Shortfall Risk Models When Information of Loss Function Is Incomplete

Erick Delage[*], Shaoyan Guo[†] and Huifu Xu[‡]

April 25, 2018

Abstract. Utility-based shortfall risk measure (SR) has received increasing attentions over the past few years for its potential to quantify more effectively the risk of large losses than conditional value at risk. In this paper we consider the case that the true loss function is unavailable either because it is difficult to be identified or the decision maker is ambiguous about disutility of the losses. We propose a preference robust SR (PRSR) model where it is possible to construct a set of utility-based loss functions from empirical data or subjective judgements and define PRSR through the worst loss function from the set in order to mitigate the effect from the ambiguity. We then demonstrate that the worst of a set of convex utility-based shortfall risk measure with some specified characteristics such as positive homogeneity and pairwise comparisons can be represented as PRSR by choosing an appropriate class of utility loss functions. The representation enables us to develop tractable formulations for optimization problem with the objective of minimizing the worst convex utility-based shortfall risk measures when the underlying probability distribution is discrete. 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.

Key words. Utility-based shortfall risk measure, preference robust optimization, loss function, tractability, sample average approximation, inverse optimization, portfolio management.

[*]Department of Decision Sciences, HEC Montréal, Montréal, Québec, H3T 2A7, Canada.(erick.delage@hec.ca).
[†]School of Mathematical Sciences, Dalian University of technology, Dalian, 116024, China.(syguo@dlut.edu.cn).
[‡]School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK. (h.xu@soton.ac.uk).

1 Introduction

Quantitative measure of risk is a key element in risk management for financial institutions and regulatory authorities. Consider a financial position represented by a random variable $Z : \Omega \to \mathbb{R}$ on a probability space $(\Omega, \mathscr{B}, P)$, where $(\Omega, \mathscr{B})$ is a measurable space with sigma algebra $\mathscr{B}$ and P is a probability measure. We consider a risk measure to be a real valued function from the space of random variables having finite p-th order moments $\rho : \mathscr{L}_p(\Omega, \mathscr{B}, P) \to \mathbb{R}$, with $p \in [1, \infty)$, which signifies the risk of the position. Based on the discussion in [10], a "good" risk measure should be sensitive to excessive losses and penalize concentration while encouraging diversification.

A well-known risk measure which has been widely used by banks under Basel II regularization is value at risk (VaR). It is defined as

$$\mathrm{VaR}_\beta(Z) := \inf\{t \in \mathbb{R} : P(-Z - t \geq 0) \leq 1 - \beta\},$$

which can be interpreted as the smallest amount of cash that needs to be added to Z such that the probability of the financial position falling into a loss does not exceed a specified probability $1 - \beta$. Unfortunately VaR is not sensitive to tail losses nor is it encouraging diversification. Various other measures are subsequently proposed to overcome the disadvantages of VaR such as conditional value at risk (CVaR) and entropic risk measure (see [10]).

In this paper, we consider the so-called "utility-based shortfall risk measure (SR)" defined as

$$\mathrm{SR}_l^P(Z) := \inf_t\{t : \mathbb{E}[l(-Z - t)] \leq \lambda\}, \tag{1.1}$$

where $l : \mathbb{R} \to \mathbb{R}$ is a non-decreasing and non-constant *loss* function and λ lies in the interior of the range of $l(\cdot)$. Its construction is intimately connected to Von Neumann-Morgenstern expected utility theory, which models an investor's risk attitude, when one considers the loss function as a disutility function for losses $l(y) := -u(-y)$ so that

$$\mathrm{SR}_l^P(Z) := \inf\{t : \mathbb{E}[u(Z + t)] \geq -\lambda\}.$$

This connection provides an intuition that $\mathrm{SR}_l^P(Z)$ measures the smallest amount of cash that must be injected to the position Z in order for the expected disutility of losses to fall below a specified target λ. In this work, we focus on the class of normalized (i.e. no cash should be injected when $Z = 0$) convex utility-based shortfall risk measures in which case one can easily show that the the target $\lambda = l(0)$ while the loss function $l(\cdot)$ must be convex and strictly increasing over some $[z_0, \infty)$ with $z_0 < 0$, see Remark 4 in [3] and further details in Section 2. Under these conditions, the utility-based shortfall risk measure reduces to

$$\mathrm{SR}_l^P(Z) := \inf_t\{t : \mathbb{E}[l(-Z - t)] - l(0) \leq 0\}. \tag{1.2}$$

The notion of utility-based shortfall risk measure was introduced by Föllmer and Schied [8] and has been studied by many others, see for example in [7, 10, 15] and references therein. It is well-known and easy to verify that $\mathrm{SR}_l^P(\cdot)$ generally satisfies monotonicity, translation invariance and law invariance, while it moreover satisfies convexity if and only if $l(\cdot)$ is convex ([8]). Additionally, one can show that this class of risk measures satisfies property such as invariance under randomization which other risk measure might not enjoy and when l is convex, it becomes the only law-invariant convex risk measure that can be used for the dynamic measurement of risk over time (see [21] and [10]). Compared to the well-known VaR and CVaR, SR is more sensitive to large losses when Z follows extreme events in the context of the heavy tailed distributions ([8, 10]). Moreover, SR 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 ([3]).

From the definition in (1.2) of SR, it is easy to observe that the loss function l plays an important role. In practice, there could be considerable ambiguity in the choice of l. For example, the decision maker hesitated about the type of loss function that best characterizes his preference, or a group of decision makers might have difficulty agreeing about which loss function to use. In these circumstances, there is no obvious choice for the "true" loss function. To overcome the risk arising from mis-specification of the decision maker's preferences regarding losses, we might exploit the idea of Armbruster and Delage [1] to construct a set of loss functions with available partial information or subjective judgements and compute the SR on the basis of the worst loss function in this set. Specifically let L be a set of the loss functions which satisfy certain conditions (to be detailed in later discussions). We define the so-called preference robust SR as follows:

$$\text{(PRSR)} \quad \mathrm{SR}_L^P(Z) := \inf \left\{ t : \sup_{l \in L} \mathbb{E}_P[l(-Z-t)] - l(0) \leq 0 \right\}, \tag{1.3}$$

where the preference robust constraint means that the expected disutility of losses $\mathbb{E}_P[l(-Z-t)]$ falls below $l(0)$ for every plausible risk attitude defined as $l \in L$.

A key element in the above model is the construction of the set L. Armbruster and Delage [1] 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 in order to obtain smaller sets of function than those typically obtained from first-order or second-order stochastic dominance constraints. 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. [12] extended this preference robust expected utility framework to cases where there is also ambiguity about the choice of a probability function. More recently, the framework was also applied by Delage and Li [6] to convex risk measures where it gives rise to the notion of a preference robust risk measure:

$$\varrho_{\mathcal{R}}(Z) := \sup_{\rho \in \mathcal{R}} \rho(Z), \tag{1.4}$$

where $\mathcal{R}$ is an ambiguous set of all convex/coherent/law invariant risk measures that satisfies a list of pairwise comparisons. However, the authors did not provide a reformulation that could exploit the fact that the measure is known to be a utility-based shortfall risk measures. Another important limitation of all methods above consist in the fact that tractable reformulations are only established under the hypothesis of a discrete outcome space. It remains to be explored as to whether there is a tractable numerical scheme for the case when the underlying random variable is continuously distributed.

In this paper, we first establish that $\mathrm{SR}_L^P(Z)$ and $\varrho_{\mathcal{R}}(Z)$ are equivalent representations when $\mathcal{R}$ only contains normalized and convex utility-based shortfall risk measures. We then describe how L can be constructed based on information about whether the risk measure is coherent or not and about a list of confidence regions for the risk of a set of random variables.

In practice, a risk measure is often associated with some decision making problems whereby the risk may be considered as the objective to minimize or as a constraint. Here, we consider an optimization problem with preference robust SR as an objective function. Specifically, let $c(x,\xi(\omega))$ a financial loss associated with decision vector $x \in X$ and random vector $\xi(\omega)$. We consider

$$\text{(PRSRP)} \quad \min_{x\in X} \mathrm{SR}_L^P(-c(x,\xi)) \quad \equiv \quad \min_{x\in X} \varrho_{\mathcal{R}}(-c(x,\xi)) \tag{1.5}$$

where $c(x,\xi): \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$ is a continuous function and ξ is a random vector mapping from probability space $(\Omega, \mathscr{B}, P)$ to $\Xi \subseteq \mathbb{R}^s$ and where $\mathcal{R} = \{\mathrm{SR}_l^P\}_{l\in L}$. In the case when $c(x,\xi)$ is a piecewise linear function, P is discrete and L is a particular class of loss functions, we show that (PRSRP) can be reformulated as a linear programming problem of finite dimension or of semi-infinite dimension for which we design an efficient column-generation method.

When P is a continuous probability measure, we propose a sample average approximation scheme

$$\text{(PRSRP-N)} \quad \min_{x\in X} \mathrm{SR}_L^{P_N}(-c(x,\xi)) \tag{1.6}$$

and establish conditions under which (PRSRP-N) converges to (PRSRP) in terms of the optimal value and optimal solutions as the sample size increases.

The rest of the paper is organized as follows. In Section 2, we present a set of useful properties satisfied by normalized utility-based shortfall risk measures and demonstrate that SR_L^P coincides with the worst risk measured by SR_l^P over $l \in L$ hence that it is equivalent to the convex risk measure $\varrho_{\mathcal{R}}$. We also construct an example showing that SR_L^P does not generally coincide with a particular utility-based shortfall risk measure SR_l^P. In Section 3, we exploit the established properties of SR_L^P to characterize the ambiguity set of loss functions L that is obtained when one wishes to account information about positive homogeneity of the risk measure, about some "confidence" intervals for the risk of a list of random variables, and finally about how sensitive the risk measure is with respect to large tail losses. In Section 4, we present tractable reformulations for (PRSRP) with the characterizations of L derived in Section 3 when the underlying random variable is finitely distributed. Finally, in Section 5, we propose a discretization scheme for (PRSRP) when the underlying random variables are continuously distributed and quantify the difference between the true (PRSRP) and its discretized counterpart in terms of the optimal value. In Section 6, we report numerical experiments which highlight the added value of the proposed (PRSRP) models. We finally conclude in Section 7.

2 Preference robust normalized utility-based shortfall risk measures

We start by specifying the set of loss functions which we will use throughout the paper to define our risk measures and describe a set of four useful properties satisfied by the functions in this set (see Appendix A.1 for a detailed proof).

Definition 2.1 *Let $\mathcal{L}$ 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 $z_0 < 0$.*

Lemma 2.1 (Properties of $l \in \mathcal{L}$) *The following assertions hold for all $l \in \mathcal{L}$:*

(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 $l(\mathbb{R})$.

Equipped with Lemma 2.1, we can now demonstrate that a normalized convex utility-based shortfall risk measure has a representation only with $l \in \mathcal{L}$ and $\lambda = l(0)$.

Proposition 2.1 (Representation of normalized utility-based shortfall risk measure) *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 $l \in \mathcal{L}$ such that*

$$\rho(Z) := \inf\{t : \mathbb{E}[l(-Z - t)] - l(0) \leq 0\}. \tag{2.7}$$

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 (1.1). By [21, 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 2.1 (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 $l \in \mathcal{L}$ 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 [21, Corollary 3.1]. Moreover, we also have that $\lambda := l(0)$ lies in the interior of $l(\mathbb{R})$ based on Lemma 2.1(iv). One can finally verify that $\rho(0) = \inf\{t : \mathbb{E}[l(-t)] - l(0) \leq 0\} = 0$ following Lemma 2.1(iii). ■

It is already clear from Proposition 2.1 that a preference robust risk measure $\varrho_{\mathcal{R}}$ which considers only normalized convex utility-based risk measures can be represented using

$$\varrho_{\mathcal{R}}(Z) = \sup_{\rho \in \mathcal{R}} \rho(Z) = \sup_{l \in \mathcal{L}_{\mathcal{R}}} \mathrm{SR}_l^P(Z),$$

where $L_{\mathcal{R}} := \{l \in \mathcal{L} \,|\, \mathrm{SR}_l^P(\cdot) \in \mathcal{R}\}$. Furthermore, this relationship is two-sided since for any set of loss functions $L \subset \mathcal{L}$,

$$\sup_{l \in L} \mathrm{SR}_l^P(Z) = \sup_{\rho \in \mathcal{R}_L} \rho(Z) = \varrho_{\mathcal{R}_L}(Z),$$

where $\mathcal{R}_L := \{\mathrm{SR}_l^P\}_{l\in L}$. The question however remains as to how to establish the connection between $\varrho_{\mathcal{R}}(Z)$ and PRSR in the form of $\mathrm{SR}_L^P(Z)$. Our success in this endeavour will be based on a set of well-known properties of normalized convex risk measures, presented in Lemma 2.2, and a property which is unique to utility-based shortfall risk measures, in Lemma 2.3. The proofs of both lemmas are deferred to appendices A.2 and A.3 given that they are either relatively easy to derive or have already appeared in the literature in one way or another.

