Utility Preference Robust Optimization with Moment-Type Information Structure

Shaoyan Guo[*] and Huifu Xu[†]

January 27, 2021

Abstract

Utility preference robust optimization (PRO) models are recently proposed to deal with decision making problems where the decision maker's true utility function is unknown and the optimal decision is based on the worst case utility function from an ambiguity set of utility functions. In this paper, we consider the case where the ambiguity set is constructed through some moment-type conditions (Hu and Mehrotra (2015)) and develop a numerical scheme for approximating the ambiguity set so that the resulting maximin optimization problem can be solved for nonconcave utility functions. To justify the approximation scheme, we derive an error bound for the approximated ambiguity set, the optimal value and optimal solutions of the resulting maximin problem. To address the data perturbation/contamination issues in the construction of the ambiguity set, we derive some stability results which quantify the variation of the ambiguity set against perturbation of the elicitation data and its propagation to the optimal value and optimal solutions of the PRO model. Finally, we extend the discussions to the case where the ambiguity set depends on the decision variables and the domain of the utility functions is unbounded. Some preliminary numerical results show that the proposed approximation schemes work very well.

keywords Piecewise linear approximation; non-concave utility functions; error bounds; data contamination; decision dependent ambiguity set, unbounded utility function

1 Introduction

The expected utility theory established by Von Neumann and Morgenstern (1947) has been a dominant normative and descriptive model of decision making. It states that any set of preferences that a decision maker may have among uncertain/risky prospects can be characterized by an expected utility function if the preferences satisfy certain reasonable axioms (i.e. completeness, transitivity, continuity and independence). Specifically, there exists a utility function

[*]School of Mathematical Sciences, Dalian University of Technology, Dalian, 116024, China. Email: syguo@dlut.edu.cn.

[†]Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong. Email: hfxu@se.cuhk.edu.hk

$u : \mathbb{R} \to \mathbb{R}$ such that the decision maker prefers prospect X over prospect Y if and only if $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$. The proof of the Von Neumann and Morgenstern (VNM) theorem provides a way to identify such a utility function when the support of the prospects is finite. In practice, however, identifying the VNM's utility function may not be an easy task either because the prospect space of a decision making problem is too large and/or too complex or information about the decision maker's preference is incomplete [1] (Karmarkar (1978) and Weber (1987)). In the literature of decision analysis and behavioural economics, a popular method is to elicit the decision maker's preferences with paired gambling approaches for preference comparisons or certainty equivalence (Farquhar (1984)) and use the elicited information to construct an approximate utility function via interpolation method, see for instance Clemen and Reilly (2001).

Armbruster and Delage (2015) take a different approach to handle the issue. Instead of trying to find an approximate VNM's utility function, they use the available information of the decision maker's preference such as preferring certain lotteries over other lotteries and being risk averse, S-shaped or prudent to construct an ambiguity set of plausible utility functions and then base the optimal decision on the worst case utility function from the ambiguity set. The approach is called preference robust optimization (PRO) as it follows the general philosophy of robust optimization. To show how the PRO problem can be solved efficiently, they derive a tractable reformulation which effectively recast the PRO as a finite dimensional linear programming problem when the ambiguity set has a specific structure.

Hu and Mehrotra (2015) also take a PRO approach to tackle the ambiguity of the true utility function but in a slightly different manner. They consider a probabilistic representation of the class of increasing concave utility functions by confining them to a compact interval and scaling them to being bounded by 1. In doing so, they propose a moment-type framework for constructing the ambiguity set of the decision maker's utility preference which covers a number of important approaches such as the certainty equivalent and pairwise comparison. Moreover, they impose a lower bound and an upper bound for the true unknown utility function and propose a step-like approximation of the functions in the moment condition for deriving tractable reformulation of the PRO model.

The research on PRO opens an interesting direction in the interdisciplinary area of robust optimization and behaviour economics. It has attracted increasing attentions over the past few years. For instances, Haskell et al. (2016) propose a robust model which handles ambiguity of the decision maker's preference and probability distribution of the underlying data in a single framework. Hu and Stepanyan (2017) propose a so-called reference-based almost stochastic dominance method for constructing a set of utility functions near a reference utility which satisfy certain stochastic dominance relationship and use the set to characterize the decision maker's preference. Hu et al. (2018) consider a PRO model with an ambiguity set of general utility functions and propose a Lagrangian function approach for solving the resulting maximin problem. More recently Vatanos et al. (2020) propose a PRO model for a recommender system which seeks to offer a user with unknown preferences an item.

