguished from those in Hu and Mehrotra (2015) whose focus is on piecewise linear approximation of the utility function of a robust preference optimization problem rather than sample average approximation of the expected utility.

4.3 Convergence of the optimal values and optimal solutions

We now return to discuss convergence of (PRSR-Opt-N) to (PRSR-Opt) in terms of the optimal values and optimal solutions. Let $v_N(x,t)$ and $v(x,t)$ be defined as in (24) and (23). We start by deriving uniform convergence of $v_N(x,t)$ to $v(x,t)$ using Lemma 1. To this end, we need to make some appropriate conditions on $c(x,\xi)$ and the set of loss functions L which correspond to the conditions that we set for $g(x,\xi)$ and $\phi \in \mathcal{W}$ in Section 5.2.

(C1) Let T be a compact set in $\mathbb{R}$. For any $\epsilon > 0$, there exists a compact set $\Xi_\epsilon \subset \Xi$ such that

$$\sup_{N,x\in X,t\in T,l\in L} \mathbb{E}_{P_N}[|l((c(x,\xi)-t)\mathbb{1}_{\Xi\setminus\Xi_\epsilon}(\xi))|] \le \epsilon.$$

(C2) For any $M > 0$, there exist positive constants κ_M (depending on M), Δ_M and $\lambda \in [-M, M]$ such that $\sup_{l\in L} |l(z_1)-l(z_2)| \le \kappa_M|z_1-z_2|, \forall z_1, z_2 \in [-M, M]$, and $\sup_{l\in L_M} \|l(\lambda)\| \le \Delta_M$ with $L_M := \{l|_{[-M,M]}(\cdot) : \forall l \in L\}$.

(C3) There exist a measurable function $r : \Xi \to \mathbb{R}_+$ and a constant $\nu > 0$ such that

$$|c(x,\xi)-c(x',\xi)| \le r(\xi)\|x-x'\|^\nu, \forall x, x' \in X, \xi \in \Xi.$$

Theorem 1 *Let conditions* (C1)–(C3) *hold. Let T be a compact set in $\mathbb{R}$. Then for any $\delta > 0$ there exist positive constants ϵ, $N(\epsilon,\delta)$, $C(\epsilon,\delta)$ and $\beta(\epsilon,\delta)$ independent of N such that for $N \ge N(\epsilon,\delta)$*

$$\mathrm{Prob}\left(\sup_{x\in X,t\in T}\left|v_N(x,t)-v(x,t)\right| \ge \delta\right) \le C(\epsilon,\delta)e^{-N\beta(\epsilon,\delta)} \tag{33}$$

We are now ready to state the main results of this section concerning convergence of (PRSR-Opt-N) to (PRSR-Opt) as the sample size N increases.

Theorem 2 *Let ϑ, ϑ_N, S and S_N be defined as in Section 5.1 and P_N be the associated empirical probability measure. Let Assumption 1 and conditions (C1)–(C3) hold. Suppose that for almost every $\xi \in \Xi$, $c(\cdot,\xi)$ is a convex function. Then*

(i) For any $\delta \le \theta$,

$$\mathrm{Prob}\left(|\vartheta_N-\vartheta| \ge \delta\right) \le C(\epsilon,\varepsilon)e^{-N\beta(\epsilon,\varepsilon)}, \tag{34}$$

for $N \ge N(\epsilon,\varepsilon)$ where $N(\epsilon,\varepsilon)$, $C(\epsilon,\varepsilon)$ and $\beta(\epsilon,\varepsilon)$ are defined as in Theorem 1 and ε is some positive constant depending on δ, and θ is given in (25).

(ii) Let $\{x_N, t_N\}$ be a sequence of optimal solutions obtained from solving (PRSR-Opt-N). Then with probability 1, a cluster point of the sequence is an optimal solution of (PRSR-Opt).

Remark 3 *Theorem 2 ensures ϑ_N converges to ϑ at exponential rate with increase of the sample size N. A key condition for the established convergence results is equi-continuity of the ambiguity set of loss functions L over any compact set. This raises a question as to whether the loss functions defined in the ambiguity sets proposed in Section 2 satisfy this condition. Following Remark 2, we know that condition (C2) is satisfied by the set of loss functions $L := L_{ubsr} \cap L_{ce}(\mathcal{W}) \cap L_{bnd}$ defined in Proposition 3. This means the bound L_{bnd} on the sensitivity of risk measures over large tail losses provides a sufficient condition for the asymptotic consistency of the optimal value and the optimal solutions of problem (PRSR-Opt-N).*

5 Numerical experiments

In this section, we repeat the experiments performed in Delage and Li (2018) regarding the comparison of different choices of performance measure one might employ in a portfolio selection problem where only partial information is available about the decision maker's preference regarding a risk measure. In particular, we consider a financial advisor that makes different hypothesis about the way an investor he is consulting with perceives risks before formulating the following portfolio optimization problem:

$$\min_{x\geq 0} \rho(\xi^T x) \quad \text{subject to} \quad \sum_{i=1}^{n} x_i = 1,$$

where each x_i is a decision variable describing the percentage of the budget invested in asset i, while $\xi \in \mathbb{R}^n$ is a random vector following a distribution P and describing the random weekly return of each asset available for investment. Note that the units of the payoff in this context need to be seen as percentage of total wealth hence the risk level and axiomatic assumptions should be interpreted accordingly.[2] A naïve approach for designing a portfolio when $\rho(\cdot)$ is unknown consists of simply minimizing the expected loss (as a percentage of initial wealth) of the portfolio, i.e. $\rho(\xi^T x) := \mathbb{E}[-\xi^T x]$, or some arbitrarily chosen expectile measure $\rho(\xi^T x) := \mathrm{SR}^P_{l_\tau}(\xi^T x)$ with $l_\tau(s) := \max(\tau s,\, (1-\tau)s)$. We compare such two approaches to an approach that is robust with respect to the limited amount of preference information. Namely, we consider two preference robust risk measures $\varrho_{\mathcal{R}_{LE}}(\cdot)$ and $\varrho_{\mathcal{R}_{CLE}}(\cdot)$ which account for the fact that the risk measure is law invariant and respectively convex or coherent as proposed in Delage and Li (2018), and two preference robust risk measures that additionally account for the fact that the true risk measure is a utility-based SR, namely $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}}(\cdot) = \mathrm{SR}^P_{L_{ubsr}\cap L_{ce}(\mathcal{W})}(\cdot)$ and $\varrho_{\mathcal{R}_{coh}\cap\mathcal{R}_{ce}}(\cdot) = \mathrm{SR}^P_{l_\tau}(\cdot)$ with the worst-case τ defined in equation (10).

Our experiments are nearly identical to the experiments described in Delage and Li (2018). Specifically, they are designed based on historical stock market data for 334 companies that were part of the S&P 500 Index during the period from January 1994 until December 2013.[3] Each experiment consists in drawing 4 assets randomly from the pool of 334 assets and a set of 13 consecutive weeks in the period from January 2004 to December 2013. We require each method to propose a portfolio that considers the weekly return vector ξ be drawn from the joint empirical distribution of the 4 assets in the selected 13 weeks. We also simulate elicitation by using a reference investor modeled using $\bar{\rho}(Z) := \mathrm{SR}^P_{l_{0.6}}(Z)$ to evaluate the certainty equivalent of up to 20 random payoffs W_k constructed based on a random choice of one asset among the 334 assets and a random set of 13 consecutive weeks in the period between January 1994 and December 2003.

Figure 2 presents the average perceived risk in lost percentage points (i.e. $\bar{\rho}(\xi^T x)\times 100$) achieved, in a set of 4000 experiments, by the portfolios obtained using either expected loss minimization, the wrong expectile measure (i.e. SR_{l_τ} with $\tau = 0.75$ instead of $\tau = 0.6$), or either of the four preference robust risk measures described above with certainty equivalent information about up to 20 random payoffs (including the null payoff). We also report the best average perceived risk that could be obtained if $\bar{\rho}$ was exactly known. Once again, in this set of experiments, we observe that preference robust risk minimization model eventually outperforms the methods that are based on the wrong risk measures (namely using the expected loss or the wrong expectile measure) after a sufficient amount of elicitation (about 10 certainty equivalent evaluations here). Interestingly, these experiments seem to indicate that information about whether the risk measure is coherent or not is more valuable than the information about whether it is a shortfall risk measure. Indeed, one can observe that the quality of portfolios only improves marginally when using $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}}$ instead of $\varrho_{\mathcal{R}_{LE}}$ (i.e. introducing the hypothesis of having a utility-based SR), whereas the improvement is much more significant when using $\varrho_{\mathcal{R}_{coh}\cap\mathcal{R}_{ce}}$ instead of $\varrho_{\mathcal{R}_{CLE}}$ (introducing the hypothesis of coherence). Additionally, one can observe that if the risk measure is coherent and it is a member of the utility-based SR, then it is already uniquely identified after a single certainty equivalent evaluation.

We also carried out the same experiments with the preference robust risk measure $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})\cap\mathcal{R}_{bnd}(\varepsilon)}$ but quickly realized that the portfolios obtained using this method were almost undistinguishable from the ones obtained using $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})}$ in our test environment. In particular, in order to obtain a piecewise linear function for $\varepsilon(\cdot)^{-1}$ that satisfies the conditions of Proposition 6, we started by identifying a value $\bar{\lambda}$ such that $\bar{\varepsilon}(y) := \exp(-\bar{\lambda}y)$ satisfies Definition 4 for our reference investor modeled with $\bar{\rho}$. As seen in Proposition 3, this can be done using any λ such that $\max(\tau y, (1-\tau)y) \leq (\exp(\lambda y) - 1)(1-\tau)$ for all $y \geq 0$. In our experiments, we used

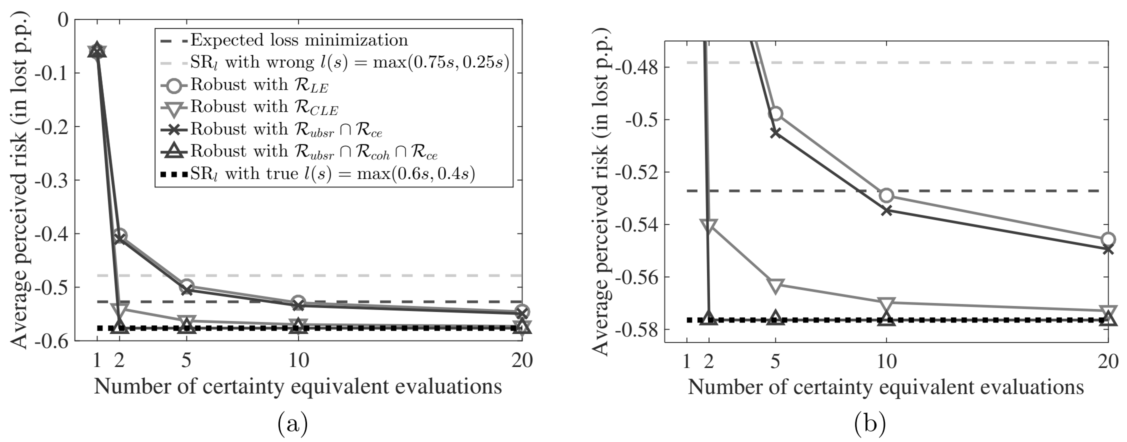

Figure 2: Comparison of the average perceived risk (in lost percentage points) for the portfolios obtained using either expected loss minimization, the wrong expectile (SR_{l_τ} with $\tau = 0.75$), or the minimization of a preference robust risk measure with certainty equivalent evaluations for up to 20 random payoffs (including the null payoff) in a set of 4000 experiments. We also report the best average perceived risk that could be obtained if the representation of this perception was exactly known. (b) presents in more detail the portion of figure (a) which achieves an average perceived risk between -0.48 and -0.58 percentage points.