Lemma 2.2 *The following properties are satisfied by any normalized utility-based shortfall risk measure ρ.*

(i) Given any constant $c \in \mathbb{R}$, $\rho(c) = -c$.

(ii) Given any random variable $Z \in \mathscr{L}_p$, one has that $-\mathrm{esssup}(Z) \le \rho(Z) \le -\mathrm{essinf}(Z)$.

Lemma 2.3 *Let $\rho(\cdot)$ be a normalized convex utility-based shortfall risk measure. Then $\rho(Z)$ is the unique solution of the equation*

$$\mathbb{E}_P[l(-Z-t)] = l(0). \tag{2.8}$$

Note that Lemma 2.3 does not hold if $\rho(\cdot)$ is not guaranteed to be finite valued. For instance, if one allows $\lambda = l(0)$ to be on the boundary of the $l(\mathbb{R})$, then $l(\cdot)$ necessarily becomes constant on all of the negatives. In that case, $\mathrm{SR}_l^P(Z)$ coincides with the "worst-case" loss measure $\rho(Z) := \mathrm{ess\,sup}(-Z)$ which could be infinite for some $Z \in \mathscr{L}_p(\Omega, \mathscr{B}, P)$ when $p < \infty$. It is clear in such cases that $\rho(Z) = \infty$ does not solve equation (2.8).

We are now ready to present the main theorem of this section, namely how $\varrho_{\mathcal{R}}(Z)$ and $\mathrm{SR}_L^P(Z)$ are equivalent representations when $\mathcal{R}$ contains only normalized convex utility-based shortfall risk measures.

Theorem 2.1 *Let $L \subseteq \mathcal{L}$. Then $\mathrm{SR}_L^P(Z) = \sup_{l\in L} \mathrm{SR}_l^P(Z)$. Hence, for any $L \subseteq \mathcal{L}$, $\mathrm{SR}_L^P(Z) = \varrho_{\mathcal{R}_L}(Z)$ with $\mathcal{R}_L := \{\mathrm{SR}_l^P\}_{l\in L}$ and conversely, for any set of normalized convex utility-based shortfall risk measures $\mathcal{R}$, $\varrho_{\mathcal{R}}(Z) = \mathrm{SR}_{L_{\mathcal{R}}}^P(Z)$ with $L_{\mathcal{R}} := \{l \in \mathcal{L} \,|\, \mathrm{SR}_l^P \in \mathcal{R}\}$.*

Proof. This proof relies on demonstrating that $\mathrm{SR}_L^P(Z) = \sup_{l\in L} \mathrm{SR}_l^P(Z)$ or equivalently

$$\inf\left\{t : \sup_{l\in L} \mathbb{E}_P[l(-Z-t)] - l(0) \le 0\right\} = \sup_{l\in L} \inf\left\{t : \mathbb{E}_P[l(-Z-t)] - l(0) \le 0\right\}. \tag{2.9}$$

Let t^* and $\hat{t}$ be the infimum and supremum achieved on the left and right-hand side of the equation above respectively, namely $t^* := \mathrm{SR}_L^P(Z)$ and $\hat{t} := \sup_{l\in L} \mathrm{SR}_l^P(Z)$. We first consider the case (a) that

$$\exists\, t_0 \in \mathbb{R}, \qquad \sup_{l\in L} \mathbb{E}_P[l(-Z-t_0)] - l(0) \le 0. \tag{2.10}$$

Under this condition, $\mathrm{SR}_L^P(Z)$ is finite. Indeed, condition (2.10) ensures that the feasible set of the minimization problem at the left hand side of (2.9) is non-empty and hence $t^* \ne +\infty$. In what follows, we show that $t^* \ne -\infty$.

Since $l(\cdot)$ is convex, by Jensen's inequality,

$$\mathbb{E}_P[l(-Z-t)] \geq l(-\mathbb{E}_P[Z]-t),$$

which enables us to deduce

$$\sup_{l\in L} \mathbb{E}_P[l(-Z-t)] - l(0) \geq \sup_{l\in L} l(-\mathbb{E}_P[Z]-t) - l(0). \tag{2.11}$$

Based on Lemma 2.1(ii), we have that $\lim_{z\to+\infty} l(z) = +\infty$, hence

$$\lim_{t\to-\infty} \sup_{l\in L} l(-\mathbb{E}_P[Z]-t) - l(0) = +\infty,$$

and through (2.11), we have

$$\lim_{t\to-\infty} \sup_{l\in L} \mathbb{E}_P[l(-Z-t)] - l(0) = +\infty > 0. \tag{2.12}$$

From this and the fact that $t^* \neq +\infty$, we can draw a conclusion that

$$\{t \in \mathbb{R} | \sup_{l\in L} \mathbb{E}_P[l(-Z-t)] - l(0) \leq 0\}$$

is non-empty and bounded below hence t^* is finite.

Now we return to show (2.9). Since for any $\epsilon > 0$, we must have that

$$\sup_{l\in L} \mathbb{E}_P[l(-Z-t^*-\epsilon)] - l(0) \leq 0,$$

then for any $l \in L$, $\mathbb{E}_P[l(-Z-t^*-\epsilon)] - l(0) \leq 0$, which means that for each $l \in L$, $t^* + \epsilon$ is a feasible solution for the inner minimization problem of the right hand side of (2.9). If we use t_l to denote the optimal value of the inner minimization problem of the right hand side of (2.9), then our discussion shows $t^* + \epsilon \geq t_l$ for all $l \in L$ and hence

$$t^* + \epsilon \geq \sup_{l\in L} t_l = \hat{t}\,, \forall \epsilon > 0 \;\Rightarrow\; t^* \geq \hat{t}, \tag{2.13}$$

which further implies that $\hat{t}$ is also finite.

Conversely, for each $l \in L$,

$$\hat{t} = \sup_{l\in L} t_l \;\Rightarrow\; \mathbb{E}_P[l(-Z-\hat{t})] - l(0) \leq \mathbb{E}_P[l(-Z-t_l)] - l(0) \leq 0\,, \forall l \in L,$$

giving $\sup_{l\in L} \mathbb{E}_P[l(-Z-\hat{t})] - l(0) \leq 0$. This shows $\hat{t}$ is a feasible solution of the left hand side of (2.9) and hence $\hat{t} \geq t^*$.

We now turn to the case (b) that condition (2.10) fails to hold, i.e., the feasible set is empty. In that case, the optimal value $t^* = +\infty$. Assume for the sake of a contradiction that $\hat{t} < +\infty$. Then $t_l \leq \hat{t} < +\infty$ for all $l \in L$ and hence for any positive number ϵ,

$$\mathbb{E}_P[l(-Z-(\hat{t}+\epsilon)) - l(0)] \leq 0, \forall l \in L.$$

The latter implies $\hat{t} + \epsilon$ is a feasible solution to the minimization problem at the left hand side of equation (2.9), a contradiction. Hence, without condition (2.10), both sides of (2.9) must be infinite and hence the equality still holds.

The rest of our claim follows straightforwardly from the representation established in Proposition 2.1. Indeed, for any $L \subseteq \mathcal{L}$, we just showed that $\mathrm{SR}_L^P(Z) = \sup_{l\in L} \mathrm{SR}_l^P(Z) = \sup_{\rho\in\mathcal{R}_L}(Z)$, while according to Proposition 2.1, $\varrho_{\mathcal{R}}(Z) = \sup_{l\in L_{\mathcal{R}}} \mathrm{SR}_l^P(Z) = \mathrm{SR}_{L_{\mathcal{R}}}^P(Z)$. ■

In practice, it might be useful to employ a preference robust risk measure in a constraint rather than in the objective. In such a circumstance, one can easily establish that the preference robust risk of a financial position Z is guaranteed to fall below a target γ if and only if the expected loss of the investment, after injecting an amount of cash equal to γ, falls below the loss of a null position for every loss function in L, see the corollary below. We refer readers to Appendix A.4 for a detailed proof of this equivalence.

Corollary 2.1 *Let $L \subseteq \mathcal{L}$. Then the following two inequalities are equivalent:*

$$\mathrm{SR}_L^P(Z) \leq \gamma \iff \sup_{l\in L} \mathbb{E}_P[l(-Z-\gamma)] - l(0) \leq 0. \tag{2.14}$$

Similarly, $\rho_{\mathcal{R}}(Z) \leq \gamma$ if and only if $\sup_{l\in L_{\mathcal{R}}} \mathbb{E}_P[l(-Z-\gamma)] - l(0) \leq 0$.

Finally, since Theorem 2.1 establishes that $\mathrm{SR}_L^P(Z)$ is a preference robust utility-based shortfall risk measure, it therefore follows that $\mathrm{SR}_L^P(Z)$ inherits all properties of convex risk measures such as monotonicity, translation invariance and convexity (see Proposition 1 in [6]), and it is necessarily normalized. One might however wonder whether it also satisfies all other properties of normalized convex utility-based shortfall risk measure. The following proposition states that it is not the case in general.

Proposition 2.2 *Let $L \subseteq \mathcal{L}$. Then $\mathrm{SR}_L^P(Z)$ is a law-invariant convex risk measure. It does not however in general coincide with a specific convex utility-based shortfall risk measure $\mathrm{SR}_l^P(Z)$.*

This result is interesting for two reasons. First, it states that in general we might need to accept to lose properties such as elicitability when making risk measurements robust amid ambiguity of the true loss function. Second, it also means that we shouldn't expect in general that the minimization of SR_L^P can be decomposed into two sub-problems: (a) identifying a universal worst-case loss function $\hat{l}$ in the first place and then (b) subsequently minimizing the risk of $\mathrm{SR}_{\hat{l}}^P$. Rather, we can expect that each random variable would suffer most from different loss functions in the set L. Exceptionally, we will see in our later discussion that in the special case when the risk measure ρ is known to be positive homogeneous, then a universal worst-case loss function $\hat{l}(\cdot)$ does arise (unless the ambiguity set for L is actually empty) thus making $\mathrm{SR}_L^P(Z)$ conveniently reduce to a specific utility-based shortfall risk measure $\mathrm{SR}_{\hat{l}}^P$.

3 Characterization of L using risk preference information

In this section, we exploit elicited preference information[1] (as proposed in [6]) to characterize the set of loss functions L based on information about the decision maker's preferences regarding

[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 [5] for a discussion).

risk. 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 3.1 *Let $\mathcal{R}$ be the set of all normalized convex utility-based shortfall risk measures, and let $\mathcal{R}_{coh} \subset \mathcal{R}$ be the set of such risk measures that are positive homogenous, i.e.,*

$$\rho(\alpha Z) = \alpha\rho(Z)\,,\ \forall\, Z \in \mathscr{L}_p\,,\ \forall\, \alpha > 0\,.$$

As mentioned in [6], identifying whether the utility-based shortfall risk measure that should be used is a member of $\mathcal{R}_{coh}$ or not reduces to establishing whether the ordinal comparison of riskiness of two random variables is affected by a uniform positive scaling of both random variables.

In order to refine the characterization of $\mathcal{R}$, and implicitly L, 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 3.2 *Let $\{W_k\}_{k=1}^K$ be a list of random variables with an associated set of "confidence" intervals $[w_k^-,\, w_k^+] \subseteq [\text{essinf}\, W_k,\, \text{esssup}\, W_k]$ for the "certainty equivalent" of each W_k. The set $\mathcal{R}_{ce}$ 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.,*

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

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 Lemma 2.2(ii)).