The PRO approach has also been successfully applied to risk management problems where a

[1] This should be distinguished from the concept of incomplete preferences which refers to the case where a pair of alternatives are not comparable, see Galaabaatar and Karni (2013) for the latter.

decision maker/an investor does not know which risk measure may be used to precisely capture his/her risk attitude but knows which one is more preferable if he/she is given a pair of relatively simple choice of random prospects/investment opportunities. For example, Delage and Li (2018) bypass the expected utility representation of an investor's risk and propose a novel approach to tackle risk management problems where the investor's risk attitude is representable by convex risk measures (Föllmer and Schied (2002)). Delage et al. (2018a) propose a minimax robust shortfall risk optimization model to deal with the case where an investor's preferences can be characterized an ambiguity set of utility loss functions. Haskell et al. (2018) take a step further to consider a general robust quasiconvex multi-attribute choice function model which allows one to tackle ambiguity in both multi-attribute utility maximization problems and nonconvex risk minimization problems (Brown and Sim (2009); Brown et al. (2006)). More recently Wang and Xu (2020) propose a robust spectral risk model where the decision maker's risk attitude is ambiguous and can be characterized by a class of distorted risk spectra/functionals.

In this paper, we concentrate on utility-based PRO models where the ambiguity set is defined by moment-type conditions and the utility functions are not necessarily concave. We propose to approximate the ambiguity set so that the resulting PRO models can be more easily solved when the utility functions are non-concave. We also consider other issues such as data perturbation/contamination in the construction of the ambiguity, dependence of ambiguity set on the decision variables and unboundedness of the domain of utility functions. The specific contributions are summarized as follows.

First, we consider a PRO model where the true unknown utility function is increasing (nondecreasing) and extend it to situations where the ambiguity set is dependent on decision variables and the utility functions have unbounded domain. To solve the resulting maxmin problem, we propose a piecewise linear approximation (PLA) scheme for the utility functions in the ambiguity set so that the inner utility minimization problem can be reduced to a linear programming problem. To justify the PLA scheme, we derive an error bound for the optimal value and optimal solutions of the approximated PRO model (Theorem 4.2). The result is built on a newly derived Hoffman's lemma for the linear system (Lemma 4.1) in the infinite dimensional space under the pseudo-metric and quantification of the difference between the ambiguity sets before and after the PLA scheme (Theorem 4.1). Second, we develop efficient numerical schemes for solving the approximated PRO model. When the utility functions in the ambiguity set are increasing, concave and independent of decision variables, we reformulate the maximin PRO model as a single linear programming problem, which achieves the same results as Armbruster and Delage (2015) and Hu and Mehrotra (2015). In the case that the utility functions are not necessarily concave, or dependent on decision variables, the PLA scheme enables us to evaluate the inner minimization problem by solving a linear programming problem. Third, to facilitate application of the PRO model in a data-driven environment, we carry out stability analysis on the optimal value and optimal solutions of the PRO model against perturbation/contamination of the data both in the definition of the ambiguity set and the exogenous uncertainty and establish the quantitative and qualitative stability results (Theorems 5.1 and 5.2). Finally, we undertake numerical tests on the proposed PRO models and numerical schemes, and report the preliminary results.

The rest of the paper is organized as follows. Section 2 introduces the PRO model and definition of the ambiguity set. Section 3 details the PLA of the ambiguity set and computational schemes for solving the approximated PRO model. Section 4 investigates the error bound of the ambiguity set resulting from the PLA scheme and its impact on the optimal value and the optimal solutions of the PRO model. Section 5 discusses the stability of the PRO model against variation of data in the definition of the ambiguity set. Section 6 extends the PRO model to situations where the utility function is defined over the whole space $\mathbb{R}$ and the ambiguity set is dependent on decision variables. Section 7 reports numerical test results on the PRO models and numerical schemes when they are applied to a portfolio optimization problem and Section 8 concludes.

2 Ambiguity set with moment-type information structure

We consider the following one-stage expected utility maximization problem

$$\max_{x \in X} \ \mathbb{E}_P[u(f(x,\xi))], \tag{2.1}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a real-valued function representing a financial position or performance of an engineering design in practice, x is a decision vector restricted to take implementable decisions over $X \subset \mathbb{R}^n$, $\xi : (\Omega, \mathcal{F}, \mathbb{P}) \to \Xi \subset \mathbb{R}^m$ is a vector of random variables representing exogenous uncertainties in the decision making problem and the expectation is taken w.r.t. the probability distribution of ξ, i.e., $P = \mathbb{P} \circ \xi^{-1}$, $u : \mathbb{R} \to \mathbb{R}$ is a real-valued increasing function which maps each value of f to a utility value of the decision maker's interest. To facilitate our discussions, we make the following assumption throughout the paper.

Assumption 2.1 *f is a continuous function over $\mathbb{R}^n \times \mathbb{R}^m$, X is a compact and convex subset of $\mathbb{R}^n$, and the support set Ξ of ξ is compact.*

Assumption 2.1 allows us to restrict the domain of the unknown true utility function to an interval denoted by $[a, b]$. Moreover, we follow Hu and Mehrotra (2015) and the literature in behavioural economics to restrict the utility function to a particular class with $u(a) = 0$ and $u(b) = 1$.