$\bar{\lambda} = \max_{y\geq 0} \ln((\tau/(\tau-1))y+1)/(y+1) \approx 0.4716$. Once $\bar{\varepsilon}(y)$ was selected, we considered $\varepsilon(y)^{-1}$ to be the piecewise linear inner approximation which matches $\bar{\varepsilon}(y)^{-1}$ exactly at the points in the discrete set $\mathcal{Y}_\varepsilon := \{1.2^k\}_{k=0}^{\infty}$. In a preliminary set of experiments, we observed that Algorithm 1 would converge in less than 6 iterations (about 3 on average) and always returned a solution that was very close (if not exactly the same) to the solution of problem (17). This motivated us to conclude that the performance of $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})\cap\mathcal{R}_{bnd}(\varepsilon)}$ in the 4000 experiments reported in Figure 2 should be considered the same as the performance of $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}}$. Moreover, this evidence seem to indicate that in a context where the distribution of ξ would be continuous and be approximated using sample average approximation, the computational cost of employing $\varrho_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})\cap\mathcal{R}_{bnd}(\varepsilon)}$ instead of $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}}$ in order to obtain a guarantee on the convergence of (PRSR-Opt-N) to (PRSR-Opt) (see Theorem 2) is in fact reasonable.

6 Conclusion

In this paper, we considered a preference robust risk minimization problem for which the risk measure is assumed to be a normalized convex utility-based shortfall risk measure. We demonstrated for the first time that $\varrho_{\mathcal{R}}(Z)$ could equivalently be represented using SR_L^P where L is the set of all plausible loss functions that could be used to characterize $\rho(\cdot)$ using SR_l^P. We also showed how this ambiguity set L can be constructed based on available information regarding positive homogeneity, pairwise comparisons involving a lottery and a certain amount, and finally the existence of a set of random variables with arbitrarily large extreme values which are considered less risky than a fixed loss of one. We established how the risk minimization problem can be reformulated as a linear program when such information is available and P is discrete. In particular, for the case of positive homogeneity the preference robust risk minimization problem reduces to a problem in which the loss function is unambiguous. We then considered the quality of solutions that can be recovered from these models using sample average approximation (SAA) schemes when the distribution P is continuously supported. While convergence analysis of the optimal values and the optimal solutions of sample average approximated problems is well documented in the literature of stochastic programming (see Ruszczyński and Shapiro (2003)), the convergence result that we established in Section 5 extends the existing results by Haskell et al. (2017) on sample average approximation of robust expected utility optimization problems to a broader class of utility functions and paves the way for application of the tractable numerical schemes developed in Section 4 to continuously distributed preference robust shortfall risk optimization problems. The discretization scheme can also be incorporated into Hu and

Mehorotra's piecewise linear utility approximation approach for solving preference robust optimization problems (Hu and Mehrotra (2015)). Finally, we presented some numerical experiments in which it is possible to quantify the value of exploiting the information that a risk measure is a utility-based SR in combination with certainty equivalent information for a small set of random payoffs.

Endnotes

1. Note that elicited preference information needs to be distinguished from the property of elicitability of risk measures. Specifically, the latter requires the existence of statistically robust procedures for estimating the measure and of a proper backtesting mechanism for prediction schemes (see Cont et al. (2010) for a discussion).

2. Alternatively, one could also redefine each x_i as the amount of actual money invested in each asset if this is needed for a more accurate interpretation of $\rho(\xi^T x)$.

3. This is the same period and same companies as in Delage and Li (2018) except for BMC software which was removed from our data set given that it was privatized in September 2013.

Proofs of statements

Appendix A1 Proofs of Section 2

A1.1 Bounds on utility-based shortfall risk

Lemma A.1 *Let ρ be a normalized utility-based shortfall risk measure ρ. Given any random variable $Z \in \mathscr{L}_p$, one has that $-\mathrm{esssup}(Z) \leq \rho(Z) \leq -\mathrm{essinf}(Z)$.*

Proof. This follows from the monotonicity of convex risk measure, namely for all $Z_1 \geq Z_2$ it must be that $\rho(Z_1) \leq \rho(Z_2)$. Specifically, if $\mathrm{essinf}(Z) \in \mathbb{R}$, since $Z \geq \mathrm{essinf}(Z)$ with probability 1, we can conclude that $\rho(Z) \leq \rho(\mathrm{essinf}(Z)) = -\mathrm{essinf}(Z)$, while otherwise $\mathrm{essinf}(Z) = -\infty$ hence $\rho(Z) \leq \infty$ follows trivially. One can establish a similar result with respect to $\mathrm{esssup}(Z)$. □

A1.2 Proof of Proposition 1

We start by describing a set of useful properties satisfied by the functions in L_{ubsr}, by normalized utility-based shortfall risk measures, and their robust versions. Readers who are familiar with the properties of utility-based risk measures may skip to our main proof in subsection A1.2.2.

A1.2.1 Useful properties of normalized utility-based shortfall risk measure and their loss functions

First, let us formally state the definition of L_{ubsr} for ease of future reference.

Definition A.1 *Let L_{ubsr} be the set of all convex non-decreasing functions $l : \mathbb{R} \to \mathbb{R}$ that are strictly increasing over $[z_0, \infty)$ for some $z_0 < 0$.*

It is well-known that loss functions in L_{ubsr} satisfy the following properties.

Lemma A.2 (Properties of $l \in L_{ubsr}$) *The following assertions hold for all $l \in L_{ubsr}$:*

(i) If there are two points $a < b$ in the domain of l such that $l(a) = l(b)$, then $l(t) = l(a) = l(b)$ for $t \in (-\infty, b]$.
(ii) $l(t) \to +\infty$ as $t \to +\infty$.
(iii) $l(t) < l(0)$ for all $t < 0$ and $l(t) > l(0)$ for all $t > 0$.
(iv) $l(0)$ lies in the interior of $l(\mathbb{R})$.

Proof. Part (i). First, since $l(\cdot)$ is non-decreasing, it is clear that $l(t) = l(a)$ for all $t \in [a, b]$ and that $l(t) \leq l(a)$ for all $t < a$. Next, let's assume that there is a point $t < a$ such that $l(t) < l(a)$, then we necessarily have that for $\theta := (a - t)/(b - t) \in [0, 1]$:

$$\theta l(b) + (1 - \theta)l(t) < \theta l(b) + (1 - \theta)l(a) = l(a) = l(\theta b + (1 - \theta)t),$$

which contradicts the fact that $l(\cdot)$ is convex.

Part (ii). Since $l(\cdot)$ is non-decreasing and non-constant, we can find two points $t_1 < t_2$ such that $l(t_1) < l(t_2)$. By the convexity of $l(\cdot)$, $l(t) - l(t_2) \geq [(l(t_2) - l(t_1))/(t_2 - t_1)](t - t_2)$ for $t > t_2$. By driving t to infinity, we immediately get the conclusion.

Part (iii). This follows directly from the fact that $l(\cdot)$ is strictly increasing over some interval $[z_0, \infty)$ with $z_0 < 0$ and non-decreasing over $(-\infty, z_0]$.

Part (iv). This follows directly from (iii). □

Equipped with Lemma A.2, one can easily derive a representation theorem for normalized convex utility-based shortfall risk measure. Note that the result is usually used without the need of a proof by the community (see for example Section 4.9 of Föllmer and Schied (2016)), yet we include a detailed argument below for completeness.

Lemma A.3 *Let $\rho(\cdot)$ be a utility-based shortfall risk measure. Then $\rho(\cdot)$ is convex and normalized (i.e. $\rho(0) = 0$) if and only if there is some $l \in L_{ubsr}$ such that*

$$\rho(Z) := \inf\{t : \mathbb{E}[l(-Z - t)] - l(0) \leq 0\}. \tag{35}$$

Proof. By the definition of SR, there is a non-decreasing function $l : \mathbb{R} \to \mathbb{R}$ which is not constant such that $\rho(Z)$ can be represented as in (1). By (Weber 2006, Corollary 3.1), $\rho(\cdot)$ is convex if and only if $l(\cdot)$ is convex. So we are left with the task to show that $\rho(0) = 0$ if and only if l is strictly increasing over $[z_0, \infty)$ for some $z_0 < 0$ and $\lambda = l(0)$.

The "only if" part. Recall that in the definition of the utility-based shortfall risk measure, λ lies in the interior of the range of $l(\cdot)$. Thus, given that real convex functions are continuous, there must exist a $t_0 \in \mathbb{R}$ such that $l(t_0) = \lambda$. By using a similar argument as in the proof of Lemma A.2 (i), we can show that there exists $\bar{t} < t_0$ such that $l(t) < l(t_0)$ for all $t \in [\bar{t}, t_0)$ and $l(\cdot)$ is strictly increasing over the interval because otherwise $l(t_0) = \lambda$ would lie at the lower boundary of the range of $l(\cdot)$. Likewise, by the non-decreasing and convex nature of $l(\cdot)$, we can show that $l(t) > l(t_0)$ for any $t > t_0$ and $l(\cdot)$ is strictly increasing over $[t_0, \infty]$. Thus, we are left with the task to demonstrate that $t_0 = 0$ but this follows from the fact that $\mathrm{SR}_l^P(0) = 0$ and that

$$\mathrm{SR}_l^P(0) = \inf\{t : \mathbb{E}[l(-t)] \leq l(t_0)\} = \inf\{t : l(-t) \leq l(t_0)\} = -t_0.$$

The " if" part is relatively easier to prove since by definition $l \in L_{ubsr}$ satisfies the conditions needed for equation to be the representation of a utility-based shortfall risk measure and since the convexity of l implies that the risk measure ρ is a convex risk measure following (Weber 2006, Corollary 3.1). Moreover, we also have that $\lambda := l(0)$ lies in the interior of $l(\mathbb{R})$ based on Lemma A.2(iv). One can finally verify that $\rho(0) = \inf\{t : \mathbb{E}[l(-t)] - l(0) \leq 0\} = 0$ following Lemma A.2(iii). □

We now turn to a useful property of normalized convex utility-based shortfall risk which will be of use later. We encourage the reader to the proof of Proposition 4.113 in Föllmer and Schied (2016) for the supporting arguments.

Lemma A.4 *Let $\rho(Z)$ be a normalized convex utility-based shortfall risk measure associated to some loss function $l \in L_{ubsr}$. Then for all $Z \in \mathscr{L}_p(\Omega, \mathscr{B}, P)$, the risk $\rho(Z)$ is equal to the unique solution t^* of the equation $\mathbb{E}_P[l(-Z - t)] = l(0)$.*

Finally, we conclude this subsection with a useful property of robust normalized convex risk measures.

Lemma A.5 *Let ρ_i with $i \in \mathcal{I}$ be a family of normalized convex risk measures on $\mathscr{L}_p(\Omega, \mathscr{B}, P)$ with associated acceptance sets $\mathcal{A}_i$ and let $\bar{\rho}(X) := \sup_{i\in\mathcal{I}} \rho_i(X)$. Then,*

$$\bar{\rho}(X) = \inf\{m \in \mathbb{R} | X + m \in \cap_{i\in\mathcal{I}}\mathcal{A}_i\}$$

Proof. Let $\rho(X) := \inf\{t \in \mathbb{R} : X + t \in \cap_{i\in\mathcal{I}}\mathcal{A}_i\}$. We want to show that $\rho(X) = \bar{\rho}(X)$.

First, since $\{t : X + t \in \cap_{i\in\mathcal{I}}\mathcal{A}_i\} \subseteq \{t : X + t \in \mathcal{A}_i\}$ for all $i \in \mathcal{I}$, we get that $\rho(X) \geq \rho_i(X)$ for all $i \in \mathcal{I}$. Hence, $\rho(X) \geq \bar{\rho}(X)$ for all $X \in \mathscr{L}_p(\Omega, \mathscr{B}, P)$.