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

$$Z_M^\varepsilon = \begin{cases} -M & \text{w.p. } \varepsilon(M), \\ 0 & \text{w.p. } 1-\varepsilon(M). \end{cases} \tag{3.15}$$

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

$$\mathcal{R}_{bnd}(\varepsilon) := \{\rho : \mathscr{L}_p \to \mathbb{R} \,|\, \rho(Z_M^\varepsilon) \leq \rho(-1), \forall\, M \geq 1\}. \tag{3.16}$$

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 $l \in \mathcal{L}$, the risk measure $\mathrm{SR}_l^P \in \mathcal{R}_{bnd}(\varepsilon^*)$ with $\varepsilon^*(M) := (l(0) - l(-1))/(l(M-1) - l(-1))$ based on the argument that:

$$\varepsilon^*(M)l(M-1) + (1-\varepsilon^*(M))l(-1) = l(0)$$
$$\Rightarrow\ \mathrm{SR}_l^P(Z_M^{\varepsilon^*}) = \inf\{t : \varepsilon^*(M)l(M-t) + (1-\varepsilon^*(M))l(-t) \leq l(0)\} = 1 = \mathrm{SR}_l^P(-1).$$

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 $\rho(Z_M^{\bar{\varepsilon}}) \leq \rho(-Z_{\bar{\lambda}}) \leq \rho(-1)$.

Proposition 3.1 (Characterization of L for $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}}(Z)$) *The preference robust risk measure $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}}(Z)$ is equivalent to $\mathrm{SR}_{L_{ce}}^P(Z)$ where*

$$L_{ce} := \left\{ l \in \mathcal{L} \,\middle|\, \begin{array}{l} \mathbb{E}_P[l(-W_k + w_k^-)] \leq l(0) \\ \mathbb{E}_P[l(-W_k + w_k^+)] \geq l(0) \end{array} ,\ \forall\, k \in \{1, 2, \ldots, K\} \right\}. \tag{3.17}$$

Proof. Based on Theorem 2.1, we only need to demonstrate that

$$L_{ce} = L_{\mathcal{R}\cap\mathcal{R}_{ce}} := \{l \in \mathcal{L} \,|\, \mathrm{SR}_l^P \in \mathcal{R} \cap \mathcal{R}_{ce}\}.$$

We start by showing that $L_{ce} \subseteq L_{\mathcal{R}\cap\mathcal{R}_{ce}}$. Namely, given any $\bar{l} \in L_{ce}$, we show that $\rho_{\bar{l}} := \mathrm{SR}_{\bar{l}}^P \in \mathcal{R} \cap \mathcal{R}_{ce}$. It is first easy to confirm that since $\bar{l} \in \mathcal{L}$, $\rho_{\bar{l}}$ is a legitimate normalized convex utility-based shortfall risk measure. We hence are left with verifying that $\rho_{\bar{l}} \in \mathcal{R}_{ce}$. To do so, we exploit the fact that $\rho_{\bar{l}}(w_k^+) = -w_k^+$ and $\rho_{\bar{l}}(w_k^-) = -w_k^-$ (see Lemma 2.2(i)). 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

$$\mathbb{E}_P[\bar{l}(-W_k + w_k^+)] \geq \bar{l}(0) \Longrightarrow \mathbb{E}_P[\bar{l}(-W_k + w_k^+ + \epsilon)] > \bar{l}(0)\,,\, \forall \epsilon > 0. \tag{3.18}$$

To see this, we note that

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

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 $\rho_{\bar{l}} \in \mathcal{R} \cap \mathcal{R}_{ce}$.

Next, we show that $L_{ce} \supseteq L_{\mathcal{R}\cap\mathcal{R}_{ce}}$. In other words, given any $\bar{\rho} \in \mathcal{R} \cap \mathcal{R}_{ce}$, there exists a $\bar{l} \in L_{ce}$ such that $\bar{\rho} = \text{SR}_{\bar{l}}^P$. First, based on Proposition 2.1, there is necessarily a $\bar{l} \in \mathcal{L}$ 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 2.3 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 $\bar{l} \in L_{ce}$ and therefore that $L_{ce} \supseteq L_{\mathcal{R}\cap\mathcal{R}_{ce}}$. ■

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

$$\min_{x\in X} \;\; \varrho_{\mathcal{R}\cap\mathcal{R}_{ce}}(-c(x,\xi)) \qquad \equiv \qquad \begin{array}{ll}\displaystyle\min_{x\in X,t} & t\\ \text{s.t.} & \mathbb{E}_P[l(c(x,\xi) - t)] \leq 0,\, \forall\, l \in L_{ce}\,.\end{array} \tag{3.19}$$

In Section 4, we will address the computational challenges arising from this reformulation by using an analysis that is similar in spirit to the one used in [1] for the case where ξ has a finite discrete distribution. Later, in Section 5, 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 [1] 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 3.2 (Characterization of L for $\varrho_{\mathcal{R}_{coh}\cap\mathcal{R}_{ce}}(Z)$) *Given that $\mathcal{R}_{coh}\cap\mathcal{R}_{ce}\neq\emptyset$, the risk measure $\varrho_{\mathcal{R}_{coh}\cap\mathcal{R}_{ce}}(Z)$ is equivalent to $\mathrm{SR}^P_{l_\tau}(Z)$ where*

$$l_\tau(s)=\max(\tau s,\,(1-\tau)s)\ \text{ for }\ \tau=\min_k\frac{\mathbb{E}[(W_k-w_k^-)^+]}{\mathbb{E}[|W_k-w_k^-|]},\tag{3.20}$$

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

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

Proof. Based on [3], it is well known that the class of utility-based shortfall risk measures that are both convex and positive homogeneous, namely $\mathcal{R}_{coh}$, coincides with

$$\mathcal{R}_{coh}=\{\mathrm{SR}^P_l\,|\,l\in L_{coh}\},$$

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 $L_{\mathcal{R}_{coh}}=L_{coh}$ and by a similar argument as in the proof of Proposition 3.1 we can also show that $L_{\mathcal{R}_{coh}\cap\mathcal{R}_{ce}}=L_{coh,ce}:=L_{coh}\cap L_{ce}$.

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 $b \le 1$ since $(W_k - w_k^-)^+ \le |W_k - w_k^-|$. ■

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

$$\min_{x\in X} \quad \varrho_{\mathcal{R}_{coh}\cap\mathcal{R}_{ce}}(-c(x,\xi)) \qquad \equiv \qquad \begin{array}{ll} \min\limits_{x\in X,t} & t \\ \text{s.t.} & \mathbb{E}_P[l_\tau(c(x,\xi)-t)] \le 0\,, \end{array} \tag{3.21}$$

with τ defined according to Proposition 3.2. In comparison with the case where the risk measure is not known to be positive homogeneous, formulation (3.21) 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 the sensitivity of the utility-based shortfall risk measure to losses of large size.

Proposition 3.3 (Characterization of L for $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}\cap\mathcal{R}_{bnd}(\varepsilon)}$) *The risk measure $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}\cap\mathcal{R}_{bnd}(\varepsilon)}(Z)$ is equivalent to $\mathrm{SR}^P_{L_{ce}\cap L_{bnd}}(Z)$ with*

$$L_{bnd} := \{l \,|\, \varepsilon(z)l(z-1) + (1-\varepsilon(z))l(-1) \le l(0)\,,\, \forall\, z \ge 1\}\,. \tag{3.22}$$

Moreover, each loss function $l \in L_{ce} \cap L_{bnd}$ satisfies

$$l(z_1) - l(0) \le \frac{1-\varepsilon(z_1+1)}{\varepsilon(z_1+1)}(l(0) - l(-1))\,,\, \forall\, z_1 \ge 0 \tag{3.23}$$

and

$$l'_+(z_1) \le \phi(z_1)(l(0)-l(-1)) \,\text{for}\, 0 \le z_1 < z_2, \tag{3.24}$$

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

$$\phi(z_1) = \inf\left\{\left(\frac{1-\varepsilon(z+1)}{\varepsilon(z+1)} - z_1\right)\frac{1}{z-z_1} \;:\; z > z_1\right\}. \tag{3.25}$$

Similarly, to the case of Proposition 3.1, this result leads to a semi-infinite reformulation for the (PRSRP) problem. Namely,

$$\min_{x\in X} \quad \varrho_{\mathcal{R}\cap\mathcal{R}_{ce}\cap\mathcal{R}_{bnd}(\varepsilon)}(-c(x,\xi)) \qquad \equiv \qquad \begin{array}{ll} \min_{x\in X,t} & t \\ \text{s.t.} & \mathbb{E}_P[l(c(x,\xi)-t)] \le 0\,,\, \forall\, l \in L_{ce}\cap L_{bnd}\,. \end{array} \tag{3.26}$$

Proof. The treatment of $\mathcal{R}_{bnd}(\varepsilon)$ is analogous to the treatment of $\mathcal{R}_{ce}$ considering that for each $M \ge 1$, we are imposing that

$$\rho(\text{esssup}(Z_M)) = \rho(0) \le \rho(Z_M^\varepsilon) \le \rho(-1)\,.$$

Hence, following similar arguments as in the proof of Proposition 3.1, we get that:

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

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 $L_{\mathcal{R}\cap\mathcal{R}_{ce}\cap\mathcal{R}_{bnd}(\varepsilon)} = L_{ce} \cap L_{bnd}$.

Furthermore, after replacing $z_1 := z-1$ in the definition of L_{bnd}, we get that all $l \in L_{ce} \cap L_{bnd}$ 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 (3.23). In what follows, we prove (3.24) also holds for such l. By exploiting the convexity of l and (3.23), we obtain

$$l(0) + (l(0)-l(-1))z \leq l(z) \leq l(0) + \frac{1-\varepsilon(z+1)}{\varepsilon(z+1)}(l(0)-l(-1)), \forall z \geq 0. \tag{3.27}$$

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

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

This completes our proof. ■

Remark 3.1 *We may draw a couple of useful conclusions from the theorem.*

(i) $\phi(z)$ is non-decreasing over $[-M, M]$. To see this, we note that the convexity of $l(\cdot)$ ensures

$$l(z) - l(0) \geq z(l(0)-l(-1)), \forall z > 0$$

and through (3.23), we have

$$\frac{1-\varepsilon(z+1)}{\varepsilon(z+1)} - z \geq 0, \forall z > z_1.$$

This implies the objective function of the minimization problem in (3.25) 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

$$\phi(z) \leq \phi(M), \forall z \in [-M, M].$$

On the other hand, since $l(\cdot)$ is convex and non-decreasing, then the inequality above and inequality (3.24) imply that

$$l'_-(z) \leq l'_+(z) \leq \phi(M)(l(0) - l(-1)), \forall z \in [-M, M],$$

where $l'_-(z)$ denotes the left derivative of $l(\cdot)$ at z. The latter ensures that every loss function in $L_{ce} \cap L_{bnd}$ is equicontinuous over any interval $[-M, M]$, i.e.

$$\frac{l(z_2) - l(z_1)}{z_2 - z_1} \leq \phi(M)(l(0) - l(-1)), \forall z_1, z_2 \in [-M, M], \forall l \in L_{ce} \cap L_{bnd},$$

which means the class of loss functions in $L_{ce} \cap L_{bnd}$ are equicontinuous over $[-M, M]$ for any $M > 0$.

4 Tractable formulation of (PRSRP)

The representations that we have developed for $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}}(Z)$ and $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}\cap\mathcal{R}_{bnd}(\varepsilon)}(Z)$ in (3.19) and (3.26) 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 (PRSRP).