In most of the existing research of expected utility optimization, the utility function is assumed to be known. Our focus here is on the situation where the decision maker does not have complete information on the utility function u, i.e., utility preference, but it is possible to elicit partial information to construct an ambiguity set of utility functions, denoted by $\mathcal{U}$, such that the true utility function which reflects precisely the decision maker's preference lies in $\mathcal{U}$ with high likelihood. Under such a circumstance, it might be sensible to consider a maximin preference robust optimization model to mitigate the risk arising from ambiguity of the true utility function

$$\text{(PRO)} \qquad \vartheta := \max_{x \in X} \min_{u \in \mathcal{U}} \ \mathbb{E}_P[u(f(x,\xi))]. \tag{2.2}$$

The structure of the (PRO) [2] model is largely determined by the structure of the ambiguity set $\mathcal{U}$ as well as the nature of the utility functions in this set. Various approaches have been proposed in the literature, see for instances Armbruster and Delage (2015); Hu and Mehrotra (2015); Hu and Stepanyan (2017). Here we consider the following moment-type information structure considered by Hu and Mehrotra (2015).

Definition 2.1 (Ambiguity with moment-type conditions) *Let $\mathscr{U}$ denote a class of non-constant increasing functions mapping from $[a,b]$ to $[0,1]$ where each function $u \in \mathscr{U}$ is piecewise continuously differentiable with finite number of non-differentiable points. The ambiguity set of utility function with moment-type information structure is defined as*

$$\mathcal{U} := \left\{ u \in \mathscr{U} : -\infty < \int_a^b \psi_j(t) du(t) \le c_j, \text{for } j = 1, \cdots, m \right\}, \tag{2.3}$$

where $\psi_j : [a,b] \to \mathbb{R}$, $j = 1, \cdots, m$ are Lebesgue integrable functions and $c_j, j = 1, \cdots, m$ are some given constants.

Since $u(\cdot)$ resembles a probability measure, we call informally the inequalities in (2.3) *moment-type conditions*. It is easy to verify that $\mathcal{U}$ is a convex set. Apart from the moment-type conditions specified in (2.3), Hu and Mehrotra (2015) also impose explicitly pointwise lower and upper bounds on the true unknown utility function and its derivative function in their definition of the ambiguity set. This is particularly sensible when the decision maker's has a nominal utility functions based on sample information or subjective judgement. Here we focus on (2.3) to facilitate our discussions.

The moment-type conditions cover a wide range of interesting cases, see Section 2 of that paper. Here we list some of them.

Example 2.1 (Hu and Mehrotra (2015)) *Let A and B be random returns of two investments with respective cumulative distribution functions (cdf for short) F_A and F_B. Suppose the supports of A and B are contained in $[a,b]$. In a survey, a decision maker is found to prefer B to A. Let u denote the decision maker's utility function. By the expected utility theory, the preference means $\mathbb{E}[u(A)] \le \mathbb{E}[u(B)]$, where the expectation is taken w.r.t. the distributions of A and B respectively, i.e.,*

$$\int_a^b u(t) dF_A(t) \le \int_a^b u(t) dF_B(t). \tag{2.4}$$

Assume that u is continuous over $[a,b]$ and is normalized with $u(a) = 0$ and $u(b) = 1$. By integration in parts, we can rewrite the inequality as

$$\int_a^b F_A(t) du(t) \ge \int_a^b F_B(t) du(t). \tag{2.5}$$

This means the decision maker's utility function lies in the ambiguity set:

$$\mathcal{U} = \left\{ u \in \mathscr{U} : \int_a^b (F_B(t) - F_A(t)) du(t) \le 0 \right\}. \tag{2.6}$$

[2] PRO is the acronym of preference robust optimization whereas (PRO) refers specifically to the model defined in (2.2).

Example 2.2 *(Armbruster and Delage (2015), Hu and Mehrotra (2015)) Consider a lottery X for an investor. Suppose the investor's certainty equivalent $\mathbb{C}_u(X) := u^{-1}(\mathbb{E}[u(X)])$ of the lottery falls into the range $[c_1, c_2]$. Then $c_1 \leq u^{-1}(\mathbb{E}[u(X)]) \leq c_2$ or equivalently $u(c_1) \leq \mathbb{E}[u(X)] \leq u(c_2)$. By integration in parts, we can rewrite the inequalities as*

$$\int_a^b (F_X(t) + \mathbb{1}_{[a,c_1]}(t))du(t) \leq 1 \text{ and } \int_a^b (-F_X(t) - \mathbb{1}_{[a,c_2]}(t))du(t) \leq -1,$$

where $\mathbb{1}_{[a,t]}(\cdot)$ denotes the indicator function over interval $[a, t]$. In this case, we can deduce from the investor's certainty equivalent that his/her utility function lies in the ambiguity set

$$\mathcal{U} = \left\{ u \in \mathscr{U} : \int_a^b (F_X(t) + \mathbb{1}_{[a,c_1]}(t))du(t) \leq 1, \int_a^b (-F_X(t) - \mathbb{1}_{[a,c_2]}(t))du(t) \leq -1 \right\}.$$

Example 2.3 *Consider the case that a decision maker has a nominal utility function $\hat{u}$ which is obtained from empirical data or subjective judgement. However, there is inadequate information to identify whether $\hat{u}$ is his/her true utility function. Additional case studies/questionnaires such as those in the previous two examples indicate that u satisfies the following inequalities*

$$r_j \leq \int_a^b g_j(t)du(t) - \int_a^b g_j(t)d\hat{u}(t) \leq R_j \text{ for } j = 1, \cdots, m, \tag{2.7}$$

where g_j, $j = 1, \cdots, m$ are some functions specified by the case studies and/or questionnaires. Since the term $\int_a^b g_j(t)d\hat{u}(t)$ can be incorporated into the constants, (2.7) is subsumed by the moment-type conditions (2.3).