To prove the converse inequality, we can first assume that there exists a $t_X \in \mathbb{R}$ for which $\bar{\rho}(X) < t_X$. We have by translation invariance that $\rho_i(X + t_X) \leq \bar{\rho}(X + t_X) < 0$ for all $i \in \mathcal{I}$, hence $X + t_X \in \cap_{i\in\mathcal{I}}\mathcal{A}_i$ so that $0 \geq \rho(X + t_X) = \rho(X) - t_X$. That is, $\rho(X) \leq t_X$. Taking the infimum over all $t_X > \bar{\rho}(X)$, we get that $\rho(X) \leq \bar{\rho}(X)$. The other case is one where $\bar{\rho}(X) = \infty$. For this, it must be that for all $t_X \in \mathbb{R}$ we have that $\bar{\rho}(X) > t_X$, which implies that there exists some $i^* \in \mathcal{I}$ that satisfies $\rho_{i^*}(X) > t_X$. In turns, this indicates that $X + t_X \notin \mathcal{A}_{i^*}$ and thus $X + t_X \notin \cap_{i\in\mathcal{I}}\mathcal{A}_i$. We can therefore say that $\rho(X) > t_X$ which allows us to conclude, since this was true for any t_X, that $\rho(X) = \bar{\rho}(X) = \infty$. □

A1.2.2 Proof of Proposition 1

Based on Lemma A.5, we have that for all $\mathcal{R} \subseteq \mathcal{R}_{ubsr}$,

$$\varrho_{\mathcal{R}}(Z) = \sup_{l\in L_{\mathcal{R}}} \mathrm{SR}_l^P(Z) = \inf\{t : \mathbb{E}_P[l(-Z - t)] - l(0) \leq 0\,,\, \forall\, l \in L_{\mathcal{R}}\} = \mathrm{SR}_{L_{\mathcal{R}}}^P(Z)\,,$$

where $L_{\mathcal{R}} := \{l \in L_{ubsr}\,|\,\mathrm{SR}_l^P \in \mathcal{R}\}$. Hence, to prove our result, we only need to demonstrate that $L_{ubsr} \cap L_{ce}(\mathcal{W}) = L_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})}$ with $L_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})} := \{l \in L_{ubsr}\,|\,\mathrm{SR}_l^P \in \mathcal{R}_{ubsr} \cap \mathcal{R}_{ce}(\mathcal{W})\}$.

We start by showing that $L_{ubsr} \cap L_{ce}(\mathcal{W}) \subseteq L_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})}$. Namely, given any $\bar{l} \in L_{ubsr} \cap L_{ce}(\mathcal{W})$, we show that $\rho_{\bar{l}} := \mathrm{SR}_{\bar{l}}^P \in \mathcal{R}_{ubsr} \cap \mathcal{R}_{ce}(\mathcal{W})$. It is easy to confirm that since $\bar{l} \in L_{ubsr}$, $\rho_{\bar{l}}$ is a legitimate normalized convex utility-based shortfall risk measure. We hence are left with verifying that $\rho_{\bar{l}} \in \mathcal{R}_{ce}(\mathcal{W})$. To do so, we exploit the fact that $\rho_{\bar{l}}(w_k^+) = -w_k^+$ and $\rho_{\bar{l}}(w_k^-) = -w_k^-$, which follows from the fact that ρ is normalized and translation invariant. Namely, by construction, we have that for $k = 1, \ldots, K$

$$\begin{aligned}\rho_{\bar{l}}(W_k) = \mathrm{SR}_{\bar{l}}^P(W_k) &= \inf\{t\,:\,\mathbb{E}_P[\bar{l}(-W_k - t)] \leq \bar{l}(0)\}\\ &= -w_k^- + \inf\{t'\,:\,\mathbb{E}_P[\bar{l}(-W_k + w_k^- - t')] \leq \bar{l}(0)\}\\ &\leq -w_k^- = \rho_{\bar{l}}(w_k^-)\,,\end{aligned}$$

where the last inequality holds because $\bar{l} \in L_{ce}$ so that $\mathbb{E}_P[\bar{l}(-W_k + w_k^-)] \leq \bar{l}(0)$. The latter implies that $t' = 0$ is a feasible solution to the minimization problem at the right hand side of the second equality.

On the other hand, since $\bar{l}(\cdot) \in \mathcal{L}$, it is strictly increasing over the positives, and essinf $W_k \leq w_k^- \leq w_k^+$, then

$$\mathbb{E}_P[\bar{l}(-W_k + w_k^+)] \geq \bar{l}(0) \Longrightarrow \mathbb{E}_P[\bar{l}(-W_k + w_k^+ + \epsilon)] > \bar{l}(0)\,,\, \forall\, \epsilon > 0. \tag{36}$$

To see this, we note that

$$\mathrm{Prob}(-W_k + w_k^+ + \epsilon/2 > 0) \geq \mathrm{Prob}(-W_k + \mathrm{essinf}\, W_k + \epsilon/2 > 0) = \mathrm{Prob}(W_k < \mathrm{essinf}\, W_k + \epsilon/2) > 0\,.$$

Hence, there is a strictly positive probability that the random variable $Y_k := -W_k + w_k^+ + \epsilon/2$ gives a strictly positive value. Moreover, since the loss function is strictly increasing in that region, we must have

$$\mathbb{E}_P[\bar{l}(-W_k + w_k^+ + \epsilon)] > \mathbb{E}_P[\bar{l}(-W_k + w_k^+ + \epsilon/2)] \geq \mathbb{E}_P[\bar{l}(-W_k + w_k^+)] \geq \bar{l}(0)\,.$$

Therefore we have

$$\begin{aligned}\rho_{\bar{l}}(W_k) &= \inf\{t \in \mathbb{R} : \mathbb{E}_P[\bar{l}(-W_k - t)] \leq \bar{l}(0)\}\\ &= -w_k^+ + \inf\{t' \in \mathbb{R} : \mathbb{E}_P[\bar{l}(-W_k + w_k^+ - t')] \leq \bar{l}(0)\}\\ &\geq -w_k^+ = \rho_{\bar{l}}(w_k^+).\end{aligned}$$

This shows that $\rho_{\bar{l}} \in \mathcal{R}_{ubsr} \cap \mathcal{R}_{ce}(\mathcal{W})$.

Next, we show that $L_{ubsr} \cap L_{ce}(\mathcal{W}) \supseteq L_{\mathcal{R}_{ubsr} \cap \mathcal{R}_{ce}(\mathcal{W})}$. In other words, given any $\bar{\rho} \in \mathcal{R}_{ubsr} \cap \mathcal{R}_{ce}(\mathcal{W})$, there exists a $\bar{l} \in L_{ubsr} \cap L_{ce}(\mathcal{W})$ such that $\bar{\rho} = \mathrm{SR}_{\bar{l}}^P$. First, based on Lemma A.3, there is necessarily a $\bar{l} \in L_{ubsr}$ such that such an equality holds. We are left with verifying that such an $\bar{l}$ satisfies

$$\mathbb{E}[\bar{l}(-W_k + w_k^-)] \leq \bar{l}(0) \;\text{ and }\; \mathbb{E}[\bar{l}(-W_k + w_k^+)] \geq \bar{l}(0),\; \text{for } k = 1, \cdots, K\,.$$

Based again on the fact that $\bar{\rho}$ is normalized and translation invariant,

$$\bar{\rho}(W_k) \leq \bar{\rho}(w_k^-) \;\Rightarrow\; \bar{\rho}(W_k) \leq -w_k^-\,.$$

Furthermore, we can exploit Lemma A.4 to show that

$$\mathbb{E}[\bar{l}(-W_k - \bar{\rho}(W_k))] = \bar{l}(0) \;\Rightarrow\; \mathbb{E}[\bar{l}(-W_k + w_k^-)] \leq \bar{l}(0)\,,$$

since $\bar{l}(\cdot)$ is non-decreasing. Likewise, since $\bar{\rho}(W_k) \geq -w_k^+$, we must have again

$$\mathbb{E}[\bar{l}(-W_k - \bar{\rho}(W_k))] = \bar{l}(0) \;\Rightarrow\; \mathbb{E}[\bar{l}(-W_k + w_k^+)] \geq \bar{l}(0)\,.$$

This completes our proof that $\bar{l} \in L_{ubsr} \cap L_{ce}(\mathcal{W})$ and therefore that $L_{ubsr} \cap L_{ce}(\mathcal{W}) \supseteq L_{\mathcal{R}_{ubsr} \cap \mathcal{R}_{ce}(\mathcal{W})}$. □

A1.3 Proof of Proposition 2

Based on Bellini and Bignozzi (2015), it is well known that the class of utility-based shortfall risk measures that are both convex and positive homogeneous, namely $\mathcal{R}_{ubsr} \cap \mathcal{R}_{coh}$, coincides with $\{\mathrm{SR}_l^P \,|\, l \in L_{coh}\}$, where $L_{coh} := \{l \,|\, \exists\, \tau \in [1/2,\; 1),\, l(s) = \max(\tau s,\, (1-\tau)s),\, \forall\, s \in \mathbb{R}\}$. Hence, we have that $L_{\mathcal{R}_{ubsr} \cap \mathcal{R}_{coh}} = L_{coh}$ and by a similar argument as in the proof of Proposition 1 we can also show that $L_{\mathcal{R}_{ubsr} \cap \mathcal{R}_{coh} \cap \mathcal{R}_{ce}(\mathcal{W})} = L_{coh,ce} := L_{ubsr} \cap L_{coh} \cap L_{ce}(\mathcal{W})$.

Given that the loss functions in L_{coh} are parametrized by τ, it is possible to simplify the representation of $L_{coh,ce}$ using the following argument. For any $l \in L_{coh,ce}$ we have that

$$\begin{aligned}\mathbb{E}_P[l(-W_k + w_k^-)] \leq l(0) \;&\Leftrightarrow\; \mathbb{E}_P[\max(\tau(-W_k + w_k^-),\, (1-\tau)(-W_k + w_k^-))] \leq 0\\ &\Leftrightarrow\; \mathbb{E}_P[\tau(-W_k + w_k^-)^+ - (1-\tau)(W_k - w_k^-)^+] \leq 0\\ &\Leftrightarrow\; \tau(\mathbb{E}_P[(-W_k + w_k^-)^+ + (W_k - w_k^-)^+] \leq \mathbb{E}_P[(W_k - w_k^-)^+]\\ &\Leftrightarrow\; \tau \leq b_k \;:=\; \mathbb{E}_P[(W_k - w_k^-)^+]/\mathbb{E}_P[|W_k - w_k^-|].\end{aligned}$$

Likewise

$$\begin{aligned}\mathbb{E}_P[l(-W_k + w_k^+)] \geq l(0) \;&\Leftrightarrow\; \mathbb{E}_P[\max(\tau(-W_k + w_k^+),\, (1-\tau)(-W_k + w_k^+))] \geq 0\\ &\Leftrightarrow\; \mathbb{E}_P[\tau(-W_k + w_k^+)^+ - (1-\tau)(W_k - w_k^+)^+] \geq 0\\ &\Leftrightarrow\; \tau \geq a_k \;:=\; \mathbb{E}_P[(W_k - w_k^+)^+]/\mathbb{E}_P[|W_k - w_k^+|]\,.\end{aligned}$$

Letting $a := \max a_k$ and $b := \min b_k$, two cases can occur. If $a > b$, then one can directly conclude that $L_{coh,ce}$ is empty. Otherwise, $L_{coh,ce}$ can be characterized as

$$L_{coh,ce} := \{l \,|\, \exists\, \tau \in [1/2, 1) \cap [a,\, b]\,,\, l(s) = \max(\tau s,\, (1-\tau)s)\,,\, \forall\, s \in \mathbb{R}\}.$$

It is finally easy to verify that for any $Z \in \mathscr{L}_p$ and any $t \in \mathbb{R}$, we have that

$$\sup_{l\in L_{coh,ce}} \mathbb{E}[l(-Z-t)]-l(0) = \sup_{\tau\in[1/2,1)\cap[a,\,b]} \mathbb{E}[\tau(-Z-t)^+-(1-\tau)(Z+t)^+] = \mathbb{E}_P[b(-Z-t)^+-(1-b)(Z+t)^+],$$

because the function $f(\tau) := \mathbb{E}[\tau(-Z-t)^+ - (1-\tau)(Z+t)^+]$ is increasing in τ, while $b \leq 1$ since $(W_k - w_k^-)^+ \leq |W_k - w_k^-|$. □

A1.4 Proof of Proposition 3

The treatment of $\mathcal{R}_{bnd}(\varepsilon)$ is analogous to the treatment of $\mathcal{R}_{ce}(\mathcal{W})$ considering that for each $M \geq 1$, we are imposing that $\rho(\text{esssup}(Z_M)) = \rho(0) \leq \rho(Z_M^\varepsilon) \leq \rho(-1)$. Hence, following similar arguments as in the proof of Proposition 1, we get that:

$$L_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})\cap\mathcal{R}_{bnd}(\varepsilon)} = L_{ubsr} \cap L_{ce}(\mathcal{W}) \cap \left\{ l \,\middle|\, \begin{array}{c} \varepsilon(z)l(z-1) + (1-\varepsilon(z))l(-1) \leq l(0) \\ \varepsilon(z)l(z) + (1-\varepsilon(z))l(0) \geq l(0) \end{array} \,, \forall\, z \geq 1 \right\}.$$

Yet, one quickly realizes that the second set of constraints in the set defined on the right is redundant since $l(\cdot)$ is non-decreasing. We can therefore conclude that $L_{\mathcal{R}_{ubsr}\cap\mathcal{R}_{ce}(\mathcal{W})\cap\mathcal{R}_{bnd}(\varepsilon)} = L_{ubsr} \cap L_{ce}(\mathcal{W}) \cap L_{bnd}$.