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

Proposition 4.1 *The (PRSRP) problem with L_{ce} can be reformulated as the following problem:*

$$\min_{x\in X,u,\gamma,\eta^{(1)},\eta^{(2)}} \quad t \tag{4.28a}$$

$$\text{s.t.} \quad p_i(c(x,\xi_i)-t) - \sum_{j=1}^{M} u_{ij} y_j \leq 0, \ \forall\, i=1,\ldots,N, \tag{4.28b}$$

$$p_i - \sum_{j=1}^{M} u_{ij} = 0, \ \forall\, i=1,\ldots,N, \tag{4.28c}$$

$$\sum_{i=1}^{N} u_{ij} + \sum_{m=1}^{M}\gamma_{jm} - \sum_{m=1}^{M}\gamma_{mj} - \sum_{k=1}^{K}\eta_k^{(1)} P(-W_k + w_k^- = y_j) + \sum_{k=1}^{K}\eta_k^{(2)} P(-W_k + w_k^+ = y_j) = 0, \ j = \{1,\ldots,M\}\setminus\{j_0, j_-\}, \tag{4.28d}$$

$$\sum_{i=1}^{N} u_{ij_-} + \sum_{m=1}^{M}\gamma_{j_-m} - \sum_{m=1}^{M}\gamma_{mj_-} - \sum_{k=1}^{K}\eta_k^{(1)} P(-W_k + w_k^- = y_{j_-}) + \sum_{k=1}^{K}\eta_k^{(2)} P(-W_k + w_k^+ = y_{j_-}) \geq 0, \tag{4.28e}$$

$$\sum_{m=1}^{M}\gamma_{mj}(y_m - y_j) \geq 0\,, \ \forall\, j=1,\ldots,M, \tag{4.28f}$$

$$u_{ij}\geq 0, \gamma_{mj}\geq 0, \eta^{(1)}\geq 0, \eta^{(2)}\geq 0, \ i=1,\ldots,N, m,j=1,\ldots,M, \tag{4.28g}$$

where $\{y_j\}_{j=1}^M$ *is an indexed list of the elements of* $\mathcal{Y} := \bigcup_{k=1}^K \operatorname{supp}(-W_k + w_k^-) \cup \operatorname{supp}(-W_k + w_k^+) \cup \{0\} \cup \{-1\}$*, 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* (4.28) *can be reformulated as a finite dimensional linear programming problem.*

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

$$\begin{aligned} \min_{x\in X,t} \quad & t \\ \text{s.t.} \quad & \sup_{l\in L_{ce}} \mathbb{E}_P[l(c(x,\xi)-t) - l(0)] \leq 0. \end{aligned} \tag{4.29}$$

Observe first that L_{ce} is a cone, thus $\sup_{l\in L_{ce}} \mathbb{E}_P[l(c(x,\xi)-t) - l(0)] = \infty$ whenever there exists a $l \in L_{ce}$ 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:

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

and show that

$$\left\{ t : \sup_{l\in L_{ce}} \mathbb{E}_P[l(c(x,\xi)-t) - l(0)] \leq 0 \right\} = \left\{ t : \sup_{l\in L'_{ce}} \mathbb{E}_P[l(c(x,\xi)-t) - l(0)] \leq 0 \right\}. \tag{4.30}$$

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

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

which means that the right-hand side of equation (4.30) is a subset of the left-hand side. The relationship (4.30) allows us to replace L_{ce} with L'_{ce} in the constraint of problem (4.29).

Step 1. We start by giving an infinite dimensional linear programming representation for the right-hand side of (4.30). First, letting $\Psi(x,t) := \sup_{l \in L'_{ce}} \mathbb{E}_P[l(c(x,\xi) - t) - l(0)]$, we can expand the Ψ operator to

$$\begin{aligned}
\Psi(x,t) = \sup_{l:\mathbb{R}\to\mathbb{R}, f:\mathbb{R}\to\mathbb{R}} \quad & \sum_{i=1}^{N} p_i l(c(x,\xi_i) - t) - l(0) && \text{(4.31a)} \\
\text{s.t.} \quad & l(y') \geq l(y) + (y' - y) f(y)\,,\ \forall\,(y,y') \in \mathbb{R} \times \mathbb{R}, && \text{(4.31b)} \\
& l(0) = 0, && \text{(4.31c)} \\
& l(-1) = -1, && \text{(4.31d)} \\
& f(y) \geq 0\,,\ \forall\, y \in \mathbb{R}, && \text{(4.31e)} \\
& \sum_{\{y:P(-W_k + w_k^- = y) > 0\}} P(-W_k + w_k^- = y) l(y) \leq l(0), && \text{(4.31f)} \\
& \sum_{\{y:P(-W_k + w_k^+ = y) > 0\}} P(-W_k + w_k^+ = y) l(y) \geq l(0). && \text{(4.31g)}
\end{aligned}$$

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

$$\begin{aligned}
\Psi(x,t) = \sup_{v \geq 0, w, l:\mathbb{R}\to\mathbb{R}, f:\mathbb{R}\to\mathbb{R}} \quad & \sum_{i=1}^{N} p_i [v_i (c(x,\xi_i) - t) + w_i] - l(0) && \text{(4.32a)} \\
\text{s.t.} \quad & v_i y + w_i \leq l(y)\,,\ \forall\, y \in \mathbb{R}, \forall i, && \text{(4.32b)} \\
& l(y') \geq l(y) + (y' - y) f(y)\,,\ \forall\,(y,y') \in \mathbb{R} \times \mathbb{R}, && \text{(4.32c)} \\
& l(0) = 0, && \text{(4.32d)} \\
& l(-1) = -1, && \text{(4.32e)} \\
& f(y) \geq 0\,,\ \forall\, y \in \mathbb{R}, && \text{(4.32f)} \\
& \sum_{\{y:P(-W_k + w_k^- = y) > 0\}} P(-W_k + w_k^- = y) l(y) \leq l(0), && \text{(4.32g)} \\
& \sum_{\{y:P(-W_k + w_k^+ = y) > 0\}} P(-W_k + w_k^+ = y) l(y) \geq l(0). && \text{(4.32h)}
\end{aligned}$$

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

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

where

$$\begin{aligned}
\tilde{\Psi}(x,t) := \sup_{v\ge 0, w, l:\mathbb{R}\to\mathbb{R}, f:\mathbb{R}\to\mathbb{R}} \quad & \sum_i p_i[v_i(c(x,\xi_i)-t)+w_i] - l(0) && (4.33a)\\
\text{s.t.} \quad & v_i y + w_i \le l(y)\,,\ \forall\, y \in \mathcal{Y}, \forall i = 1,\cdots,N, && (4.33b)\\
& l(y') \ge l(y) + (y'-y)f(y)\,,\ \forall\,(y,y') \in \mathcal{Y}\times\mathcal{Y}, && (4.33c)\\
& l(0) = 0, && (4.33d)\\
& l(-1) = -1, && (4.33e)\\
& f(y) \ge 0\,,\ \forall\, y \in \mathcal{Y}, && (4.33f)\\
& \sum_{\{y: P(-W_k+w_k^-=y)>0\}} P(-W_k+w_k^- = y)l(y) \le l(0), && (4.33g)\\
& \sum_{\{y: P(-W_k+w_k^+=y)>0\}} P(-W_k+w_k^+ = y)l(y) \ge l(0), && (4.33h)
\end{aligned}$$

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 $\tilde{\Psi}(x,t) \le 0 \Rightarrow \Psi(x,t) \le 0$. This follows straightforwardly from the fact that the feasible set of problem (4.32) is smaller than that of problem (4.33) and consequently $\Psi(x,t) \le \tilde{\Psi}(x,t)$.

To see the reverse implication $\Psi(x,t) \le 0 \Rightarrow \tilde{\Psi}(x,t) \le 0$, 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 (4.32) 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 (4.33) which achieves a strictly positive objective value. We construct $\hat{l}$ as

$$\hat{l}(y) := \begin{cases} \sup_{v,w: v\ge 0,\, vy'+w \le \bar{l}(y')\ \forall y' \in \mathcal{Y}} \ vy + w & \text{if } y \le y_*, \\ \max\{\max_i \bar{v}_i y + \bar{w}_i\,,\ \hat{v}y + \hat{w}\} & \text{otherwise,} \end{cases} \tag{4.34}$$

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 $y \le y_*$, 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 (4.31b)-(4.31g) hold, for some non-negative sub-derivative function $\hat{f}$. We are left with the task to check that the objective value of (4.31) gives a strictly positive value. In particular,

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

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

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

while when $c(x,\xi_i)-t>y^*$, again by construction $\hat{l}(c(x,\xi_i)-t)\ge\ \max_{i'}\{\bar{v}_{i'}(c(x,\xi_i)-t)+\bar{w}_{i'}\}$. This allows us to conclude that if $\tilde{\Psi}(x,t)>0$ then $\Psi(x,t)>0$ meaning that $\Psi(x,t)\le 0\Rightarrow\tilde{\Psi}(x,t)\le 0$. We complete this first step by arguing that since in problem (4.33), 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

$$\begin{aligned}\tilde{\Psi}(x,t)=\sup_{v\ge 0,w,l:\mathcal{Y}\to\mathbb{R},f:\mathcal{Y}\to\mathbb{R}} \quad & \sum_i p_i[v_i(c(x,\xi_i)-t)+w_i]-l(0)\\ \text{s.t.}\quad & v_iy+w_i\le l(y)\,,\ \forall\, y\in\mathcal{Y},\forall i=1,\cdots,N\\ & l(y')\ge l(y)+(y'-y)f(y)\,,\ \forall\,(y,y')\in\mathcal{Y}\times\mathcal{Y},\\ & l(0)=0,\\ & l(-1)=-1,\\ & f(y)\ge 0\,,\ \forall\, y\in\mathcal{Y},\\ & \sum_{\{y:P(-W_k+w_k^-=y)>0\}} P(-W_k+w_k^-=y)l(y)\le l(0),\\ & \sum_{\{y:P(-W_k+w_k^+=y)>0\}} P(-W_k+w_k^+=y)l(y)\ge l(0)\,.\end{aligned}$$

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 $k=1,\dots,K$ 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 $j=1,\dots,M$.

Then $\tilde{\Phi}(x,t)$ can be rewritten as

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

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

$$
\begin{aligned}
\min_{u,\gamma,\nu_0,\nu_-,\lambda,\eta^{(1)},\eta^{(2)}} \quad & \nu_- \\
\text{s.t.} \quad & p_i(c(x,\xi_i)-t) - \sum_{j=1}^{M} u_{ij} y_j \leq 0,\ \forall\, i=1,\dots,N, \\
& p_i - \sum_{j=1}^{M} u_{ij} = 0,\ \forall\, i=1,\dots,N, \\
& \sum_{i=1}^{N} u_{ij} + \sum_{m=1}^{M} \gamma_{jm} - \sum_{m=1}^{M} \gamma_{mj} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_j) \\
& \qquad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_j) = 0,\ j = \{1,\dots,M\} \setminus \{j_0, j_-\}, \\
& \sum_{i=1}^{N} u_{ij_-} + \sum_{m=1}^{M} \gamma_{j_-m} - \sum_{m=1}^{M} \gamma_{mj_-} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_{j_-}) \\
& \qquad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_{j_-}) = -\nu_- \\
& \sum_{i=1}^{N} u_{ij_0} + \sum_{m=1}^{M} \gamma_{j_0m} - \sum_{m=1}^{M} \gamma_{mj_0} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_{j_0}) \\
& \qquad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_{j_0}) = \nu_0, \\
& \lambda_j - \sum_{m=1}^{M} \gamma_{mj}(y_m - y_j) = 0 \,,\ \forall\, j=1,\dots,M, \\
& u_{ij} \geq 0, \gamma_{mj} \geq 0, \lambda \geq 0, \eta^{(1)} \geq 0, \eta^{(2)} \geq 0,\ i=1,\dots,N, m,j=1,\dots,M.
\end{aligned}
$$

Realizing that ν_0 and ν_- both only appear in one of the constraints allows us to simplify the model slightly to