Let $\mathscr{G} := \{g_1, \cdots, g_m\}$ and consider the case that $R_j = -r_j = r > 0$. Then (2.7) can be equivalently written as

$$\sup_{g \in \mathscr{G}} \left| \int_a^b g(t)du(t) - \int_a^b g(t)d\hat{u}(t) \right| \leq r. \tag{2.8}$$

The left hand side of inequality (2.8) defines a kind of "distance" between u and $\hat{u}$ under some pseudo-metric $\mathsf{dl}_{\mathscr{G}}$ generated by $\mathscr{G}$. We will give a formal definition of the pseudo-metric in Section 4.1. In this case the ambiguity set defined by (2.8) may be regarded as a ball of utility functions centred at $\hat{u}$ with radius r in the space of utility functions $\mathscr{U}$ equipped with the pseudo-metric $\mathsf{dl}_{\mathscr{G}}$.

Note that there may be other ways to construct the ambiguity set depending on availability of the true utility function. Consider the case that a decision maker has J candidate utility functions denoted by $u_1, \cdots, u_J$, which could be obtained from past experience or from stakeholders involved in the decision making problem. In such a case, the decision maker may construct the ambiguity set by convex combination of the utility functions, i.e.,

$$\mathcal{U} := \left\{ u = \sum_{j=1}^J \alpha_j u_j : (\alpha_1, \cdots, \alpha_J) \in \mathcal{A} \right\}, \tag{2.9}$$

where $\mathcal{A} := \{(\alpha_1, \cdots, \alpha_J) \in \mathbb{R}^J : \alpha_j \geq 0, j = 1, \cdots, J, \sum_{j=1}^J \alpha_j = 1\}$. *Consequently, the PRO model can be written as*

$$\vartheta := \max_{x \in X} \min_{(\alpha_1, \cdots, \alpha_J) \in \mathcal{A}} \mathbb{E}_P \left[\sum_{j=1}^J \alpha_j u_j(f(x, \xi)) \right] = \max_{x \in X} \min_{(\alpha_1, \cdots, \alpha_J) \in \mathcal{A}} \sum_{j=1}^J \alpha_j \mathbb{E}_P[u_j(f(x, \xi))].$$

3 Approximation of the ambiguity set

One of the main tasks of this paper is to develop efficient computational schemes for solving the (PRO) model. In the literature of preference robust optimization, tractable reformulations are often developed by utilization of support functions, see Armbruster and Delage (2015); Hu and Mehrotra (2015) for concave utility case and Haskell et al. (2018) for quasi-concave utility case. Here we take a different strategy.

3.1 Piecewise linear approximation

We propose to develop a piecewise linear approximation of the utility functions defined in the ambiguity set. Let $t_1 < \cdots < t_N$ be an ordered sequence of points in $[a, b]$ and $T := \{t_1, \cdots, t_N\}$ with $t_1 = a$ and $t_N = b$. For the convenience of citation in the later discussions, we call T the set of grid points. These points may be preset, e.g., evenly spread over $[a, b]$, or determined in the process of utility information elicitation (design of questionnaires, see Section 7.2).

Definition 3.1 *Let* $\mathscr{U}_N$ *be a subset of* $\mathscr{U}$ *where each function is a piecewise linear function with turning points on* T. *Define the ambiguity set of piecewise linear utility functions satisfying the moment-type conditions:*

$$\mathcal{U}_N := \left\{ u_N \in \mathscr{U}_N : \int_a^b \psi_j(t) du_N(t) \leq c_j, \text{ for } j = 1, \cdots, m \right\}. \tag{3.1}$$

We propose to use $\mathcal{U}_N$ to approximate $\mathcal{U}$. Since $\mathscr{U}_N \subset \mathscr{U}$, then $\mathcal{U}_N \subset \mathcal{U}$. Conversely, for any $u \in \mathcal{U}$, one can construct a piecewise linear utility function $u_N \in \mathscr{U}_N$ by connecting the function values at the ends of interval $[t_{i-1}, t_i]$ for each i. In general $u_N \notin \mathcal{U}_N$ but the inclusion may hold under some special cases. The following proposition states this.