Furthermore, after replacing $z_1 := z - 1$ in the definition of L_{bnd}, we get that all $l \in L_{ubsr} \cap L_{ce}(\mathcal{W}) \cap L_{bnd}$ should satisfy $\varepsilon(z_1+1)l(z_1) + (1-\varepsilon(z_1+1))l(-1) \leq l(0), \forall z_1 \geq 0$, which is equivalent to (13). In what follows, we prove (14) also holds for such l. By exploiting the convexity of l and (13), we obtain

$$l(0) + (l(0) - l(-1))z \leq l(z) \leq l(0) + \frac{1-\varepsilon(z+1)}{\varepsilon(z+1)}(l(0) - l(-1)), \forall z \geq 0. \tag{37}$$

Thus for any $z_2 > z_1 \geq 0$,

$$\begin{aligned}
\frac{l(z_2) - l(z_1)}{z_2 - z_1} &\leq \frac{1}{z_2 - z_1}\left[l(0) + \frac{1-\varepsilon(z_2+1)}{\varepsilon(z_2+1)}(l(0) - l(-1)) - l(z_1)\right] \\
&\leq \frac{1}{z_2 - z_1}\left[\frac{1-\varepsilon(z_2+1)}{\varepsilon(z_2+1)}(l(0) - l(-1)) - (l(0) - l(-1))z_1\right] \\
&= \frac{1}{z_2 - z_1}\left(\frac{1-\varepsilon(z_2+1)}{\varepsilon(z_2+1)} - z_1\right)(l(0) - l(-1)),
\end{aligned}$$

which gives rise to

$$l'_+(z_1) \leq \inf_{z>z_1} \frac{1}{z - z_1}\left(\frac{1-\varepsilon(z+1)}{\varepsilon(z+1)} - z_1\right)(l(0) - l(-1)) = \phi(z_1)(l(0) - l(-1)).$$

This completes our proof. □

Appendix A2 Proofs of Section 3

A2.1 Proof of Proposition 4

We proceed the proof in two steps. First, we identify a linear program which represents the constraint $\sup_{l\in L_{ubsr}\cap L_{ce}(\mathcal{W})} \mathbb{E}_P[l(c(x,\xi) - t) - l(0)] \leq 0$. Second, we derive the dual formulation for this linear program and introduce it in the representation of (PRSR-Opt) problem that takes the form:

$$\begin{array}{ll} \min\limits_{x\in X,t} & t \\ \text{s.t.} & \sup\limits_{l\in L_{ubsr}\cap L_{ce}(\mathcal{W})} \mathbb{E}_P[l(c(x,\xi) - t) - l(0)] \leq 0. \end{array} \tag{38}$$

Observe first that $L_{ubsr} \cap L_{ce}(\mathcal{W})$ is a cone, thus $\sup_{l \in L_{ubsr} \cap L_{ce}(\mathcal{W})} \mathbb{E}_P[l(c(x,\xi)-t)-l(0)] = \infty$ whenever there exists a $l \in L_{ce}(\mathcal{W})$ such that $\mathbb{E}_P[l(c(x,\xi)-t)-l(0)] > 0$. This motivates us to consider an equivalent representation of the inequality constraint. To this end, we consider the set of loss functions:

$$L_{ce}(\mathcal{W})' = \left\{ l : \begin{array}{l} l(y') \geq l(y) + (y'-y)f(y), \,\forall\, (y,y') \in \mathbb{R}\times\mathbb{R}, l(0)=0, l(-1)=-1, f(y)\geq 0\,,\,\forall\, y \in \mathbb{R} \\ \mathbb{E}_P[l(-W_k + w_k^-)] \leq l(0), \mathbb{E}_P[l(-W_k+w_k^+)] \geq l(0)\,,\,\forall\, k \in \{1,2,\ldots,K\} \end{array} \right\}$$

and show that

$$\left\{ t : \sup_{l \in L_{ubsr}\cap L_{ce}(\mathcal{W})} \mathbb{E}_P[l(c(x,\xi)-t)-l(0)] \leq 0 \right\} = \left\{ t : \sup_{l \in L_{ce}(\mathcal{W})'} \mathbb{E}_P[l(c(x,\xi)-t)-l(0)] \leq 0 \right\}. \quad (39)$$

To see this, notice that since $L_{ce}(\mathcal{W})' \subset L_{ubsr} \cap L_{ce}(\mathcal{W})$, then the set at the left-hand side is contained in that of the right-hand side. On the other hand, for any $l \in L_{ubsr}\cap L_{ce}(\mathcal{W})$, if we subtract it by $l(0)$ and scale it by $(l(0)-l(-1))^{-1}$, then we obtain a function $\tilde{l} \in L_{ce}(\mathcal{W})'$ with the following properties:

$$\begin{aligned} \mathbb{E}_P[\tilde{l}(c(x,\xi)-t)-\tilde{l}(0)] \leq 0 \;&\Leftrightarrow\; \mathbb{E}_P\Big[\frac{l(c(x,\xi)-t)-l(0)}{l(0)-l(-1)} - \frac{l(0)-l(0)}{l(0)-l(-1)}\Big] \leq 0 \\ &\Leftrightarrow\; \mathbb{E}_P[l(c(x,\xi)-t)-l(0)] \leq 0, \end{aligned}$$

which means that the right-hand side of equation (39) is a subset of the left-hand side. The relationship (39) allows us to replace $L_{ubsr}\cap L_{ce}(\mathcal{W})$ with $L_{ce}(\mathcal{W})'$ in the constraint of problem (38).

Step 1. We start by giving an infinite dimensional linear programming representation for the right-hand side of (39). First, letting $\Psi(x,t) := \sup_{l\in L_{ce}(\mathcal{W})'} \mathbb{E}_P[l(c(x,\xi)-t)-l(0)]$, we can expand the Ψ operator to

$$\begin{aligned} \Psi(x,t) = \sup_{l:\mathbb{R}\to\mathbb{R}, f:\mathbb{R}\to\mathbb{R}} \quad & \sum_{i=1}^N p_i l(c(x,\xi_i)-t) - l(0) && (40a)\\ \text{s.t.} \quad & l(y') \geq l(y) + (y'-y)f(y)\,,\,\forall\,(y,y') \in \mathbb{R}\times\mathbb{R}, && (40b)\\ & l(0) = 0, && (40c)\\ & l(-1) = -1, && (40d)\\ & f(y) \geq 0\,,\,\forall\, y \in \mathbb{R}, && (40e)\\ & \sum_{\{y: P(-W_k+w_k^- = y)>0\}} P(-W_k + w_k^- = y) l(y) \leq l(0), && (40f)\\ & \sum_{\{y: P(-W_k+w_k^+ = y)>0\}} P(-W_k + w_k^+ = y) l(y) \geq l(0). && (40g) \end{aligned}$$

By exploiting the first and fourth constraints, we may conclude that

$$l(y) = \sup_{v\geq 0, w: vy'+w \leq l(y')\,,\,\forall\, y'\in\mathbb{R}} vy + w.$$

Moreover, since only its value at ξ_i with positive probability affects the objective in the problem above, we may rewrite the optimization problem equivalently as

$$\begin{aligned} \Psi(x,t) = \sup_{v\geq 0, w, l:\mathbb{R}\to\mathbb{R}, f:\mathbb{R}\to\mathbb{R}} \quad & \sum_{i=1}^N p_i[v_i(c(x,\xi_i)-t)+w_i] - l(0) && (41a)\\ \text{s.t.} \quad & v_i y + w_i \leq l(y)\,,\,\forall\, y\in\mathbb{R}, \forall i, && (41b)\\ & (40b) - (40g), \end{aligned}$$

where $v \in \mathbb{R}^N$ and $w \in \mathbb{R}^N$. Furthermore, we can demonstrate that

$$\Psi(x,t) \leq 0 \;\Leftrightarrow\; \tilde{\Psi}(x,t) \leq 0\,,$$

where

$$\tilde{\Psi}(x,t) := \sup_{v\geq 0, w, l:\mathbb{R}\to\mathbb{R}, f:\mathbb{R}\to\mathbb{R}} \quad \sum_i p_i[v_i(c(x,\xi_i)-t)+w_i] - l(0) \tag{42a}$$

$$\text{s.t.} \quad v_i y + w_i \leq l(y)\,,\ \forall\, y\in\mathcal{Y}, \forall i=1,\cdots,N, \tag{42b}$$

$$l(y') \geq l(y) + (y'-y)f(y)\,,\ \forall\,(y,y')\in\mathcal{Y}\times\mathcal{Y}, \tag{42c}$$

$$f(y)\geq 0\,,\ \forall\, y\in\mathcal{Y}, \tag{42d}$$

$$(40c)-(40d),(40f)-(40g)$$

with $\mathcal{Y} := \bigcup_{k=1}^K \operatorname{supp}(-W_k + w_k^-) \cup \operatorname{supp}(-W_k + w_k^+) \cup \{0\} \cup \{-1\}$. A clear benefit of the latter formulation is that it contains a finite number of constraints. We will further show that the decision space can be reduced to finite dimensional so that $l(y)$ and $f(y)$ are only defined on a finite number of points $y\in\mathcal{Y}$.

We start by showing that $\tilde{\Psi}(x,t)\leq 0 \Rightarrow \Psi(x,t)\leq 0$. This follows straightforwardly from the fact that the feasible set of problem (41) is smaller than that of problem (42) and consequently $\Psi(x,t)\leq\tilde{\Psi}(x,t)$.