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

and reintegrate it in the (PRSRP). The proof is complete. ■

We now move on to discuss a finite dimensional reformulation of (PRSRP) with $L_{ce} \cap L_{bnd}$ 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 4.2 *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 (PRSRP) problem with $L_{ce} \cap L_{bnd}$ is equivalent to*

$$
\begin{aligned}
\min_{x,t,u,\sigma,\gamma,\rho,\theta,\eta^{(1)},\eta^{(2)}} \quad & t && (4.35a)\\
\text{s.t.} \quad & t_- \le t \le t_+ && (4.35b)\\
-\Bigg(\sum_{i=1}^{N} u_{ij_-} + \sum_{m=1}^{M} \gamma_{j_-m} - \sum_{m=1}^{M} \gamma_{mj_-} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_{j_-}) & \\
+ \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_{j_-})\Bigg) + \sum_{j\in\mathcal{I}} \theta_{j-M}(z_j - 1) + \sum_{j=j_0+1}^{M} \rho_{j-j_0}(z_j - 1) \le 0, & && (4.35c)\\
& p_i(c(x,\xi_i) - t) - \sum_{j=1}^{M} u_{ij} y_j - \sigma_i \le 0,\ \forall\, i = 1,\dots,N, && (4.35d)\\
& p_i - \sum_{j=1}^{M} u_{ij} = 0,\ \forall\, i = 1,\dots,N, && (4.35e)\\
& \sum_{i=1}^{N} u_{ij} + \sum_{m=1}^{M} \gamma_{jm} - \sum_{m=1}^{M} \gamma_{mj} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_j) && (4.35f)\\
& \qquad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_j) = 0,\ j = \{1,\dots,j_0 - 1\} \setminus \{j_-\}, && \\
& \sum_{i=1}^{N} u_{ij} + \sum_{m=1}^{M} \gamma_{jm} - \sum_{m=1}^{M} \gamma_{mj} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_j) && (4.35g)\\
& \qquad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_j) = -\rho_{j-j_0}, \forall\, j = j_0 + 1,\dots,M - 1, && \\
& \sum_{i=1}^{N} u_{iM} + \sum_{m=1}^{M} \gamma_{Mm} - \sum_{m=1}^{M} \gamma_{mM} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_M) && (4.35h)\\
& \qquad + \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_M) = -\rho_{M-j_0} - \sum_{j\in\mathcal{I}} \theta_{j-M}, && \\
& \sum_{m=1}^{M} \gamma_{mj}(y_m - y_j) \ge 0\,,\ \forall\, j = 1,\dots,M - 1, && (4.35i)\\
& \sum_{m=1}^{M} \gamma_{mM}(y_m - y_M) \ge \sum_{i=1}^{N} \sigma_i - \sum_{j\in\mathcal{I}} \theta_{j-M}(y_j - y_M), && (4.35j)\\
& u_{ij} \ge 0, \gamma_{mj} \ge 0, \forall\, i = 1,\dots,N, m, j = 1,\dots,M, && (4.35k)\\
& \sigma \ge 0, \rho \ge 0, \theta \ge 0, \eta^{(1)} \ge 0, \eta^{(2)} \ge 0, && (4.35l)
\end{aligned}
$$

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 $\mathcal{Y}'_\varepsilon := \{y \in \Re \,|\, y+1 \in \mathcal{Y}_\varepsilon\}$ *and where*

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

and where 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* $j = j_0+1, j_0+2, \ldots$.

Proof. We can apply similar analysis as in the proof of Proposition 4.1 for $\Phi(x,t) := \sup_{l\in L'_{ce}\cap L_{bnd}} \mathbb{E}_P[l(c(x,\xi)-t) - l(0)]$ which appears in the (PRSRP) problem presented as :

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

which can be reduced to

$$\min_{x\in X, t\in[t_-,\,t_+]} t \quad \text{s.t.} \quad \Phi(x,t) \le 0\,, \tag{4.36}$$

since for any $x \in X$

$$\mathrm{SR}^P_L(-c(x,\xi)) \le \mathrm{SR}^P_L(-\max_{x\in X}\max_i c(x,\xi_i)) = t_+$$

and

$$\mathrm{SR}^P_L(-c(x,\xi)) \ge \mathrm{SR}^P_L(-\min_{x\in X}\min_i c(x,\xi_i)) = t_-\,.$$

Note that we can exploit once more the fact that for all $x \in X$ and $t \in [t_-,\, t_+]$:

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

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

$$\begin{aligned}
\Phi(x,t) = \sup_{l:\mathbb{R}\to\mathbb{R}, f:\mathbb{R}\to\mathbb{R}} \quad & \sum_{i=1}^N p_i l(c(x,\xi_i)-t) - l(0) && (4.37a)\\
\text{s.t.} \quad & l(y') \ge l(y) + (y'-y)f(y)\,,\ \forall\,(y,y') \in \mathbb{R}\times\mathbb{R}, && (4.37b)\\
& l(y) \le \varepsilon(y+1)^{-1} - 1\,,\ \forall\, y \ge 0, && (4.37c)\\
& l(0) = 0, && (4.37d)\\
& l(-1) = -1, && (4.37e)\\
& f(y) \ge 0\,,\ \forall\, y \in \mathbb{R}, && (4.37f)\\
& \sum_{\{y:P(-W_k+w_k^-=y)>0\}} P(-W_k + w_k^- = y) l(y) \le l(0), && (4.37g)\\
& \sum_{\{y:P(-W_k+w_k^+=y)>0\}} P(-W_k + w_k^+ = y) l(y) \ge l(0)\,, && (4.37h)
\end{aligned}$$

where we were able to simply replace $l(0) = 0$ and $l(-1) = -1$ in constraint (3.22). Using the fact that l is a convex function, we once again replace the objective function to obtain

$$\begin{aligned}
\Phi(x,t) = \sup_{v\geq 0, w, l:\mathbb{R}\to\mathbb{R}, f:\mathbb{R}\to\mathbb{R}} \quad & \sum_i p_i[v_i(c(x,\xi_i)-t)+w_i] - l(0) && (4.38a)\\
\text{s.t.} \quad & v_i y + w_i \leq l(y)\,,\ \forall\, y\in\mathbb{R}, \forall i, && (4.38b)\\
& l(y') \geq l(y) + (y'-y)f(y)\,,\ \forall\,(y,y')\in\mathbb{R}\times\mathbb{R}, && (4.38c)\\
& l(y) \leq \varepsilon(y+1)^{-1} - 1\,,\ \forall\, y\geq 0, && (4.38d)\\
& l(0) = 0, && (4.38e)\\
& l(-1) = -1, && (4.38f)\\
& f(y)\geq 0\,,\ \forall\, y\in\mathbb{R}, && (4.38g)\\
& \sum_{\{y:P(-W_k+w_k^-=y)>0\}} P(-W_k+w_k^-=y)l(y) \leq l(0), && (4.38h)\\
& \sum_{\{y:P(-W_k+w_k^+=y)>0\}} P(-W_k+w_k^+=y)l(y) \geq l(0)\,. && (4.38i)
\end{aligned}$$

Moreover, one can establish that the following two constraints can be added to problem (4.38) without affecting the supremum value:

$$\begin{aligned}
& v_i \leq f(y^*)\,,\ \forall\, i, && (4.39a)\\
& l(y^*) + f(y^*)(y-y^*) \leq \varepsilon(y+1)^{-1} - 1\,,\ \forall\, y\in \mathcal{Y}'_\varepsilon \cap]y^*,\infty[\,. && (4.39b)
\end{aligned}$$

Namely, constraint (4.39b) 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 (4.38c) and (4.38d). On the other hand, while constraint (4.39a) 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,

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

where we first used the fact that $l(y^*) \geq v_i y^* + w_i$, 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

$$
\begin{aligned}
\tilde{\Phi}(x,t) := \sup_{v\geq 0, w, l:\mathcal{Y}'\to\mathbb{R}, f:\mathcal{Y}'\to\mathbb{R}} \quad & \sum_i p_i[v_i(c(x,\xi_i)-t)+w_i] - l(0) && (4.40a)\\
\text{s.t.} \quad & v_i y + w_i \leq l(y)\,,\ \forall\, y\in\mathcal{Y}', \forall i, && (4.40b)\\
& v_i \leq f(y^*)\,,\ \forall\, i, && (4.40c)\\
& l(y') \geq l(y) + (y'-y)f(y)\,,\ \forall\,(y,y')\in\mathcal{Y}'\times\mathcal{Y}', && (4.40d)\\
& l(y) \leq \varepsilon(y+1)^{-1}-1\,,\ \forall\, y\in\mathcal{Y}'\cap\mathbb{R}^+, && (4.40e)\\
& l(y^*)+f(y^*)(y-y^*) \leq \varepsilon(y+1)^{-1}-1\,,\ \forall\, y\in\mathcal{Y}'_\varepsilon\cap(y^*,\infty), && (4.40f)\\
& l(0)=0\ ,\ l(-1)=-1, && (4.40g)\\
& f(y)\geq 0\,,\ \forall\, y\in\mathcal{Y}', && (4.40h)\\
& \sum_{\{y:P(-W_k+w_k^-=y)>0\}} P(-W_k+w_k^-=y)l(y) \leq l(0), && (4.40i)\\
& \sum_{\{y:P(-W_k+w_k^+=y)>0\}} P(-W_k+w_k^+=y)l(y) \geq l(0), && (4.40j)
\end{aligned}
$$

where by definition $y^* = \max_{y\in\mathcal{Y}'} y$.

While $\tilde{\Phi}(x,t)\leq 0 \Rightarrow \Phi(x,t)\leq 0$ 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 (4.40) and one can therefore construct a loss function $\hat{l}$ according to

$$
\hat{l}(y) := \begin{cases} \sup_{v,w:v\geq 0,\, vy'+w\leq\bar{l}(y')\ \ \forall y'\in\mathcal{Y}'} \ vy+w & \text{if } y\leq y^*,\\ \bar{l}(y^*)+\bar{f}(y^*)(y-y^*) & \text{otherwise}. \end{cases} \tag{4.41}
$$

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 (4.37b) and (4.37d)-(4.37h) for some non-negative sub-derivative function $\hat{f}(y)$. We can also verify as in the proof of Proposition 4.1 that it returns a strictly positive objective value. Namely,

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

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

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

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

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

Before concluding that $\tilde{\Phi}(x,t) > 0 \Rightarrow \Phi(x,t) > 0$, we must confirm that $\hat{l}(y)$ also satisfies constraint (4.37c). First, in the case that $0 \leq y \leq y^*$, then either $y \in \mathcal{Y}'$ and (4.37c) is satisfied since $\hat{l}(y) = \bar{l}(y) \leq \varepsilon(y+1)^{-1} - 1$, or by construction

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

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

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

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 (4.37c) is satisfied by $\hat{l}(y)$.

To complete this proof, one simply needs to confirm (4.35) by strong duality for problem (4.40). Recall that $\mathcal{Y}'$ includes the union of the set $\mathcal{Y}$ defined in Proposition 4.1, $\{y_*, y^*\}$ 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:

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

$$u_{ij} \geq 0, \gamma_{mj} \geq 0, \forall\, i = 1, \ldots, N, m, j = 1, \ldots, M,$$
$$\sigma \geq 0, \rho \geq 0, \theta \geq 0, \lambda \geq 0, \eta^{(1)} \geq 0, \eta^{(2)} \geq 0.$$

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 slightly to

$$\min_{u,\sigma,\gamma,\rho,\theta,\eta^{(1)},\eta^{(2)}} \quad -\Bigg(\sum_{i=1}^{N} u_{ij_-} + \sum_{m=1}^{M} \gamma_{j_- m} - \sum_{m=1}^{M} \gamma_{mj_-} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_{j_-}) \tag{4.42a}$$
$$+ \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_{j_-})\Bigg) + \sum_{j\in\mathcal{I}} \theta_{j-M}(z_j - 1) + \sum_{j=j_0+1}^{M} \rho_{j-j_0}(z_j - 1)$$
$$\text{s.t.} \quad p_i(c(x,\xi_i) - t) - \sum_{j=1}^{M} u_{ij} y_j - \sigma_i \leq 0,\ \forall\, i = 1, \ldots, N, \tag{4.42b}$$
$$p_i - \sum_{j=1}^{M} u_{ij} = 0,\ \forall\, i = 1, \ldots, N, \tag{4.42c}$$
$$\sum_{i=1}^{N} u_{ij} + \sum_{m=1}^{M} \gamma_{jm} - \sum_{m=1}^{M} \gamma_{mj} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_j) \tag{4.42d}$$
$$+ \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_j) = 0,\ j = \{1, \ldots, j_0 - 1\} \setminus \{j_-\},$$
$$\sum_{i=1}^{N} u_{ij} + \sum_{m=1}^{M} \gamma_{jm} - \sum_{m=1}^{M} \gamma_{mj} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_j) \tag{4.42e}$$
$$+ \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_j) = -\rho_{j-j_0}, \forall\, j = j_0 + 1, \ldots, M - 1,$$
$$\sum_{i=1}^{N} u_{iM} + \sum_{m=1}^{M} \gamma_{Mm} - \sum_{m=1}^{M} \gamma_{mM} - \sum_{k=1}^{K} \eta_k^{(1)} P(-W_k + w_k^- = y_M) \tag{4.42f}$$
$$+ \sum_{k=1}^{K} \eta_k^{(2)} P(-W_k + w_k^+ = y_M) = -\rho_{M-j_0} - \sum_{j\in\mathcal{I}} \theta_{j-M},$$
$$\sum_{m=1}^{M} \gamma_{mj}(y_m - y_j) \geq 0\,,\ \forall\, j = 1, \ldots, M - 1, \tag{4.42g}$$
$$\sum_{m=1}^{M} \gamma_{mM}(y_m - y_M) \geq \sum_{i=1}^{N} \sigma_i - \sum_{j\in\mathcal{I}} \theta_{j-M}(y_j - y_M), \tag{4.42h}$$
$$u_{ij} \geq 0, \gamma_{mj} \geq 0, \forall\, i = 1, \ldots, N, m, j = 1, \ldots, M, \tag{4.42i}$$
$$\sigma \geq 0, \rho \geq 0, \theta \geq 0, \eta^{(1)} \geq 0, \eta^{(2)} \geq 0, \tag{4.42j}$$

which can be reinserted in the (PRSRP) problem (4.36). ■

Note that in Proposition 4.2 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 4.2 assumes $\mathcal{Y}_\varepsilon$ to be unbounded, it must be that $\mathcal{I}$ is an infinite set thus the linear program (4.35) 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 [16] and reference therein).