Proposition 3.1 **($\mathcal{U}_N$ vs $\mathcal{U}$)** *If* $\psi_j(t)$ *is a step function over* $[a, b]$ *with jumps at* t_i *for* $i = 1, \cdots, N$, *then for any* $u \in \mathcal{U}$, *there exists a function* $u_N \in \mathscr{U}_N$ *with* $u_N(t_i) = u(t_i)$ *for* $i = 1, \cdots, N$ *such that* $u_N \in \mathcal{U}_N$. *Specifically, such* u_N *can be constructed as*

$$u_N(t) := u(t_{i-1}) + \frac{u(t_i) - u(t_{i-1})}{t_i - t_{i-1}}(t - t_{i-1}) \text{ for } t \in [t_{i-1}, t_i], \ i = 2, \cdots, N. \tag{3.2}$$

Proof. Observe that

$$\int_{t_1}^{t_N} \psi_j(t)du(t) = \sum_{i=2}^{N} \psi_j(t_i)(u(t_i) - u(t_{i-1})) = \int_{t_1}^{t_N} \psi_j(t)du_N(t), \tag{3.3}$$

where the last equality is due to the fact that integral only involves the values of u at the turning points. □

With $\mathcal{U}_N$, we propose an approximation of the PRO model by

$$\text{(PRO-}N\text{)} \qquad \vartheta_N := \max_{x\in X} \min_{u\in\mathcal{U}_N} \ \mathbb{E}_P[u(f(x,\xi))]. \tag{3.4}$$

The following theorem says that (PRO-N) coincides with (PRO) in concave utility function case.

Proposition 3.2 ((PRO) vs (PRO-N)) *If the utility functions in $\mathcal{U}$ are concave and $\psi_j(t)$, $j = 1,\cdots,m$, are step-like over $[a,b]$ with jumps at $t_i \in T$ for $i = 1,\cdots,N$, then $\vartheta_N = \vartheta$.*

Proof. Since $\mathcal{U}_N \subset \mathcal{U}$, then $\vartheta_N \geq \vartheta$, which means ϑ_N provides an upper bound for ϑ. It suffices to show that $\vartheta_N \leq \vartheta$. Let ϵ be a small positive number and $u_\epsilon^* \in \mathcal{U}$ be such that

$$\max_{x\in X} \mathbb{E}_P[u_\epsilon^*(f(x,\xi))] \leq \vartheta + \epsilon.$$

By Proposition 3.1, we can find a piecewise linear concave function $u_\epsilon^N \in \mathcal{U}_N$ such that $u_\epsilon^N(t) \leq u_\epsilon^*(t)$ for $t \in [a,b]$. Consequently we have

$$\vartheta_N \leq \max_{x\in X} \mathbb{E}_P[u_\epsilon^N(f(x,\xi))] \leq \max_{x\in X} \mathbb{E}_P[u_\epsilon^*(f(x,\xi))] \leq \vartheta + \epsilon.$$

The conclusion follows as ϵ can be arbitrarily small. □

Remark 3.1 *(i) Note that if the ambiguity set is constructed thorough pairwise comparison between two lotteries which follow discrete distributions, then function ψ_j is a step function. If, in addition, the utility functions in the ambiguity set are concave and the set of grid points T is selected from these discrete outcomes, then the conditions in Proposition 3.2 are satisfied, which means piecewise linear approximation does not incur any additional approximation errors. In general case without concavity, we may not be able to find a piecewise linear function $u_N \in \mathcal{U}_N$ such that $u_N \leq u^*$, in which case the (PRO) model is not equivalent to the (PRO-N) model.*

(ii) In the case when the lotteries in the pairwise comparison are continuously distributed, Hu and Mehrotra (2015) propose an approach which approximates ψ_j by a step function. The step-like approximation of ψ_j and piecewise linear approximation of u may have the same effect if they are set up properly. To see this, let u_N be a piecewise linear approximation of u with kinks $t_2,\cdots,t_{N-1}$. Then

$$\begin{aligned}
\int_{t_1}^{t_N} \psi_j(t)du_N(t) &= \sum_{i=2}^{N} \int_{t_{i-1}}^{t_i} \psi_j(t)du_N(t) = \sum_{i=2}^{N} \frac{u(t_i)-u(t_{i-1})}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} \psi_j(t_i)dt \\
&= \sum_{i=2}^{N} (u(t_i)-u(t_{i-1})) \frac{\int_{t_{i-1}}^{t_i} \psi_j(t_i)dt}{t_i - t_{i-1}} = \sum_{i=2}^{N} (u(t_i)-u(t_{i-1}))\psi_j(t_i'),
\end{aligned}$$

where $t'_i \in [t_{i-1}, t_i]$*. On the other hand, let* ψ_j^N *be a step-like approximation of* ψ_j *resulting from a discretization approach with grid points* $t_2, \cdots, t_{N-1}$ *such that* $\psi_j^N(t) = \psi_j(t'_i)$, for $t \in [t_{i-1}, t_i]$. *Then*

$$\int_{t_1}^{t_N} \psi_j^N(t)du(t) \;=\; \sum_{i=2}^{N} \int_{t_{i-1}}^{t_i} \psi_j^N(t)du(t) = \sum_{i=2}^{N} \int_{t_{i-1}}^{t_i} \psi_j(t'_i)du(t) = \sum_{i=2}^{N} \psi_j(t'_i)(u(t_i) - u(t_{i-1})).$$