To see the reverse implication $\Psi(x,t)\leq 0 \Rightarrow \tilde{\Psi}(x,t)\leq 0$, it suffices to show that $\tilde{\Psi}(x,t)>0 \Rightarrow \Psi(x,t)>0$. In other words, if $\tilde{\Psi}(x,t)>0$, there is a loss function, denoted by $\hat{l}$, such that $\hat{l}$ is feasible in problem (41) and it achieves a strictly positive objective value. Let $(\bar{v},\bar{w},\bar{l},\bar{f})$ be any tuplet that defines a feasible solution of problem (42) which achieves a strictly positive objective value. We construct $\hat{l}$ as

$$\hat{l}(y) := \begin{cases} \sup_{v,w:v\geq 0,\, vy'+w\leq \bar{l}(y')\ \forall y'\in\mathcal{Y}} \ vy+w & \text{if } y\leq y_*, \\ \max\{\max_i \bar{v}_i y + \bar{w}_i\,,\ \hat{v}y+\hat{w}\} & \text{otherwise,} \end{cases} \tag{43}$$

where $y_* := \max_{y\in\mathcal{Y}} y$, $\hat{v} := (\bar{l}(y_*)-\bar{l}(y_{**}))/(y_*-y_{**})$ and $\hat{w} := \bar{l}(y_*)$, with $y_{**} := \max_{\{y\in\mathcal{Y}:y<y_*\}} y$. In words, $\hat{l}(y)$ is the convex envelope of the points $\{(y,\bar{l}(y)\}_{y\in\mathcal{Y}}$ in the region where $y\leq y_*$, while outside the region, it is the maximum between the linear extrapolation of this envelope based on the segment between y_{**} and y_*, which ensures continuity at y_*, and the piecewise linear function defined by the supporting planes parameterized by $(\bar{v}_i,\bar{w}_i)$.

Note that the function $\hat{l}$ is convex, non-decreasing, and achieves the same value as $\bar{l}(y)$ when $y\in\mathcal{Y}$. This implies that constraints (40b)–(40g) hold, for some non-negative sub-derivative function $\hat{f}$. We are left with the task to check that the objective value of (40) gives a strictly positive value. In particular,

$$\begin{aligned} \sum_{i=1}^N p_i\hat{l}(c(x,\xi_i)-t)-\hat{l}(0) &= \sum_{i:c(x,\xi_i)-t\leq y^*} p_i\hat{l}(c(x,\xi_i)-t) + \sum_{i:c(x,\xi_i)-t>y^*} p_i\hat{l}(c(x,\xi_i)-t)-\bar{l}(0) \\ &\geq \sum_i p_i \max_{i'}\{\bar{v}_{i'}(c(x,\xi_i)-t)+\bar{w}_{i'}\} - \bar{l}(0) \\ &\geq \sum_i p_i\{\bar{v}_i(c(x,\xi_i)-t)+\bar{w}_i\} - \bar{l}(0)\,, \end{aligned}$$

where we exploit the fact that when $c(x,\xi_i)-t\leq y^*$, by construction

$$\hat{l}(c(x,\xi_i)-t) = \sup_{v,w:v\geq 0,\, vy'+w\leq\bar{l}(y')\ \forall y'\in\mathcal{Y}} v(c(x,\xi_i)-t)+w \geq \max_{i'}\{\bar{v}_{i'}(c(x,\xi_i)-t)+\bar{w}_{i'}\}$$

while when $c(x,\xi_i)-t>y^*$, again by construction $\hat{l}(c(x,\xi_i)-t)\geq \max_{i'}\{\bar{v}_{i'}(c(x,\xi_i)-t)+\bar{w}_{i'}\}$. This allows us to conclude that if $\tilde{\Psi}(x,t)>0$ then $\Psi(x,t)>0$ meaning that $\Psi(x,t)\leq 0\Rightarrow\tilde{\Psi}(x,t)\leq 0$. We complete this first step by arguing that since in problem (42), the decision functions $l:\mathbb{R}\to\mathbb{R}$ and $f:\mathbb{R}\to\mathbb{R}$ are only evaluated at $y\in\mathcal{Y}$, we can reduce the representation of $\tilde{\Psi}(x,t)$ to

$$\begin{aligned} \tilde{\Psi}(x,t) = \sup_{v\geq 0,w,l:\mathcal{Y}\to\mathbb{R},f:\mathcal{Y}\to\mathbb{R}} \quad & \sum_i p_i[v_i(c(x,\xi_i)-t)+w_i]-l(0) \\ \text{s.t.} \quad & (42b)-(42d),(40c)-(40d),(40f)-(40g). \end{aligned}$$

Step 2. Recall that $\mathcal{Y}$ includes the union of the support set of all random variables $-W_k + w_k^-$ and $-W_k + w_k^+$ for $k = 1, \ldots, K$ as well as 0 and -1. For the simplicity of notation, let M denote the size of $\mathcal{Y}$, and y_j the j-th smallest element in $\mathcal{Y}$. Moreover, let j_0 and j_- be the indexes in this list such that $y_{j_0} = 0$ and $y_{j_-} = -1$. Let $\alpha_j := l(y_j)$ and $\beta_j := f(y_j)$ for $j = 1, \ldots, M$. Then $\tilde{\Phi}(x, t)$ can be rewritten as

$$\begin{aligned}
\tilde{\Psi}(x,t) = \sup_{v\geq 0, w, \alpha, \beta} \quad & \sum_{i=1}^{N} p_i [v_i(c(x,\xi_i) - t) + w_i] \\
\text{s.t.} \quad & v_i y_j + w_i \leq \alpha_j \,,\ \forall\, j = 1, \ldots, M, i = 1, \ldots, N, \\
& \alpha_m \geq \alpha_j + (y_m - y_j)\beta_j \,,\ \forall\, m, j = 1, \ldots, M, \\
& \alpha_{j_0} = 0, \\
& \alpha_{j_-} = -1, \\
& \beta_j \geq 0 \,,\ \forall\, j = 1, \ldots, M, \\
& \sum_{j=1}^{M} P(-W_k + w_k^- = y_j)\alpha_j \leq 0, k = 1, \ldots, K, \\
& \sum_{j=1}^{M} P(-W_k + w_k^+ = y_j)\alpha_j \geq 0, k = 1, \ldots, K.
\end{aligned}$$

By introducing the dual variables $u \in \mathbb{R}^{N \times M}$, $\gamma \in \mathbb{R}^{M \times M}$, $\nu_0, \nu_- \in \mathbb{R}$, $\lambda \in \mathbb{R}^M$, and $\eta^{(1)}, \eta^{(2)} \in \mathbb{R}^K$, we obtain that the dual formulation of the problem above takes the form

$$\begin{aligned}
\min_{u,\gamma,\nu_0,\nu_-,\lambda,\eta^{(1)},\eta^{(2)}} \quad & \nu_- \\
\text{s.t.} \quad & p_i(c(x,\xi_i) - t) - \sum_{j=1}^{M} u_{ij} y_j \leq 0, \ \forall\, i = 1, \ldots, N, \\
& p_i - \sum_{j=1}^{M} u_{ij} = 0, \ \forall\, i = 1, \ldots, N, \\
& \sum_{i=1}^{N} u_{ij} + \sum_{m=1}^{M} \gamma_{jm} - \sum_{m=1}^{M} \gamma_{mj} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_j) \\
& \qquad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_j) = 0, \ j = \{1, \ldots, M\} \setminus \{j_0, j_-\}, \\
& \sum_{i=1}^{N} u_{ij_-} + \sum_{m=1}^{M} \gamma_{j_- m} - \sum_{m=1}^{M} \gamma_{m j_-} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_{j_-}) \\
& \qquad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_{j_-}) = -\nu_- \\
& \sum_{i=1}^{N} u_{ij_0} + \sum_{m=1}^{M} \gamma_{j_0 m} - \sum_{m=1}^{M} \gamma_{m j_0} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_{j_0}) \\
& \qquad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_{j_0}) = \nu_0, \\
& \lambda_j - \sum_{m=1}^{M} \gamma_{mj}(y_m - y_j) = 0 \,,\ \forall\, j = 1, \ldots, M, \\
& u_{ij} \geq 0, \gamma_{mj} \geq 0, \lambda \geq 0, \eta^{(1)} \geq 0, \eta^{(2)} \geq 0, \ i = 1, \ldots, N, m, j = 1, \ldots, M.
\end{aligned}$$

Realizing that ν_0 and ν_- both only appear in one of the constraints allows us to simplify the model and reintegrate it in the (PRSR-Opt). The proof is complete. □

A2.2 Proof of Proposition 5

We can apply similar analysis as in the proof of Proposition 4 for $\Phi(x,t) := \sup_{l\in L_{ce}(\mathcal{W})'\cap L_{bnd}} \mathbb{E}_P[l(c(x,\xi)-t)-l(0)]$ which appears in the (PRSR-Opt) problem presented as :

$$\min_{x\in X,t} t \quad \text{s.t.} \quad \Phi(x,t)\leq 0\,,$$

which can be reduced to

$$\min_{x\in X,t\in[t_-,\,t_+]} t \quad \text{s.t.} \quad \Phi(x,t)\leq 0\,, \tag{44}$$

since for any $x\in X$ $\mathrm{SR}_L^P(-c(x,\xi))\leq \mathrm{SR}_L^P(-\max_{x\in X}\max_i c(x,\xi_i)) = t_+$ and $\mathrm{SR}_L^P(-c(x,\xi))\geq \mathrm{SR}_L^P(-\min_{x\in X}\min_i c(x,\xi_i)) = t_-$. Note that we can exploit once more the fact that for all $x\in X$ and $t\in[t_-,\,t_+]$:

$$\left\{t : \sup_{l\in L_{ubsr}\cap L_{ce}(\mathcal{W})\cap L_{bnd}} \mathbb{E}_P[l(c(x,\xi)-t)-l(0)]\leq 0\right\} = \left\{t : \sup_{l\in L_{ce}(\mathcal{W})'\cap L_{bnd}} \mathbb{E}_P[l(c(x,\xi)-t)-l(0)]\leq 0\right\}.$$

When expanding the operator $\Phi(x,t)$, we now obtain

$$\begin{aligned}
\Phi(x,t) = \sup_{l:\mathbb{R}\to\mathbb{R},f:\mathbb{R}\to\mathbb{R}} \quad & \sum_{i=1}^N p_i l(c(x,\xi_i)-t) - l(0) && \text{(45a)}\\
\text{s.t.} \quad & l(y)\leq \varepsilon(y+1)^{-1}-1\,,\ \forall\, y\geq 0, && \text{(45b)}\\
& (40b)-(40g),
\end{aligned}$$

where we were able to simply replace $l(0)=0$ and $l(-1)=-1$ in constraint (12). Using the fact that l is a convex function, we once again replace the objective function to obtain

$$\begin{aligned}
\Phi(x,t) = \sup_{v\geq 0,w,l:\mathbb{R}\to\mathbb{R},f:\mathbb{R}\to\mathbb{R}} \quad & \sum_i p_i[v_i(c(x,\xi_i)-t)+w_i] - l(0) && \text{(46a)}\\
\text{s.t.} \quad & v_i y + w_i \leq l(y)\,,\ \forall\, y\in\mathbb{R}, \forall i, && \text{(46b)}\\
& l(y)\leq \varepsilon(y+1)^{-1}-1\,,\ \forall\, y\geq 0, && \text{(46c)}\\
& (40b)-(40g),
\end{aligned}$$

Moreover, one can establish that the following two constraints can be added to problem (46) without affecting the supremum value:

$$\begin{aligned}
& v_i\leq f(y^*)\,,\ \forall\, i, && \text{(47a)}\\
& l(y^*)+f(y^*)(y-y^*)\leq \varepsilon(y+1)^{-1}-1\,,\ \forall\, y\in\mathcal{Y}'_\varepsilon\cap]y^*,\infty[\,. && \text{(47b)}
\end{aligned}$$