Algorithm 1 Column generation for solving (PRSRP) problem (4.35)

1: $\mathcal{I}' \leftarrow \emptyset$
2: **repeat**
3: Solve problem (4.35) with $\mathcal{I}$ restricted to $\mathcal{I}'$
4: $\alpha_M^* \leftarrow$ dual variables for constraint (4.35h) at optimum of restricted problem
5: $\beta_M^* \leftarrow$ dual variables for constraint (4.35j) 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 3.2) are used which should make constraint (4.37c) 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 4.3 *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*

$$y^*_{\alpha,\beta} := \arg \min_{y\in\mathcal{Y}'_\varepsilon \cap]y^*,\,\infty[} \varepsilon(y+1)^{-1} - 1 - \alpha - \beta(y - y^*) \tag{4.43}$$

can be found by solving

$$\bar{y}^*_{\alpha,\beta} := \arg \min_{y\in[y^*,\,\infty[} \bar{\varepsilon}(y+1)^{-1} - 1 - \alpha - \beta(y - y^*) \tag{4.44}$$

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 4.3, we can solve (4.43) by solving problem (4.44) 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.

Proof. 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. ∎

We conclude this section with an example involving $\bar{\varepsilon}(y) := \exp(-\bar{\lambda}y)$ as described in Section 3. In this case, it is clear that $\bar{\varepsilon}(y)^{-1} = \exp(\bar{\lambda}y)$ does not satisfy the condition imposed in Proposition 4.2, 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 $\mathcal{Y}_\varepsilon := \{2^k\}_{k=0}^\infty$.

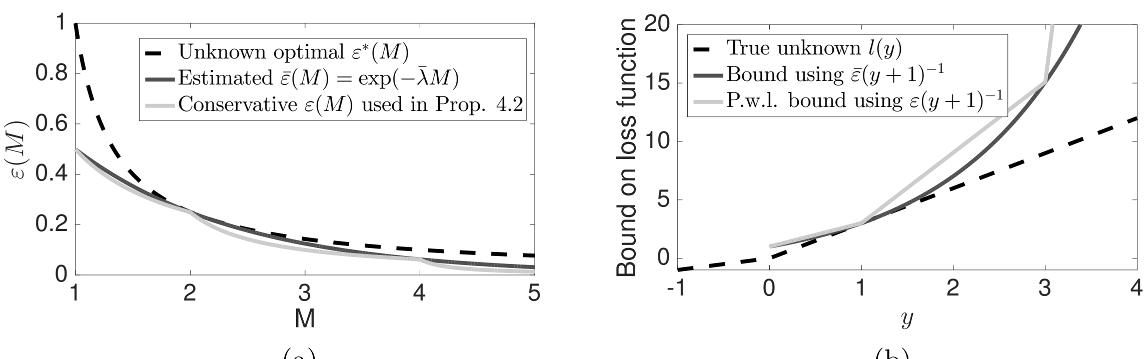


Figure 1: Illustration of the effect of manipulating $\bar{\varepsilon}$ in order to satisfy the assumptions made in propositions 4.2 and 4.3 thus making the resolution of problem (4.35) more tractable using Algorithm 1.

With this particular choice of characterization for $\varepsilon(y)$, Algorithm 1 can be used to solve the (PRSRP) problem (4.35). Following the result of Proposition 4.3, 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.

5 Discrete approximation of (PRSRP) when P is continuous

A key condition for tractable reformulation of problem (1.5) 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.

5.1 Problem set-up

By the definition of the preference robust normalized risk measure, we can write problem (1.5) as

$$\begin{array}{ll} \min\limits_{x\in X, t\in\mathbb{R}} & t \\ \text{s.t.} & \sup\limits_{l\in L} \mathbb{E}_P[l(c(x,\xi)-t)] - l(0) \leq 0. \end{array} \tag{5.45}$$

Let P_N be a discrete approximation of P. We consider

$$\begin{array}{ll} \min\limits_{x\in X, t\in\mathbb{R}} & t \\ \text{s.t.} & \sup\limits_{l\in L} \mathbb{E}_{P_N}[l(c(x,\xi)-t)] - l(0) \leq 0. \end{array} \tag{5.46}$$

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

$$v(x,t) := \sup_{l\in L} \mathbb{E}_P[l(c(x,\xi)-t)] - l(0) \tag{5.47}$$

and

$$v_N(x,t) := \sup_{l\in L} \mathbb{E}_{P_N}[l(c(x,\xi)-t)] - l(0). \tag{5.48}$$

Consequently, we can rewrite (5.45) and (5.46) as

$$\text{(PRSRP)} \qquad \begin{array}{cl} \displaystyle\min_{x\in X, t\in\mathbb{R}} & t \\ \text{s.t.} & v(x,t) \le 0, \end{array} \tag{5.49}$$

and

$$\text{(PRSRP-N)} \qquad \begin{array}{cl} \displaystyle\min_{x\in X, t\in\mathbb{R}} & t \\ \text{s.t.} & v_N(x,t) \le 0. \end{array} \tag{5.50}$$

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

Assumption 5.1 *We assume: (a) X is a compact set, (b) $c(\cdot,\cdot)$ is a continuous function on $X \times \Xi$, (c) (PRSRP) satisfies Slater condition, i.e., there exist a positive constant number θ, and $x_0 \in X, t_0 \in \mathbb{R}$ such that*

$$\sup_{l\in L} \mathbb{E}_P[l(c(x_0,\xi) - t_0)] - l(0) \le -\theta, \tag{5.51}$$

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

Under Assumption 5.1, the optimal value of PRSRP is finite. To see this, we note that

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

Following a similar proof to (2.12), we can show

$$\lim_{t\to-\infty} \sup_{l\in L} \mathbb{E}_P[l(Z - t)] = +\infty,$$

which implies the t component of the feasible set $\mathcal{F}$ must have a lower bound and hence the optimal value $\vartheta > -\infty$. On the other hand, condition (5.51) ensures $\vartheta \le t_0$.

5.2 Sample average approximation

Let $\xi^1, \ldots, \xi^N$ be iid samples of ξ and

$$\mathbb{1}_{\xi^k}(\xi(\omega)) := \begin{cases} 1, & \text{if}\, \xi(\omega) = \xi^k, \\ 0, & \text{if}\, \xi(\omega) \ne \xi^k. \end{cases}$$

Let

$$P_N(\cdot) := \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{\xi^k}(\cdot). \tag{5.52}$$

Instead of deriving uniform approximation specifically for $v_N(x,t)$ defined in (5.48), 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 [19, 22] 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 5.2 *Let Ξ be the support set of ξ.*

(a) For any $\epsilon > 0$, there exists a compact set $\Xi_\epsilon \subset \Xi$ such that

$$\sup_{N, x \in X, \phi \in \mathcal{W}} \mathbb{E}_{P_N}[|\phi(g(x,\xi) \mathbb{1}_{\Xi \setminus \Xi_\epsilon}(\xi))|] \leq \epsilon, \text{ w.p.1.} \tag{5.53}$$

(b) For any $M > 0$, there exist positive constants κ_M, Δ_M (both depending on M) and $\lambda \in [-M, M]$ such that

$$\sup_{\phi \in \mathcal{W}} |\phi(z_1) - \phi(z_2)| \leq \kappa_M |z_1 - z_2|, \forall z_1, z_2 \in [-M, M], \tag{5.54}$$

and $\sup_{\phi \in \mathcal{W}_M} \|\phi(\lambda)\| \leq \Delta_M$, where κ_M increases as M increases and $\mathcal{W}_M$ is defined as follows:

$$\mathcal{W}_M := \{\phi|_{[-M,M]}(\cdot) : \forall \phi(\cdot) \in \mathcal{W}\}, \tag{5.55}$$

where $\phi|_{[-M,M]}(\cdot)$ stands for the restriction of function $\phi(z)$ to being defined over interval $[-M, M]$.

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

$$|g(x,\xi) - g(x',\xi)| \leq r(\xi) \|x - x'\|^\nu, \forall x, x' \in X, \xi \in \Xi. \tag{5.56}$$

Assumption 5.2 (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 [2]. Condition (b) requires the class of functions in $\mathcal{W}$ to be equi-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 [19].

Under Assumption 5.2 (a)-(c), the set $\mathcal{W}_M$ is bounded by

$$\sup_{\phi \in \mathcal{W}_M} \|\phi\|_\infty \leq \Delta_M + 2\kappa_M M. \tag{5.57}$$

By Ascoli-Arzela Theorem ([4]), the equi-Lipschitz continuity condition (5.54) and the unform boundedness (5.57) 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

$$\mathcal{W}_M = \bigcup_{k=1}^{K} (\mathcal{W}_M)_k^{\epsilon} \tag{5.58}$$

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

Lemma 5.1 *Let Assumption 5.2 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*

$$\text{Prob}\left(\sup_{x\in X} \left|\Phi_N(x) - \Phi(x)\right| \geq \delta\right) \leq C(\epsilon, \delta) e^{-N\beta(\epsilon,\delta)}, \tag{5.59}$$

where 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 $\mathscr{B} \times \mathscr{B} \times \cdots$.

Proof. Under Assumption 5.2 (a), for any $\epsilon > 0$, there exists a compact set $\Xi_\epsilon \subset \Xi$ such that (5.53) holds. Let

$$M_\epsilon := \sup_{x\in X, \xi\in\Xi_\epsilon} |g(x, \xi)|$$

and $\mathcal{W}_{M_\epsilon}$ be defined as in (5.55). 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 (5.53),

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

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 [20, Lemma 1] or [11, Lemma 2.1] that for any $x \in X, \phi \in \mathcal{W}_{M_\epsilon}$,

$$\mathbb{E}_P[\phi(g(x, \xi)\mathbb{1}_{\Xi\backslash\Xi_\epsilon}(\xi))] = \lim_{N\to\infty} \mathbb{E}_{P_N}[\phi(g(x, \xi)\mathbb{1}_{\Xi\backslash\Xi_\epsilon}(\xi))] \leq \epsilon,$$

which 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 $\Phi(x)$,

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

For any $\delta>0$, we may set ϵ sufficiently small such that $\epsilon<\delta/4$. Under Assumption 5.2 (b) and (c), for any $\phi\in\mathcal{W}_{M_\epsilon}$,

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

It follows from [19, Theorem 5.1] that for each k there exist positive constants $C(\epsilon,\delta,\phi_k)$ and $\beta(\epsilon,\delta,\phi_k)$ such that

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

Hence, we have

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

which implies (5.59). ■

A crucial requirement in Lemma 5.1 is Assumption 5.2 (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 5.1 *Let Ξ be compact and Assumptions 5.2 (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*

$$\mathrm{Prob}\left(\sup_{x\in X}\big|\Phi_N(x)-\Phi(x)\big|\ge\delta\right)\le C(\epsilon,\delta)e^{-N\beta(\epsilon,\delta)}.$$

Proof. The result follows from Lemma 5.1 by setting $M := \sup_{x\in X,\xi\in\Xi}|g(x,\xi)|$ and $\Xi_\epsilon = \Xi$. ■

Note that Haskell et al. [13] 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 (PRSRP) 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 Mehorotra [14] 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.