(iii) When the utility functions are restricted to concave, Armbruster and Delage (2015); Hu and Mehrotra (2015) demonstrate that the PRO model can be reformulated as a linear programming problem. However, when the utility functions are general increasing function, the tractable reformulations do not exist. Indeed, the fundamental idea behind their tractable formulation scheme is that the worst-case utility function in $\mathcal{U}$ *is achieved by a piecewise linear concave function and this is not correct in general when the utility functions are merely increasing. In our approach, the worst-case utility function is chosen from all piecewise linear utility functions satisfying the moment-type conditions and the inner minimization problem of the (PRO-N) model is a linear programming problem regardless of whether the utility functions in* $\mathcal{U}_N$ *are concave or not. This is the main reason that we propose the piecewise linear approximation of the utility functions. Moreover, the piecewise linear approximation also facilitates us to establish some quantitative analysis results as opposed to qualitative convergence results in Hu and Mehrotra (2015).*

In the rest of the section, we discuss numerical schemes for solving the (PRO-N) model. We divide our discussion into two parts: increasing concave utility function case and increasing utility function case. To this end, we need to restrict our discussion to the case that ξ is discretely distributed.

Assumption 3.1 *P follows a discrete distribution with $P(\xi = \xi^k) = p_k$ for $k = 1, \cdots, K$.*

Under Assumption 3.1, we can rewrite the (PRO-N) model (3.4) as

$$\max_{x \in X} \min_{u \in \mathcal{U}_N} \sum_{k=1}^{K} p_k u(f(x, \xi^k)). \tag{3.5}$$

If ξ is continuously distributed, then we may regard (3.5) as an approximation to the (PRO-N) model. We will discuss convergence of the such approximation in Section 5.2.

3.2 Increasing concave utility function case

We start with the case that $\mathcal{U}$ is a class of increasing concave functions.

Assumption 3.2 *Each function $u \in \mathscr{U}$ is concave, where $\mathscr{U}$ is defined as in Definition 2.1.*

Recall that a concave function mapping from $\mathbb{R}^n$ to $\mathbb{R}$ can be represented as an upper envelope of linear functions and given a set of points in $\mathbb{R}^n \times \mathbb{R}$, the upper envelope is a piecewise linear concave function passing through some of the points. The following lemma gives a precise statement on this.

Lemma 3.1 *(Haskell et al., 2018, Theorem 4.2). Let $f : \mathbb{R}^n \to \mathbb{R}$. Then the following assertions hold.*

(i) f is concave if and only if

$$f(x) = \inf_{j\in\mathcal{J}} h_j(x), \forall x \in dom\, f, \tag{3.6}$$

where $\mathcal{J}$ is possibly infinite, $dom\, f$ denotes the effective domain of f and $h_j(x) := \langle a_j, x\rangle + b_j$ for all $j \in \mathcal{J}$.

(ii) For any finite set $\mathcal{O} \subset \mathbb{R}^n$ and values $\{v_s\}_{s\in\mathcal{O}} \subset \mathbb{R}$, the function $\hat{f} : \mathbb{R}^n \to \mathbb{R}$, defined by

$$\begin{aligned} \hat{f}(x) := \min_{a_0\geq 0, b_0} \quad & \langle a_0, x\rangle + b_0 \\ \text{s.t.} \quad & \langle a_0, s\rangle + b_0 \geq v_s, \forall s \in \mathcal{O}, \end{aligned}$$

is increasing and concave. Furthermore, $\hat{f} \leq \tilde{f}$ over $\mathbb{R}^n$ for all increasing and concave functions $\tilde{f}$ with $\tilde{f}(s) \geq v_s$.

Assumption 3.3 *Each function $u \in \mathscr{U}$ is Lipschitz continuous over $[a, b]$ with modulus being bounded by L.*

The assumption is needed in Section 4 when we derive the error bound of the ambiguity sets resulting from piecewise linear approximation. We present it here as some of the conditions should be included in the reformulation of the (PRO-N) model. The constant L gives an upper bound for the marginal utility value $u'(t)$ over $[a, b]$. The assumption is satisfied by many practically interesting utility functions such as dual-power functions $u(t) = 1 - (1 - t)^\nu$ defined over $[0, 1]$, where $\nu \geq 1$. The piecewise continuously differentiability also allows us to accommodate S-shaped utility function

$$u(t) = \begin{cases} (1 - e^{-\alpha t})/\alpha & \text{if } t \geq 0, \\ (\lambda * (e^{\beta t} - 1))/\beta & \text{otherwise}, \end{cases} \tag{3.7}$$

where $\lambda \geq 1$ and $\alpha, \beta \in (0, \infty)$ and $u(t)$ may be non-differentiable at 0.