Namely, constraint (47b) is simply redundant given that for all $y\geq 0$

$$l(y^*)+f(y^*)(y-y^*)\leq l(y)\leq \varepsilon(y+1)^{-1}-1\,,$$

based on constraints (40c) and (46c). On the other hand, while constraint (47a) is not redundant, one can show that if there is an i for which $v_i\geq f(y^*)$ then the objective value can be improved by replacing $v_i' := f(y^*)$ and $w_i' := l(y^*)-y^*f(y^*)$. In particular,

$$\begin{aligned}
v_i'(c(x,\xi_i)-t)+w_i' \;&=\; l(y^*)+f(y^*)(c(x,\xi_i)-t-y^*) \;\geq\; v_iy^*+w_i+f(y^*)(c(x,\xi_i)-t-y^*)\\
&\geq\; v_iy^*+w_i+v_i(c(x,\xi_i)-t-y^*) \;=\; v_i(c(x,\xi_i)-t)+w_i\,,
\end{aligned}$$

where we first used the fact that $l(y^*)\geq v_iy^*+w_i$, then the fact that both $v_i\geq f(y^*)$ and $c(x,\xi_i)-t\leq y^*$. Similarly as in the previous proof, we will show that $\Phi(x,t)\leq 0$ if and only if $\tilde{\Phi}(x,t)\leq 0$ with

$$\tilde{\Phi}(x,t) := \sup_{v\geq 0,w,l:\mathcal{Y}'\to\mathbb{R},f:\mathcal{Y}'\to\mathbb{R}} \quad \sum_i p_i[v_i(c(x,\xi_i)-t)+w_i] - l(0) \tag{48a}$$

$$\begin{aligned}
\text{s.t.}\quad & v_i y + w_i \le l(y)\,,\ \forall\, y \in \mathcal{Y}', \forall i, && (48b)\\
& v_i \le f(y^*)\,,\ \forall\, i, && (48c)\\
& l(y') \ge l(y) + (y'-y)f(y)\,,\ \forall\,(y,y') \in \mathcal{Y}' \times \mathcal{Y}', && (48d)\\
& l(y) \le \varepsilon(y+1)^{-1} - 1\,,\ \forall\, y \in \mathcal{Y}' \cap \mathbb{R}^+, && (48e)\\
& l(y^*) + f(y^*)(y-y^*) \le \varepsilon(y+1)^{-1} - 1\,,\ \forall\, y \in \mathcal{Y}'_\varepsilon \cap (y^*, \infty), && (48f)\\
& f(y) \ge 0\,,\ \forall\, y \in \mathcal{Y}', && (48g)\\
& (40c) - (40d), (40f) - (40g)
\end{aligned}$$

where by definition $y^* = \max_{y \in \mathcal{Y}'} y$. While $\tilde{\Phi}(x,t) \le 0 \Rightarrow \Phi(x,t) \le 0$ is again straightforward, the converse requires a slightly modified argument. Indeed, we argue again that if $\tilde{\Phi}(x,t) > 0$ then there must exist an $(\bar{v}, \bar{w}, \bar{l}, \bar{f})$ that satisfy the constraints described in problem (48) and one can therefore construct a loss function $\hat{l}$ according to

$$\hat{l}(y) := \begin{cases} \sup_{v,w:v\ge 0,\, vy'+w\le \bar{l}(y')\ \ \forall y' \in \mathcal{Y}'} \ vy + w & \text{if } y \le y^*, \\ \bar{l}(y^*) + \bar{f}(y^*)(y - y^*) & \text{otherwise}. \end{cases} \quad (49)$$

Given that once again $\hat{l}(y)$ is convex, non-decreasing and achieves the same value as $\bar{l}(y)$ when $y \in \mathcal{Y}'$, it necessarily satisfies constraints (40b)-(40g) for some non-negative sub-derivative function $\hat{f}(y)$. We can also verify as in the proof of Proposition 4 that it returns a strictly positive objective value. Namely,

$$\begin{aligned}
\sum_{i=1}^{N} p_i \hat{l}(c(x,\xi_i) - t) - \hat{l}(0) &= \sum_{i:c(x,\xi_i)-t \le y^*} p_i \hat{l}(c(x,\xi_i) - t) + \sum_{i:c(x,\xi_i)-t > y^*} p_i \hat{l}(c(x,\xi_i) - t) - \bar{l}(0)\\
&\ge \sum_i p_i \max_{i'} \{\bar{v}_{i'}(c(x,\xi_i) - t) + \bar{w}_{i'}\} - \bar{l}(0)\\
&\ge \sum_i p_i \bar{v}_i (c(x,\xi_i) - t) + \bar{w}_i - \bar{l}(0)\,,
\end{aligned}$$

where for all i such that $c(x,\xi_i) - t \le y^*$ once again by construction we have that

$$\hat{l}(c(x,\xi_i) - t) = \sup_{v,w:v\ge 0,\, vy'+w\le \bar{l}(y')\ \ \forall y' \in \mathcal{Y}'} v(c(x,\xi_i) - t) + w \ge \max_{i'} \{\bar{v}_{i'}(c(x,\xi_i) - t) + \bar{w}_{i'}\}$$

while for all i such that $c(x,\xi_i) - t > y^*$ we exploit the following:

$$\hat{l}(c(x,\xi_i) - t) = \bar{l}(y^*) + \bar{f}(y^*)(c(x,\xi_i) - t - y^*) \ge \bar{v}_i y^* + \bar{w}_i + \bar{v}_i(c(x,\xi_i) - t - y^*) = \bar{v}_i(c(x,\xi_i) - t) + \bar{w}_i\,.$$

Before concluding that $\tilde{\Phi}(x,t) > 0 \Rightarrow \Phi(x,t) > 0$, we must confirm that $\hat{l}(y)$ also satisfies constraint (45b). First, in the case that $0 \le y \le y^*$, then either $y \in \mathcal{Y}'$ and (45b) is satisfied since $\hat{l}(y) = \bar{l}(y) \le \varepsilon(y+1)^{-1} - 1$, or by construction

$$\hat{l}(y) = (1-\theta)\bar{l}(y_-) + \theta\bar{l}(y_+) \le (1-\theta)(\varepsilon(y_- + 1)^{-1} - 1) + \theta(\varepsilon(y_+ + 1)^{-1} - 1) = \varepsilon(y+1)^{-1} - 1$$

with $y_- := \sup\{y' \in \mathcal{Y}' : y' < y\}$, $y_+ := \inf\{y' \in \mathcal{Y}' : y' > y\}$, and $\theta := (y - y_-)/(y_+ - y_-) \in]0, 1[$, and where the last equality comes from the fact that $\varepsilon(y+1)^{-1} - 1$ is linear on the interval $[y_-, y_+]$.

Secondly, we should confirm the same fact for $y > y^*$. Indeed, a similar argument can be used here. By construction, we have that

$$\begin{aligned}
\hat{l}(y) &= \bar{l}(y^*) + \bar{f}(y^*)(y - y^*)\\
&= (1-\theta)(\bar{l}(y^*) + \bar{f}(y^*)(y_- - y^*)) + \theta(\bar{l}(y^*) + \bar{f}(y^*)(y_+ - y^*))\\
&\le (1-\theta)(\varepsilon(y_- + 1)^{-1} - 1) + \theta(\varepsilon(y_+ + 1)^{-1} - 1) = \varepsilon(y+1)^{-1} - 1\,,
\end{aligned}$$

with $y_- := \sup\{y' \in \mathcal{Y}'_\varepsilon \cap [y^*, \infty) : y' < y\}$, $y_+ := \inf\{y' \in \mathcal{Y}'_\varepsilon \cap [y^*, \infty) : y' > y\}$, and θ as before. Hence, we can conclude that constraint (45b) is satisfied by $\hat{l}(y)$.

To complete this proof, one simply needs to confirm (18) by strong duality for problem (48). Recall that $\mathcal{Y}'$ includes the union of the set $\mathcal{Y}$ defined in Proposition 4, $\{y_*, y^*\}$ and $\{\mathcal{Y}'_\varepsilon \cap [y_*, y^*]\}$. For the simplicity of notations, let M denote the size of $\mathcal{Y}'$, and y_j the $j-$th smallest element in $\mathcal{Y}'$. Moreover, let j_0 and j_- be the indexes in this list such that $y_{j_0} = 0$, $y_{j_-} = -1$. Then $y_* = y_1$, $y^* = y_M$, $y_1 < y_2 < \cdots < y_{j_-} = -1 < \cdots < y_{j_0} = 0 < \cdots < y_M$, and $\mathcal{Y}' \cap \mathbb{R}^+ = \{y_{j_0+1}, \ldots, y_M\}$. Let $\mathcal{Y}'_\varepsilon \cap (y^*, \infty) := \{y_j\}_{j\in\mathcal{I}}$ with $\mathcal{I} := \{M+1, M+2, \ldots\}$ the index set of the ordered version of this list. Finally, we let $\alpha_j := l(y_j)$ and $\beta_j := f(y_j)$ for $j = 1, \ldots, M$, and consider $z_j := \varepsilon(y_j+1)^{-1}$ for $j = j_0+1, j_0+2, \ldots$. Using this new notation, we obtain that

$$
\begin{aligned}
\tilde{\Phi}(x,t) = \sup_{v\geq 0, w, \alpha, \beta} \quad & \sum_i p_i[v_i(c(x,\xi_i) - t) + w_i] \\
\text{s.t.} \quad & v_i y_j + w_i \leq \alpha_j \,,\ \forall\, i = 1, \ldots, N, j = 1, \ldots, M, \\
& v_i \leq \beta_M \,,\ \forall\, i = 1, \ldots, N, \\
& \alpha_m \geq \alpha_j + (y_m - y_j)\beta_j \,,\ \forall\, m, j = 1, \ldots, M \\
& \alpha_j \leq z_j - 1 \,,\ \forall\, j = j_0+1, \ldots, M, \\
& \alpha_M + \beta_M(y_j - y_M) \leq z_j - 1, \forall\, j \in \mathcal{I}, \\
& \alpha_{j_0} = 0 \\
& \alpha_{j_-} = -1 \\
& \beta_j \geq 0 \,,\ \forall\, j = 1, \ldots, M \\
& \sum_{j=1}^{M} P(-W_k + w_k^- = y_j)\alpha_j \leq 0, k = 1, \ldots, K, \\
& \sum_{j=1}^{M} P(-W_k + w_k^+ = y_j)\alpha_j \geq 0, k = 1, \ldots, K.
\end{aligned}
$$

By introducing the dual variables $u \in \mathbb{R}^{N\times M}, \sigma \in \mathbb{R}^N, \gamma \in \mathbb{R}^{M\times M}$, $\rho \in \mathbb{R}^{M-j_0}$, $\theta \in \mathbb{R}^{(\mathcal{I})}$, $v_0 \in \mathbb{R}$, $v_- \in \mathbb{R}$, $\lambda \in \mathbb{R}^M$, $\eta^{(1)} \in \mathbb{R}^K$ and $\eta^{(2)} \in \mathbb{R}^K$, we can derive the dual formulation of the above linear program:

$$
\begin{aligned}
\min_{u,\sigma,\gamma,\rho,\theta,\nu_0,\nu_-,\lambda,\eta^{(1)},\eta^{(2)}} \quad & \nu_- + \sum_{j\in\mathcal{I}} \theta_{j-M}(z_j - 1) + \sum_{j=j_0+1}^{M} \rho_{j-j_0}(z_j - 1) \\
\text{s.t.} \quad & p_i(c(x,\xi_i) - t) - \sum_{j=1}^{M} u_{ij} y_j - \sigma_i \leq 0,\ \forall\, i = 1, \ldots, N, \\
& p_i - \sum_{j=1}^{M} u_{ij} = 0,\ \forall\, i = 1, \ldots, N, \\
& \sum_{i=1}^{N} u_{ij} + \sum_{m=1}^{M} \gamma_{jm} - \sum_{m=1}^{M} \gamma_{mj} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_j) \\
& \qquad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_j) = 0,\ j = \{1, \ldots, j_0 - 1\} \setminus \{j_-\}, \\
& \sum_{i=1}^{N} u_{ij_-} + \sum_{m=1}^{M} \gamma_{j_- m} - \sum_{m=1}^{M} \gamma_{m j_-} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_{j_-}) \\
& \qquad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_{j_-}) = -\nu_- \\
& \sum_{i=1}^{N} u_{is} + \sum_{m=1}^{M} \gamma_{sm} - \sum_{m=1}^{M} \gamma_{ms} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_{j_0})
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_{j_0}) = \nu_0, \\
&\sum_{i=1}^{N} u_{ij} + \sum_{m=1}^{M} \gamma_{jm} - \sum_{m=1}^{M} \gamma_{mj} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_j) \\
&\quad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_j) = -\rho_{j-j_0}, \forall\, j = j_0+1, \ldots, M-1, \\
&\sum_{i=1}^{N} u_{iM} + \sum_{m=1}^{M} \gamma_{Mm} - \sum_{m=1}^{M} \gamma_{mM} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_M) \\
&\quad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_M) = -\rho_{M-j_0} - \sum_{j\in\mathcal{I}} \theta_{j-M}, \\
&\lambda_j - \sum_{m=1}^{M} \gamma_{mj}(y_m - y_j) = 0\,,\ \forall\, j = 1, \ldots, M-1, \\
&\lambda_M - \sum_{m=1}^{M} \gamma_{mM}(y_m - y_M) = \sum_{j\in\mathcal{I}} \theta_{j-M}(y_j - y_M) - \sum_{i=1}^{N} \sigma_i, \\
&u_{ij} \geq 0, \gamma_{mj} \geq 0, \forall\, i = 1, \ldots, N, m, j = 1, \ldots, M, \\
&\sigma \geq 0, \rho \geq 0, \theta \geq 0, \lambda \geq 0, \eta^{(1)} \geq 0, \eta^{(2)} \geq 0.
\end{aligned}
$$