5.3 Convergence of the optimal values and optimal solutions

We now return to discuss convergence of (PRSRP-N) to (PRSRP) in terms of the optimal values and optimal solutions. Let $v_N(x,t)$ and $v(x,t)$ be defined as in (5.48) and (5.47). We start by deriving uniform convergence of $v_N(x,t)$ to $v(x,t)$ using Lemma 5.1. 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 $\sup_{l\in L_M}\|l(\lambda)\|\le\Delta_M$.

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

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

Following Remark 3.1, we know that condition (C2) is satisfied by the loss functions from $L_{ce}\cap L_{bnd}$ and l_τ defined in (3.20).

Theorem 5.1 *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*

$$\text{Prob}\left(\sup_{x\in X,t\in T}\left|v_N(x,t)-v(x,t)\right|\ge\delta\right)\le C(\epsilon,\delta)e^{-N\beta(\epsilon,\delta)} \quad (5.60)$$

for $N\ge N(\epsilon,\delta)$.

Proof. 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 5.1, 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. ■

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

Theorem 5.2 *Let ϑ, ϑ_N, S and S_N be defined as in Section 5.1 and P_N be defined as in (5.52). Let Assumption 5.1 and conditions (C1)-(C3) hold. Suppose that for almost every $\xi \in \Xi$, $c(\cdot,\xi)$ is a convex function. Then*

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

$$\text{Prob}\left(|\vartheta_N - \vartheta| \geq \delta\right) \leq C(\epsilon,\varepsilon)e^{-N\beta(\epsilon,\varepsilon)}, \tag{5.61}$$

for $N \geq N(\epsilon,\varepsilon)$ where $N(\epsilon,\varepsilon)$, $C(\epsilon,\varepsilon)$ and $\beta(\epsilon,\varepsilon)$ are defined as in Theorem 5.1 and ε is some positive constant depending on δ, and θ is given in (5.51).

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

Proof. Part (i). Let $t^* = \vartheta$. Following the discussions immediately after Assumption 5.1, we know that t^* is finite and $t^* \leq t_0$. Let θ be defined as in Assumption 5.1 and δ be given as in Theorem 5.1 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

$$\inf_{x\in X} v(x, t^* - c_\eta) \geq \eta. \tag{5.62}$$

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 (5.62) holds.

Let T in Theorem 5.1 be chosen such that $[t^* - c_\eta, t_0] \subset T$. Then by Theorem 5.1

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

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 (PRSRP-N). The inequality above shows

$$v_N(x_N, t^* - c_\eta) \geq \inf_{x\in X} v_N(x, t^* - c_\eta) > \eta - \delta/2, \tag{5.63}$$

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 (5.51) and Theorem 5.1,

$$\sup_{l\in L} \mathbb{E}_{P_N}[l(c(x_0, \xi) - t_0) - l(0)] < -\theta + \delta/2 \leq -\delta/2. \tag{5.64}$$

with probability at least $1-C(\epsilon, \delta/2)e^{-N\beta(\epsilon,\delta/2)}$ for $N \geq N(\epsilon, \delta/2)$. The inequality (5.64) implies (x_0, t_0) is a feasible solution to (PRSRP-N) and hence $t_N \leq t_0$. Summarizing the discussions above, we have

$$t_N \in [t^* - c_\eta, t_0] \tag{5.65}$$

with probability at least $1 - C(\epsilon, \delta/2)e^{-N\beta(\epsilon,\delta/2)}$.

Let us now consider the systems of inequalities

$$v(x, t) \leq 0, (x, t) \in X \times T$$

and

$$v_N(x, t) \leq 0, (x, t) \in X \times T.$$

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 (5.51), we may use Robinson's error bound theorem for convex systems ([17]) to establish that,

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

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 (5.64) to obtain

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

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

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

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

$$|\vartheta_N - \vartheta| = |t_N - t^*| \leq \mathbb{H}(\mathcal{F}_N \cap (X \times T), \mathcal{F} \cap (X \times T)) \leq \frac{2\Delta}{\delta} \sup_{x \in X, t \in T} |v_N(x,t) - v(x,t)|. \quad (5.66)$$

Let $\varepsilon := \min(\frac{\delta^2}{2\Delta}, \frac{\delta}{2})$. We deduce from (5.60) and (5.66)

$$\begin{aligned} \text{Prob}(|\vartheta_N - \vartheta| \geq \delta) &\leq \text{Prob}\left(\sup_{x \in X, t \in T} |v_N(x,t) - v(x,t)| \geq \varepsilon \right) \\ &\leq C(\epsilon, \varepsilon) e^{-N\beta(\epsilon,\varepsilon)} \end{aligned}$$

for $N \geq N(\epsilon, \varepsilon)$.

Part (ii). The exponential rate of convergence (5.61) implies $t_N \to t^*$ almost surely. Moreover, since $v_N(x_N, t_N) \leq 0$ and v_N converges uniformly to v over $X \times T$, then $v(\hat{x}, t^*) \leq 0$ for every cluster point $\hat{x}$ of $\{x_N\}$. ■

Theorem 5.2 ensures ϑ_N converges to ϑ at exponential rate with increase of the sample size N.

6 Numerical Experiments

In this section, we repeat the experiments performed in [6] 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:

$$\begin{aligned} \min_x \quad & \rho(\xi^T x) \\ \text{s.t.} \quad & \sum_{i=1}^n x_i = 1, \\ & x \geq 0\,, \end{aligned}$$

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 $\rho(\xi^T x) := \text{SR}^P_{l_\tau}(\xi^T x)$ 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 [6], and two preference robust

[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 $\rho(\xi^T x)$.

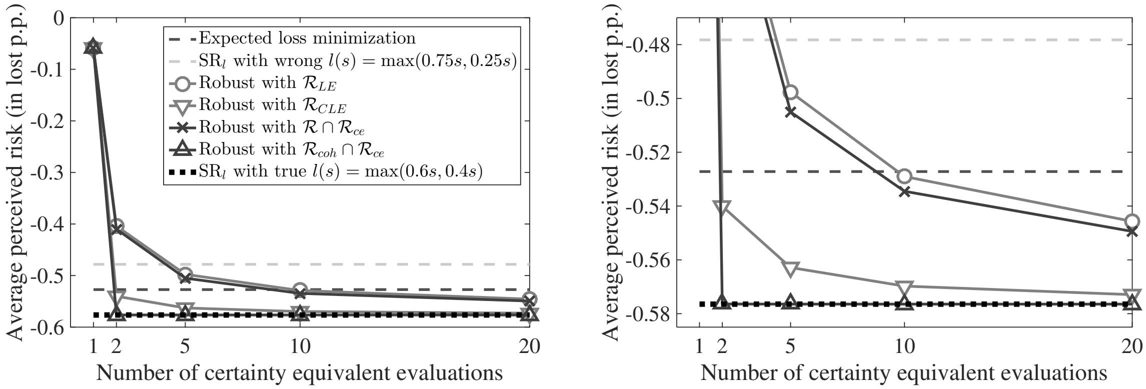
risk measures that additionally account for the fact that the true risk measure is a utility-based SR, namely $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}}(\cdot) = \mathrm{SR}^P_{L_{ce}}(\cdot)$ and $\varrho_{\mathcal{R}_{coh}\cap\mathcal{R}_{ce}}(\cdot) = \mathrm{SR}^P_{l_\tau}(\cdot)$ with the worst-case τ defined in equation (3.20).

Our experiments are nearly identical to the experiments described in [6]. 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 $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}\cap\mathcal{R}_{bnd}(\varepsilon)}$ but quickly realized that the portfolios obtained using this method were almost undistinguishable from the ones obtained using $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}}$ in our test environment. In particular, in order to obtain a piecewise linear function for $\varepsilon(\cdot)^{-1}$ that satisfies the conditions of Proposition 4.3, we started by identifying a value $\bar{\lambda}$ such that $\bar{\varepsilon}(y) := \exp(-\bar{\lambda}y)$ satisfies Definition 3.3 for our reference investor modeled with $\bar{\rho}$. As seen in Proposition 3.3, 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 $\bar{\lambda} = \max_{y\geq 0} \ln((\tau/(\tau-1))y+1)/(y+1) \approx 0.4716$. 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 (4.28). This motivated us to conclude that the performance of $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}\cap\mathcal{R}_{bnd}(\varepsilon)}$ 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

[3] This is the same period and same companies as in [6] except for BMC software which was removed from our data set given that it was privatized in September 2013.



(a) (b)

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.

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 $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}\cap\mathcal{R}_{bnd}(\varepsilon)}$ instead of $\varrho_{\mathcal{R}\cap\mathcal{R}_{ce}}$ in order to obtain a guarantee on the convergence of (PRSRP-N) to (PRSRP) (see Theorem 5.2) is in fact reasonable.

7 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 [18]), the convergence result that we established in Section 5 extends the existing results by Haskell et al. [13] 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 ([14]). 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.

References

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

[2] Billingsley P (1999), *Convergence of Probability Measures*, Wiley, New York.

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

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

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

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

[7] Dunkel J, Weber S (2010), Stochastic root finding and efficient estimation of convex risk measures, *Operations Research*, 58(5): 1505–1521.

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

[9] Föllmer H, Schied A (2011), *Stochastic Finance*, 3rd ed, Berlin: de Gruyter.

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

[11] 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.

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

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

[14] 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.

[15] Hu Z, Zhang D (2016), Convex risk measures: efficient computations via monte carlo, *manuscript*.

[16] 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.

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

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

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

[20] Sun H, Xu H (2015), Convergence analysis for distributionally robust optimization and equilibrium problems, *Mathematics of Operations Research*, 41(2): 377–401.

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

[22] Xu H (2010), Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, *Journal of Mathematical Analysis and Applications*, 368(2): 692–710.

A Proofs of Section 2

A.1 Proof of Lemma 2.1

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 $l(\cdot)$,

$$l(t) - l(t_2) \geq \frac{l(t_2) - l(t_1)}{t_2 - t_1}(t - t_2)$$

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). ■

A.2 Proof of Lemma 2.2

Property (i) follows the translation invariance of convex risk measures, namely $\rho(Z+t) = \rho(Z)-t$ for all $t \in \mathbb{R}$, and the fact that the risk measure is normalized hence $\rho(c) = \rho(0+c) = \rho(0) - c = -c$. Property (ii) 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 $\text{essinf}(Z) \in \mathbb{R}$, since $Z \geq \text{essinf}(Z)$, we can conclude that

$$\rho(Z) \leq \rho(\text{essinf}(Z)) = -\text{essinf}(Z)$$

while otherwise $\text{essinf}(Z) = -\infty$ hence $\rho(Z) \leq \infty$ follows trivially. One can establish a similar result with respect to $\text{esssup}(Z)$. ■

A.3 Proof of Lemma 2.3

This result is covered by [9, Proposition 4.113]. Here we include a proof for completeness as we need it in a number of proofs later on. Let $h(t) := \mathbb{E}_P[l(-Z-t)]$ and t_0 be such that

$$h(t_0) = \mathbb{E}_P[l(-Z-t_0)] = l(0). \tag{A.67}$$

Existence of t_0 is guaranteed by the fact that: (a) the fact that $\rho(Z)$ is finite implies that the feasible set $\{t \,|\, \mathbb{E}[l(-Z-t)] - l(0) \leq 0\}$ is nonempty and hence there exists a $\underline{t}$ such that

$$h(\underline{t}) = \mathbb{E}_P[l(-Z-\underline{t})] \leq l(0);$$

(b) when $t \to -\infty$, we have that $h(t) \to +\infty$ based on Lemma 2.1(ii). Hence, we have that t_0 must exist since $h(t)$ is continuous.