Using Lemma 3.1, we can reformulate the maximin problem (PRO-N) as a single linear programming problem. The proposition below states this and its proof gives some details of derivation.

Proposition 3.3 *Under Assumptions 2.1-3.3, problem (3.5) can be reformulated as the following maximization problem which is a linear programming problem when $f(x, \xi)$ is affine in x:*

$$
\begin{aligned}
\max_{x\in X,\theta,v,\eta,\lambda,\mu} \quad & -\sum_{j=1}^{m}\lambda_j c_j+\theta_{N-1}+\sum_{k=1}^{K}\mu_{Nk}-L\sum_{j=1}^{N-1}\eta_j \\
\text{s.t.} \quad & p_k f(x,\xi^k)-\sum_{i=1}^{N}\mu_{ik}t_i\geq 0, k=1,\cdots,K, \\
& p_k-\sum_{i=1}^{N}\mu_{ik}=0, k=1,\cdots,K, \\
& \theta_i t_i-\theta_i t_{i+1}+v_{i-1}(t_i-t_{i-1})+\sum_{j=1}^{m}\lambda_j\int_{t_i}^{t_{i+1}}\psi_j(t)dt+\eta_i\geq 0, i=2,\cdots,N-2, \\
& \theta_1 t_1-\theta_1 t_2+\sum_{j=1}^{m}\lambda_j\int_{t_1}^{t_2}\psi_j(t)dt+\eta_1\geq 0, \\
& \theta_{N-1}t_{N-1}-\theta_{N-1}t_N+v_{N-2}(t_{N-1}-t_{N-2})+\sum_{j=1}^{m}\lambda_j\int_{t_{N-1}}^{t_N}\psi_j(t)dt+\eta_{N-1}\geq 0, \\
& \theta_{i-1}-\theta_i+\sum_{k=1}^{K}\mu_{ik}-v_{i-1}+v_i=0, i=2,\cdots,N-2, \\
& \theta_{N-2}-\theta_{N-1}+\sum_{k=1}^{K}\mu_{(N-1)k}-v_{N-2}=0, \\
& v_i\geq 0, i=1,\cdots,N-2, \\
& \eta_i\geq 0, i=1,\cdots,N-1, \\
& \lambda_j\geq 0, j=1,\cdots,m, \\
& \mu_{ik}\geq 0, i=1,\cdots,N, k=1,\cdots,K,
\end{aligned}
$$

where $\theta\in\mathbb{R}^{N-1}, v\in\mathbb{R}^{N-2}, \eta\in\mathbb{R}^{N-1}, \lambda\in\mathbb{R}^m$ *and* $\mu\in\mathbb{R}^{N\times K}$.

The proof is similar to those in Armbruster and Delage (2015) and Hu and Mehrotra (2015), we defer it to the appendix. Note that here the tractable reformulation does not require ψ_j to be step-like. Note also that no matter how large K is, the number of pieces of the optimal piecewise linear utility function in this formulation is limited to $N-1$, which means some (a_k,b_k), $k=1,\cdots,K$ may coincide. This is a departure from the support function approach in Armbruster and Delage (2015); Hu and Mehrotra (2015) and we believe it is an advantage of adopting the piecewise linear approximations.

3.3 Increasing utility function case

We now move on to the case that $\mathscr{U}$ is a class of increasing and Lipschitz continuous functions which are not necessarily concave. Our plan here is to use existing derivative-free algorithms such as COBYLA (Powell (1994)) in the open-source library and NLopt for nonlinear optimization (Johnson (2007)) for solving (PRO-N). To this effect, we discuss how to solve the inner minimization problem

$$u \in \arg\min_{u \in \mathcal{U}_N} \mathbb{E}_P[u(f(x,\xi))]. \tag{3.9}$$

Problem (3.9) is a convex programming problem in that $\mathcal{U}_N$ is a compact and convex set. Indeed, by writing each utility function $u \in \mathcal{U}_N$ as

$$u(t) = (a_1 t + b_1)\mathbb{1}_{[t_1,t_2]}(t) + \sum_{j=2}^{N-1}(a_j t + b_j)\mathbb{1}_{(t_j,t_{j+1}]}(t)$$

for $t \in [a,b]$, we may reformulate problem (3.9) as

$$\begin{aligned}
\arg \min_{(a_j,b_j),j=1,\cdots,N-1} \quad & \sum_{k=1}^{K} p_k\{(a_1 f(x,\xi^k) + b_1)\mathbb{1}_{[t_1,t_2]}(f(x,\xi^k)) \\
& + \sum_{j=2}^{N-1}(a_j f(x,\xi^k) + b_j)\mathbb{1}_{(t_j,t_{j+1}]}(f(x,\xi^k))\} \\
\text{s.t.} \quad & a_{j-1}t_j + b_{j-1} = a_j t_j + b_j, j = 2,\cdots,N-1, && (3.10a) \\
& a_1 t_1 + b_1 = 0, && (3.10b) \\
& a_{N-1}t_N + b_{N-1} = 1, && (3.10c) \\
& 0 \le a_j \le L, j = 1,\cdots,N-1, && (3.10d) \\
& \sum_{j=1}^{N-1} a_j \int_{t_j}^{t_{j+1}} \psi_i(t)dt \le c_i,\ i = 1,\cdots,m. && (3.10e)
\end{aligned}$$