Realizing that ν_0, ν_-, and λ all only appear in only one of the constraints (besides $\lambda \geq 0$) allows us to simplify the model and reintegrate it in the (PRSR-Opt) problem (44). □

A2.3 Proof of Proposition 6

Letting $\psi(y) := \varepsilon(y+1)^{-1} - 1 - \alpha - \beta(y - y^*)$ and $\bar{\psi} := \bar{\varepsilon}(y+1)^{-1} - 1 - \alpha - \beta(y - y^*)$, we first look at the case where $\bar{y}^*_{\alpha,\beta} \in \mathcal{Y}'_\varepsilon$, then $\lfloor \bar{y}^*_{\alpha,\beta} \rfloor = \lceil \bar{y}^*_{\alpha,\beta} \rceil = \bar{y}^*_{\alpha,\beta}$. Necessarily, this implies that

$$
\psi(\bar{y}^*_{\alpha,\beta}) = \bar{\psi}(\bar{y}^*_{\alpha,\beta}) \leq \min_{y \in \mathcal{Y}'_\varepsilon \cap]y^*,\, \infty[} \bar{\psi}(y) \leq \min_{y \in \mathcal{Y}'_\varepsilon \cap]y^*,\, \infty[} \psi(y)\,.
$$

Second, if $\bar{y}^*_{\alpha,\beta} \notin \mathcal{Y}'_\varepsilon$, $\mathcal{Y}'_\varepsilon \cap (y^*,\, \bar{y}^*_{\alpha,\beta}) \neq \emptyset$, and $\psi(\lfloor \bar{y}^*_{\alpha,\beta} \rfloor) \leq \psi(\lceil \bar{y}^*_{\alpha,\beta} \rceil)$, then since $\psi(y)$ is linear between $\lfloor \bar{y}^*_{\alpha,\beta} \rfloor$ and $\lceil \bar{y}^*_{\alpha,\beta} \rceil$, we must have that for all $y \geq \lfloor \bar{y}^*_{\alpha,\beta} \rfloor$:

$$
\psi(y) \geq \psi(\lfloor \bar{y}^*_{\alpha,\beta} \rfloor) + \frac{\psi(\lceil \bar{y}^*_{\alpha,\beta} \rceil) - \psi(\lfloor \bar{y}^*_{\alpha,\beta} \rfloor)}{\lceil \bar{y}^*_{\alpha,\beta} \rceil - \lfloor \bar{y}^*_{\alpha,\beta} \rfloor}(y - \lfloor \bar{y}^*_{\alpha,\beta} \rfloor) \geq \psi(\lfloor \bar{y}^*_{\alpha,\beta} \rfloor)\,.
$$

On the other hand, for all $y \leq \lfloor \bar{y}^*_{\alpha,\beta} \rfloor$:

$$
\psi(y) \geq \bar{\psi}(y) \geq \bar{\psi}(\lfloor \bar{y}^*_{\alpha,\beta} \rfloor) + \frac{\bar{\psi}(\bar{y}^*_{\alpha,\beta}) - \bar{\psi}(\lfloor \bar{y}^*_{\alpha,\beta} \rfloor)}{\bar{y}^*_{\alpha,\beta} - \lfloor \bar{y}^*_{\alpha,\beta} \rfloor}(y - \lfloor \bar{y}^*_{\alpha,\beta} \rfloor) \geq \bar{\psi}(\lfloor \bar{y}^*_{\alpha,\beta} \rfloor) = \psi(\lfloor \bar{y}^*_{\alpha,\beta} \rfloor)\,.
$$

The third case describes a situation where $\bar{y}^*_{\alpha,\beta} \notin \mathcal{Y}'_\varepsilon$, $\mathcal{Y}'_\varepsilon \cap (y^*,\, \bar{y}^*_{\alpha,\beta}) \neq \emptyset$, and $\psi(\lfloor \bar{y}^*_{\alpha,\beta} \rfloor) \geq \psi(\lceil \bar{y}^*_{\alpha,\beta} \rceil)$ yet the conclusion is entirely analogous to the second case that was just analysed.

Finally, if $\mathcal{Y}'_\varepsilon \cap (y^*,\, \bar{y}^*_{\alpha,\beta}) = \emptyset$, i.e., $\bar{y}^*_{\alpha,\beta} \in (y^*, \lceil \bar{y}^*_{\alpha,\beta} \rceil)$, we have for all $y \geq \lceil \bar{y}^*_{\alpha,\beta} \rceil$:

$$
\psi(y) \geq \bar{\psi}(y) \geq \bar{\psi}(\lceil \bar{y}^*_{\alpha,\beta} \rceil) + \frac{\bar{\psi}(\bar{y}^*_{\alpha,\beta}) - \bar{\psi}(\lceil \bar{y}^*_{\alpha,\beta} \rceil)}{\bar{y}^*_{\alpha,\beta} - \lceil \bar{y}^*_{\alpha,\beta} \rceil}(y - \lceil \bar{y}^*_{\alpha,\beta} \rceil) \geq \bar{\psi}(\lceil \bar{y}^*_{\alpha,\beta} \rceil) = \psi(\lceil \bar{y}^*_{\alpha,\beta} \rceil)\,.
$$

This completes our proof. □

Appendix A3 Proofs of Section 4

A3.1 Proof of Lemma 1

Under Assumption 2 (a), for any $\epsilon > 0$, there exists a compact set $\Xi_\epsilon \subset \Xi$ such that (26) holds. Let $M_\epsilon := \sup_{x\in X, \xi\in\Xi_\epsilon} |g(x,\xi)|$ and $\mathcal{W}_{M_\epsilon}$ be defined as in (29). Then $\mathcal{W}_{M_\epsilon}$ is relatively compact. Let $r \geq 2\sup_{\phi\in\mathcal{W}_{M_\epsilon}} \sup_{z\in[-M_\epsilon, M_\epsilon]} |\phi(z)|$. Then $\Xi_\epsilon \subset \{\xi \in \Xi : |\phi(g(x,\xi))| < r\}$, $\forall x \in X, \phi \in \mathcal{W}$ and hence $\{\xi \in \Xi : |\phi(g(x,\xi))| \geq r\} \subset \Xi\backslash\Xi_\epsilon$, $\forall x \in X, \phi \in \mathcal{W}$. Under condition (26),

$$\sup_{N,x\in X,\phi\in\mathcal{W}} \int_{\{\xi\in\Xi:|\phi(g(x,\xi))|\geq r\}} |\phi(g(x,\xi))|P_N(d\xi) \leq \sup_{N,x\in X,\phi\in\mathcal{W}} \int_{\Xi\backslash\Xi_\epsilon} |\phi(g(x,\xi))|P_N(d\xi) \leq \epsilon.$$

Since P is assumed to be nonatomic, the Lebesgue measure of the set of points where the indicator function $\mathbb{1}_{\Xi\backslash\Xi_\epsilon}(\xi)$ is discontinuous is zero. Together with the above uniform integrability condition, this enables us to claim through (Guo et al. 2017, Lemma 2.1) that for any $x \in X, \phi \in \mathcal{W}$, $\mathbb{E}_P[\phi(g(x,\xi)\mathbb{1}_{\Xi\backslash\Xi_\epsilon}(\xi))] = \lim_{N\to\infty} \mathbb{E}_{P_N}[\phi(g(x,\xi)\mathbb{1}_{\Xi\backslash\Xi_\epsilon}(\xi))]$, together with Assumption 2 (a), this implies $\sup_{x\in X,\phi\in\mathcal{W}} \mathbb{E}_P[|\phi(g(x,\xi)\mathbb{1}_{\Xi\backslash\Xi_\epsilon}(\xi))|] \leq \epsilon$. By the definition of $\Phi_N(x)$ and $\Phi(x)$,

$$\begin{aligned}
|\Phi_N(x) - \Phi(x)| &= |\sup_{\phi\in\mathcal{W}} \mathbb{E}_{P_N}[\phi(g(x,\xi))] - \sup_{\phi\in\mathcal{W}} \mathbb{E}_P[\phi(g(x,\xi))]| \\
&\leq |\sup_{\phi\in\mathcal{W}} \mathbb{E}_{P_N}[\phi(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))] - \sup_{\phi\in\mathcal{W}} \mathbb{E}_P[\phi(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))]| + 2\epsilon \\
&= |\sup_{\phi\in\mathcal{W}_{M_\epsilon}} \mathbb{E}_{P_N}[\phi(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))] - \sup_{\phi\in\mathcal{W}_{M_\epsilon}} \mathbb{E}_P[\phi(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))]| + 2\epsilon \\
&= |\sup_{k\in K} \sup_{\phi\in(\mathcal{W}_{M_\epsilon})_k^\epsilon} \mathbb{E}_{P_N}[\phi(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))] - \sup_{k\in K} \sup_{\phi\in(\mathcal{W}_{M_\epsilon})_k^\epsilon} \mathbb{E}_P[\phi(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))]| + 2\epsilon \\
&\leq \sup_{k\in K} \sup_{\phi\in(\mathcal{W}_{M_\epsilon})_k^\epsilon} |\mathbb{E}_{P_N}[\phi(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi)) - \phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi)) + \phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))] \\
&\quad -\mathbb{E}_P[\phi(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi)) - \phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi)) + \phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))]| + 2\epsilon \\
&\leq 4\epsilon + \sup_{k\in K} |\mathbb{E}_{P_N}[\phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))] - \mathbb{E}_P[\phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))]|.
\end{aligned}$$

For any $\delta > 0$, we may set ϵ sufficiently small such that $\epsilon < \delta/4$. Under Assumption 2 (b) and (c), for any $\phi \in \mathcal{W}_{M_\epsilon}$,

$$\begin{aligned}
|\phi(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi)) - \phi(g(x',\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))| &\leq \kappa_M |g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi) - g(x',\xi)\mathbb{1}_{\Xi_\epsilon}(\xi)| \\
&\leq \kappa_M r(\xi)\mathbb{1}_{\Xi_\epsilon}(\xi)|x - x'|^\nu, \ \forall \xi \in \Xi.
\end{aligned}$$

It follows from (Shapiro and Xu 2008, Theorem 5.1) that for each k there exist positive constants $C(\epsilon,\delta,\phi_k)$ and $\beta(\epsilon,\delta,\phi_k)$ such that

$$\text{Prob}\left(\sup_{x\in X} |\mathbb{E}_{P_N}[\phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))] - \mathbb{E}_P[\phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))]| \geq \delta - 4\epsilon\right) \leq C(\epsilon,\delta,\phi_k)e^{-N\beta(\epsilon,\delta,\phi_k)}.$$