Now, we show that t_0 is actually unique by demonstrating that $h(\cdot)$ is strictly decreasing near t_0. Let Ω_+ and Ω_- be a partition of Ω such that

$$-Z(\omega) - t_0 > z_0, \forall \omega \in \Omega_+, \tag{A.68}$$

and

$$-Z(\omega) - t_0 \leq z_0, \forall \omega \in \Omega_-, \tag{A.69}$$

where z_0 is defined as in Definition 2.1. We necessarily have that $P(\Omega_+) > 0$ otherwise we would have $P(\Omega_-) = 1$ and then

$$\mathbb{E}_P[l(-Z-t_0)] \leq \mathbb{E}_P[l(z_0)] = l(z_0) < l(0) = \mathbb{E}_P[l(-Z-t_0)]\,,$$

which is a contradiction. Now, given any $\epsilon > 0$, we can establish that

$$l(-Z(\omega) - t_0 + \epsilon) > l(-Z(\omega) - t_0), \forall \omega \in \Omega_+\,, \tag{A.70}$$

so that

$$\begin{aligned}
h(t_0 - \epsilon) &= \mathbb{E}_P[l(-Z - t_0 + \epsilon)] \\
&= \int_{\Omega_+} l(-Z(\omega) - t_0 + \epsilon) P(d\omega) + \int_{\Omega_-} l(-Z(\omega) - t_0 + \epsilon) P(d\omega) \\
&> \int_{\Omega_+} l(-Z(\omega) - t_0) P(d\omega) + \int_{\Omega_-} l(-Z(\omega) - t_0 + \epsilon) P(d\omega) \\
&\geq l(0) = h(t_0).
\end{aligned}$$

In the other direction, we can more carefully first assess that since $P(\Omega_+) > 0$ there must necessarily exist a small enough $\delta > 0$ such that

$$P(-Z(\omega) - t_0 - \epsilon > z_0) > 0\,, \forall\, \epsilon \in [0,\, \delta]\,.$$

Letting Ω_+^ϵ be this specific set of outcomes, it follows that for all $\epsilon \in [0, \delta]$,

$$\begin{aligned} h(t_0+\epsilon) &= \mathbb{E}_P[l(-Z-t_0-\epsilon)] \\ &= \int_{\Omega_+^\epsilon} l(-Z(\omega)-t_0-\epsilon)P(d\omega) + \int_{\Omega/\Omega_+^\epsilon} l(-Z(\omega)-t_0-\epsilon)P(d\omega) \\ &< \int_{\Omega_+^\epsilon} l(-Z(\omega)-t_0)P(d\omega) + \int_{\Omega/\Omega_+^\epsilon} l(-Z(\omega)-t_0-\epsilon)P(d\omega) \\ &\leq l(0) = h(t_0). \end{aligned}$$

This allows us to conclude that $h(t)$ is strictly increasing around t_0 and therefore that t_0 is unique.

We now return to showing that $t^* := \rho(Z)$ is the unique solution of equation (2.8). Since ρ is a normalized convex utility-based shortfall risk measure, by Proposition 2.1, it has representation (2.7). Moreover, since $\rho(Z)$ is finite and the feasible set of the minimization problem (2.7) is closed, then the infimum is achieved hence $h(t^*) \leq l(0)$. The strict inequality does not hold because otherwise t^* could be further reduced and hence would not be optimal. This therefore confirms that t^* is the unique solution of equation (2.8). ■

A.4 Proof of Corollary 2.1

To obtain the equivalence portrayed in equation (2.14), we follow the steps below:

$$\begin{aligned} \mathrm{SR}_L^P(Z) \leq \gamma &\iff \sup_{l\in L} \inf\{t : \mathbb{E}_P[l(-Z-t)] - l(0) \leq 0\} \leq \gamma && \text{(A.71a)} \\ &\iff \inf\{t : \mathbb{E}_P[l(-Z-t)] - l(0) \leq 0\} \leq \gamma \,,\, \forall\, l \in L && \text{(A.71b)} \\ &\iff \mathbb{E}_P[l(-Z-\gamma)] - l(0) \leq 0 \,,\, \forall\, l \in L && \text{(A.71c)} \\ &\iff \sup_{l\in L}\{\mathbb{E}_P[l(-Z-\gamma)] - l(0)\} \leq 0 \,, && \text{(A.71d)} \end{aligned}$$

where the first equivalence comes from Theorem 2.1 and the second and fourth exploit a well known property of supremums. The third equivalence follows from Lemma 2.3 since, letting t^* taking the value of the infimum in equation (A.71b), we can check that for any $l \in L$,

$$\inf\{t : \mathbb{E}_P[l(-Z-t)] - l(0) \leq 0\} \leq \gamma \;\;\Rightarrow\;\; \mathbb{E}_P[l(-Z-\gamma)] - l(0) \leq \mathbb{E}_P[l(-Z-t^*)] - l(0) = 0\,,$$

since $l(\cdot)$ is non-decreasing, while

$$\mathbb{E}_P[l(-Z-\gamma)] - l(0) \leq 0 \;\;\Rightarrow\;\; \inf\{t : \mathbb{E}_P[l(-Z-t)] - l(0) \leq 0\} \leq \gamma\,.$$

This completes our proof. ■

A.5 Proof of Proposition 2.2

One can indeed easily demonstrate that the supremum achieved over a set of risk measures that all satisfy law-invariance, monotonicity, convexity, and translation invariance must necessarily satisfy these same properties. In order to show that $\mathrm{SR}_L^P(Z)$ is not in general a utility-based shortfall risk measure, we will identify a counter example. Namely, let the ambiguity set for $\hat{L}$ be defined as follows:

$$\hat{L} := \left\{ l \in \mathcal{L} \;\middle|\; \begin{array}{l} (1/2)(l(-3) + l(1)) = l(0) \\ (1/2)(l(-35) + l(5)) = l(0) \\ l(0) = 0 \end{array} \right\}.$$

One cans easily first check that set $\hat{L}$ is non-empty given that the loss function $\bar{l}(y) := \max(y/3,\, (8/3)y-(5/3))$ is a member.

Our claim is that $\mathrm{SR}^P_{\hat{L}}(Z)$ is not a utility-based shortfall risk measure, or equivalently there is no $\hat{l}(\cdot)$ such that for all random variable Z

$$\mathrm{SR}^P_{\hat{L}}(Z) = \mathrm{SR}^P_{\hat{l}}(Z)\,.$$

In particular, let us assume that such a $\hat{l}(\cdot)$ would exist for the following pair of random variables:

$$Z_1 := \left\{ \begin{array}{ll} 3 & \text{w.p. } 35/38, \\ -5 & \text{w.p. } 3/38, \end{array} \right. \text{ and } Z_2 := \left\{ \begin{array}{ll} 35 & \text{w.p. } 1/6, \\ -1 & \text{w.p. } 5/6. \end{array} \right.$$

We will first establish that $\mathrm{SR}^P_{\hat{L}}(Z_1) = \mathrm{SR}^P_{\hat{L}}(Z_2) = 0$. We then provide arguments that indicate that for $\mathrm{SR}^P_{\hat{l}}(Z_1) = \mathrm{SR}^P_{\hat{l}}(Z_2) = 0$ to hold it must be that $\hat{l}(1) = 0$ which is a contradiction since $\hat{l}(\cdot)$ should be strictly increasing over the positives.

Step 1: We start by showing that $\mathrm{SR}^P_{\hat{L}}(Z_1) = \mathrm{SR}^P_{\hat{L}}(Z_2) = 0$. To first show that for both Z_1 and Z_2 the risk does not exceed 0, we simply exploit the constraints that must be satisfied by all loss functions $l \in \hat{L}$. In particular, by convexity we have that

$$\begin{aligned} l(0) \geq l(-3) + 3(l(-3) - l(-35))/32 \;&\Rightarrow\; 35l(-3) \leq 3l(-35) \\ &\Rightarrow\; (35/38)l(-3) + (3/38)l(5) \leq (3/38)l(-35) + (3/38)l(5) = 0 \\ &\Rightarrow\; \sup_{l\in\hat{L}} \mathbb{E}_P[l(-Z_1)] - l(0) \leq 0 \\ &\Rightarrow\; \mathrm{SR}^P_{\hat{L}}(Z_1) = \inf\{t : \sup_{l\in\hat{L}} \mathbb{E}_P[l(-Z_1 - t)] - l(0) \leq 0\} \leq 0\,, \end{aligned}$$

where the last step follows from Corollary 2.1. Similarly, we have by convexity that for all $l \in \hat{L}$

$$\begin{aligned} l(5) \geq l(1) + 4(l(1) - l(0)) \;&\Rightarrow\; 5l(1) \leq l(5) \\ &\Rightarrow\; (1/6)l(-35) + (5/6)l(1) \leq (1/6)l(-35) + (1/6)l(5) = 0 \\ &\Rightarrow\; \mathrm{SR}^P_{l}(Z_2) = \inf\{t : \sup_{l\in\hat{L}} \mathbb{E}_P[l(-Z_2 - t)] - l(0) \leq 0\} \leq 0\,. \end{aligned}$$

We can also show that the opposite inequality also holds for both Z_1 and Z_2. In the case of Z_1, we can use the function $\bar{l}(y) := \max(y/3,\, (8/3)y - (5/3))$, which can be shown to belong to $\hat{L}$, and confirm that

$$\mathbb{E}_P[\bar{l}(-Z_1)] - \bar{l}(0) = (35/38)(-1) + (3/38)(35/3) = 0 \;\Rightarrow\; \mathrm{SR}^P_{\hat{L}}(Z_1) \geq \mathrm{SR}^P_{\bar{l}}(Z_1) = 0\,,$$

based on Lemma 2.3. In the case of Z_2, a similar result is established using $\bar{l}(y) := \max((1/8)y-(5/8),\, y)$, namely that $\mathrm{SR}^P_{\hat{L}}(Z_2) \geq 0$. Hence, $\mathrm{SR}^P_{\hat{L}}(Z_1) = \mathrm{SR}^P_{\hat{L}}(Z_2) = 0$.

Step 2: We now show that $\text{SR}_{\hat{l}}^P(Z_1) = \text{SR}_{\hat{l}}^P(Z_2) = 0$ implies that $\hat{l}(1) = 0$. Based on Lemma 2.3, these equalities imply that

$$(35/38)\hat{l}(-3) + (3/38)\hat{l}(5) = \hat{l}(0),\ (1/6)\hat{l}(-35) + (5/6)\hat{l}(1) = \hat{l}(0)\,. \tag{A.72}$$

Based on the definition of $\hat{L}$, we further have that for

$$Z_3 := \begin{cases} 3 & \text{w.p. } 1/2, \\ -1 & \text{w.p. } 1/2, \end{cases} \quad \text{and } Z_4 := \begin{cases} 35 & \text{w.p. } 1/2, \\ -5 & \text{w.p. } 1/2, \end{cases}$$

one can establish that

$$\begin{aligned} \mathbb{E}_P[l(-Z_3)] - l(0) = 0\,,\ \forall\, l \in \hat{L} \quad &\Rightarrow \quad \text{SR}_l^P(Z_3) = 0\,,\ \forall\, l \in \hat{L} \\ &\Rightarrow \quad \text{SR}_{\hat{l}}^P(Z_3) = \sup_{l \in \hat{L}} \text{SR}_l^P(Z_3) = 0\,, \end{aligned}$$

where we used Lemma 2.3 and the definition of $\hat{l}$. One can draw a similar conclusion about Z_4, namely that $\text{SR}_{\hat{l}}^P(Z_4) = 0$.

Again according to Lemma 2.3, this implies that

$$(1/2)\hat{l}(-3) + (1/2)\hat{l}(1) = \hat{l}(0),\ (1/2)\hat{l}(-35) + (1/2)\hat{l}(5) = \hat{l}(0). \tag{A.73}$$

The four equality constraints described in equation (A.72) and (A.73) can be used to conclude that

$$\begin{aligned} -(3/7)(\hat{l}(1) - \hat{l}(0)) &= -(3/35)(\hat{l}(0) - \hat{l}(-35)) = -(3/35)(\hat{l}(5) - \hat{l}(0)) \\ &= -(\hat{l}(0) - \hat{l}(-3)) = -(\hat{l}(1) - \hat{l}(0))\,, \end{aligned}$$

which can only be true if $\hat{l}(1) - \hat{l}(0) = 0$ which contradicts the fact that $\hat{l}(\cdot)$ should be increasing over the positive. ■