Hence, we have

$$\begin{aligned}
& \text{Prob}\left(\sup_{x\in X} |\Phi_N(x) - \Phi(x)| \geq \delta\right) \\
\leq\ & \text{Prob}\left(\sup_{x\in X} \sup_{k\in K} |\mathbb{E}_{P_N}[\phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))] - \mathbb{E}_P[\phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))]| \geq \delta - 4\epsilon\right) \\
=\ & \text{Prob}\left(\sup_{k\in K} \sup_{x\in X} |\mathbb{E}_{P_N}[\phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))] - \mathbb{E}_P[\phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))]| \geq \delta - 4\epsilon\right) \\
\leq\ & \sum_{k\in K} \text{Prob}\left(\sup_{x\in X} |\mathbb{E}_{P_N}[\phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))] - \mathbb{E}_P[\phi_k(g(x,\xi)\mathbb{1}_{\Xi_\epsilon}(\xi))]| \geq \delta - 4\epsilon\right) \\
\leq\ & \sum_{k\in K} C(\epsilon,\delta,\phi_k)e^{-N\beta(\epsilon,\delta,\phi_k)},
\end{aligned}$$

which implies (32). □

A3.2 Proof of Corollary 1

The result follows from Lemma 1 by setting $M := \sup_{x\in X,\xi\in\Xi} |g(x,\xi)|$ and $\Xi_\epsilon = \Xi$. □

A3.3 Proof of Theorem 1

Observe that $|v_N(x,t) - v(x,t)| \leq \sup_{l\in L} |\mathbb{E}_{P_N}[l(c(x,\xi)-t)] - \mathbb{E}_P[l(c(x,\xi)-t)]|$. Following similar analysis as in Lemma 1, for any $\delta > 0$, there exist positive constants ϵ, $C(\epsilon,\delta)$ and $\beta(\epsilon,\delta)$, independent of N such that

$$\text{Prob}\left(\sup_{x\in X,t\in T}\sup_{l\in L} |\mathbb{E}_{P_N}[l(c(x,\xi)-t)] - \mathbb{E}_P[l(c(x,\xi)-t)]| \geq \delta\right) \leq C(\epsilon,\delta)e^{-N\beta(\epsilon,\delta)}$$

when N is sufficiently large. □

A3.4 Proof of Theorem 2

Part (i). Let $t^* = \vartheta$. Following the discussions immediately after Assumption 1, we know that t^* is finite and $t^* \leq t_0$. Let θ be defined as in Assumption 1 and δ be given as in Theorem 1 with $\delta \leq \theta$ and η be any fixed positive constant such that $\eta \geq \delta$. Then there exists a constant $c_\eta > 0$ such that

$$\inf_{x\in X} v(x, t^* - c_\eta) \geq \eta. \tag{50}$$

To see the existence, notice that

$$\begin{aligned}
\inf_{x\in X} v(x,t^*-c_\eta) &= \inf_{x\in X}\sup_{l\in L} \mathbb{E}_P[l(c(x,\xi)-(t^*-c_\eta))] - l(0)\\
&\geq \inf_{x\in X} \mathbb{E}_P[l_0(c(x,\xi)-(t^*-c_\eta))] - l_0(0) \text{ (for any fixed } l_0\in L)\\
&\geq \inf_{x\in X} l_0(\mathbb{E}_P[c(x,\xi)]-(t^*-c_\eta)) - l_0(0) \text{ (by convexity of } l_0)\\
&= l_0\left(\inf_{x\in X}\mathbb{E}_P[c(x,\xi)]-(t^*-c_\eta)\right) - l_0(0) \text{ (by monotonicity of } l_0).
\end{aligned}$$

Since X is compact and $\mathbb{E}_P[c(x,\xi)]$ is continuous, then $\inf_{x\in X}\mathbb{E}_P[c(x,\xi)]$ is bounded. Moreover, since $\lim_{t\to+\infty} l(t) = +\infty$, the last term goes beyond η for a sufficiently large c_η and hence (50) holds. Let T in Theorem 1 be chosen such that $[t^*-c_\eta, t_0] \subset T$. Then by Theorem 1

$$\begin{aligned}
\inf_{x\in X} v_N(x,t^*-c_\eta) &= \inf_{x\in X} v(x,t^*-c_\eta) + \inf_{x\in X} v_N(x,t^*-c_\eta) - \inf_{x\in X} v(x,t^*-c_\eta)\\
&\geq \eta - \sup_{x\in X}|v_N(x,t^*-c_\eta) - v(x,t^*-c_\eta)| > \eta - \delta/2
\end{aligned}$$

with probability at least $1 - C(\epsilon,\delta/2)e^{-N\beta(\epsilon,\delta/2)}$ for $N \geq N(\epsilon,\delta/2)$. Let $(x_N,t_N) \in S_N$ be the optimal solution of (PRSR-Opt-N). The inequality above shows

$$v_N(x_N, t^*-c_\eta) \geq \inf_{x\in X} v_N(x,t^*-c_\eta) > \eta - \delta/2, \tag{51}$$

which implies $t_N > t^* - c_\eta$ because $v_N(x_N,t_N) \leq 0$ and $v_N(x_N,\cdot)$ is non-increasing.

On the other hand, it follows by (25) and Theorem 1,

$$\sup_{l\in L}\mathbb{E}_{P_N}[l(c(x_0,\xi)-t_0) - l(0)] < -\theta + \delta/2 \leq -\delta/2. \tag{52}$$

with probability at least $1 - C(\epsilon,\delta/2)e^{-N\beta(\epsilon,\delta/2)}$ for $N \geq N(\epsilon,\delta/2)$. The inequality (52) implies (x_0,t_0) is a feasible solution to (PRSR-Opt-N) and hence $t_N \leq t_0$. Summarizing the discussions above, we have $t_N \in [t^*-c_\eta, t_0]$ with probability at least $1 - C(\epsilon,\delta/2)e^{-N\beta(\epsilon,\delta/2)}$.

Let us now consider the systems of inequalities $v(x,t) \le 0, (x,t) \in X \times T$ and $v_N(x,t) \le 0, (x,t) \in X \times T$. Let $\mathcal{F}$ and $\mathcal{F}_N$ be defined as in Section 5.1. Then the set of solutions to the systems of inequalities are equal to $\mathcal{F} \cap (X \times T)$ and $\mathcal{F}_N \cap (X \times T)$ respectively. Since $l(c(x,\xi) - t)$ is convex in (x,t), both $v(x,t)$ and $v_N(x,t)$ are convex functions. By the Slater condition (25), we may use Robinson's error bound theorem for convex systems (Robinson (1975)) to establish that,

$$d((x,t), \mathcal{F} \cap (X \times T)) \le \frac{\Delta}{\delta} \max(v(x,t), 0), \forall (x,t) \in X \times \mathbb{R},$$

where Δ denotes the diameter of $\mathcal{F} \cap X \times T$ and we write $d(a,A)$ for the distance from a point a to a set A. Likewise, we can utilize the Slater condition (52) to obtain

$$d((x,t), \mathcal{F}_N \cap (X \times T)) \le \frac{2\Delta}{\delta} \max(v_N(x,t), 0), \forall (x,t) \in X \times \mathbb{R}$$

with probability at least $1 - C(\epsilon, \delta/2)e^{-N\beta(\epsilon,\delta/2)}$ for $N \ge N(\epsilon, \delta/2)$. Combining the two error bounds, we effectively obtain

$$\mathbb{H}(\mathcal{F}_N \cap (X \times T), \mathcal{F} \cap (X \times T)) \le \frac{2\Delta}{\delta} \sup_{x \in X, t \in T} |v_N(x,t) - v(x,t)|,$$

where $\mathbb{H}$ denotes the Hausdorff distance. Thus

$$|\vartheta_N - \vartheta| = |t_N - t^*| \le \mathbb{H}(\mathcal{F}_N \cap (X \times T), \mathcal{F} \cap (X \times T)) \le \frac{2\Delta}{\delta} \sup_{x \in X, t \in T} |v_N(x,t) - v(x,t)|. \quad (53)$$

Let $\varepsilon := \min(\frac{\delta^2}{2\Delta}, \frac{\delta}{2})$. We deduce from (33) and (53) that for $N \ge N(\epsilon, \varepsilon)$

$$\text{Prob}(|\vartheta_N - \vartheta| \ge \delta) \le \text{Prob}\left(\sup_{x \in X, t \in T} |v_N(x,t) - v(x,t)| \ge \varepsilon\right) \le C(\epsilon, \varepsilon)e^{-N\beta(\epsilon,\varepsilon)}.$$

Part (ii). The exponential rate of convergence (34) implies $t_N \to t^*$ almost surely. Moreover, since $v_N(x_N, t_N) \le 0$ and v_N converges uniformly to v over $X \times T$, then $v(\hat{x}, t^*) \le 0$ for every cluster point $\hat{x}$ of $\{x_N\}$. □

References

Armbruster B, Delage E (2015), Decision making under uncertainty when preference information is incomplete, Management Science, 61(1): 111–128.

Artzner, P., Delbaen, F. and Eber, J.-M., Heath, D (1999), Coherent Measures of Risk, Mathematical Finance, 9(3):203–228.

Billingsley P (1999), Convergence of Probability Measures, Wiley, New York.

Bellini F, Bignozzi V (2015), On elicitable risk measures, Quantitative Finance, 15(5): 725–733.

Ben-Tal A and Teboulle M, An old-new concept of convex risk measures: the optimized certainty equivalent, Mathematical Finance, 17(3): 449–476.

Brown R F (2004), A Topological Introduction to Nonlinear Analysis, Spring, New York.

Buhlmann H (1970), Mathematical Methods in Risk Theory, Berlin: Springer.

Cont R, Deguest R, Scandolo G (2010), Robustness and sensitivity analysis of risk measurement procedures, Quantitative Finance, 10(6): 593–606.

Delage E, Li J. Y. (2018), Minimizing risk exposure when the choice of a risk measure is ambiguous, Management Science, 64(1): 327–344.

Desrosiers J, Lübbecke M C (2005), A Primer in Column Generation, Column Generation, G. Desaulniers, J. Desrosiers, and M.M. Solomon (Eds.), Springer, 1–32.

Föllmer H, Schied A. (2002), Convex measures of risk and trading constraints, Finance and Stochastic, 6(4): 429–447.

Föllmer H. and Schied A. (2016), Stochastic Finance: An Introduction in Discrete Time, 4th ed, Berlin: de Gruyter.

Giesecke K, Schmidt T, Weber S (2008), Measuring the risk of large losses, Journal of Investment Management, 6(4): 1–15.

Guo S, Xu H (2019), Distributionally Robust Shortfall Risk Optimization Model and Its Approximation, Mathematical Programming, 174:473–498.

Guo S, Xu H, Zhang L (2017), Convergence analysis for mathematical programs with distributionally robust chance constraint, SIAM Journal on optimization, 27(2): 784–816.

Haskell B, Fu L, Dessouky M (2016), Ambiguity in risk preferences in robust stochastic optimization, European Journal of Operational Research, 254(1): 214–225.

Haskell W, Shanthikumar J, Shen Z (2017), Aspects of optimization with stochastic dominance, Annals of Operations Research, 253(1): 247–273.

Hu J, Mehrotra S (2015), Robust decision making over a set of random targets or risk-averse utilities with an application to portfolio optimization, IIE Transactions, 47(4): 358–372.

Robinson S M (1975), An application of error bounds for convex programming in a linear space, SIAM Journal on Control, 13(2): 271–273.

Rockafellar R T and Uryasev S (2013), The fundamental risk quadangle in risk management, optimization and statistical estimation, Surveys in Operations Research and Management Science, 18: 33–53.

Rusczyński A, Shapiro A (2003), Stochastic Programming, in: Handbooks Operations Research and Management Science, volume 10, North-Holland Publishing Company, Amsterdam.

Shapiro A, Xu H (2008), Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation, Optimization, 57(3): 395–418.

Weber S (2006), Distribution-invariant risk measures, information, and dynamic consistency, Mathematical Finance, 16(2): 419–441.