Obviously the problem above is a linear programming problem in that the objective function is linear in a_j, b_j and all constraints are linear as well. Constraint (3.10a) requires the piecewise linear utility function to be continuous at the kinks, constraints (3.10b) and (3.10c) are normalization conditions, constraint (3.10d) imposes Lipschitz condition where L is a positive constant, and finally constraint (3.10e) is the condition from the ambiguity set.

From the above reformulation, piecewise linear approximation scheme enables us to evaluate the inner utility minimization problem when the utility functions in the ambiguity set are merely increasing. This allows us to use the derivative-free algorithms such as COBYLA in NLopt to solve the maximin problems.

4 Error bounds of the approximation

In this section, we investigate the difference between the (PRO) model and the (PRO-N) model in terms of their optimal values and optimal solutions in order to provide theoretical grounding for us to use the latter to approximate the former. Since the two maximin optimization problems are identical except for the ambiguity sets, it suffices to look into the difference between $\mathcal{U}_N$ and $\mathcal{U}$, and its impact on the optimal values and optimal solutions. To this end, we start by deriving a kind of Hoffman's lemma for the linear system in (2.3) which quantifies the deviation of any $u \in \mathscr{U}$ from set $\mathcal{U}$ by the residual of the linear system defining $\mathcal{U}$.

4.1 Pseudo metric

For each fixed $u \in \mathscr{U}$, we write integral $\int_a^b \psi_j(t)du(t)$ in a bilinear form $\langle \psi_j, u\rangle$. Let $\Psi := (\psi_1, \cdots, \psi_m)$, $C := (c_1, \cdots, c_m)$. Then we can rewrite (2.3) succinctly as

$$\mathcal{U} = \{u \in \mathscr{U} : \langle \Psi, u\rangle \leq C\}, \tag{4.1}$$

where the inner product of Ψ and u is calculated componentwise, i.e., $\langle \Psi, u\rangle = (\langle \psi_1, u\rangle, \cdots, \langle \psi_m, u\rangle)$. Note that the ambiguity set $\mathcal{U}$ is indeed the solution set of a system of linear inequalities defined in the functional space $\mathscr{U}$. In what follows, we introduce a kind of "distance" which measures the discrepancy of any two sets in $\mathscr{U}$. Let $\mathscr{G}$ be a set of measurable functions defined over $[a, b]$. For $u, v \in \mathscr{U}$, define the pseudo-metric between u and v by

$$\mathsf{dl}_{\mathscr{G}}(u, v) := \sup_{g\in\mathscr{G}} |\langle g, u\rangle - \langle g, v\rangle|.$$

It is easy to observe that $\mathsf{dl}_{\mathscr{G}}(u, v) = 0$ if and only if $\langle g, u\rangle = \langle g, v\rangle$ for all $g \in \mathscr{G}$. In practice, we may regard $\mathscr{G}$ as a set of "test functions" associated with some prospects and interpret u as a kind of utility measure, the pseudo-metric means that if u and v give the same average value for each $g \in \mathscr{G}$, then they are regarded as "equal" under $\mathsf{dl}_{\mathscr{G}}$ albeit they may not be identical. Thus $\mathsf{dl}_{\mathscr{G}}$ is a kind of pseudo-metric defined over the space of "utility measures" $\mathscr{U}$.

This definition is in parallel to a similar definition in probability theory, where u and v are in a position of probability measures and the corresponding pseudo-metric is known as ζ-metric, see Römisch (2003) for an overview. Here we continue to adopt the terminology although the background is different. Moreover, instead of using infinity norm or L_p norm in functional analysis straightforwardly, we adopt the pseudo-metric in that the information on the utility functions often comes out in a bilinear form as in (4.1). Note also that $\mathscr{G}$ must be restricted to ensure $\mathsf{dl}_{\mathscr{G}}(u, v)$ to be well defined, that is, $\mathsf{dl}_{\mathscr{G}}(u, v) < \infty$ for each pair of $u, v \in \mathscr{U}$. We state this in the following assumption.

Assumption 4.1 *The set $\mathscr{G}$ is chosen such that $\mathsf{dl}_{\mathscr{G}}(u, v)$ is finite-valued for any $u, v \in \mathscr{U}$.*

Note that if there exists a positive number θ such that $\sup_{g\in\mathscr{G}} \int_a^b |g(t)|dt \leq \theta$, then under Assumption 3.3, $\mathsf{dl}_{\mathscr{G}}(u, v) \leq 2L\theta$. Assumption 4.1 is also satisfied when $\mathscr{G}$ takes some specific structures as in the next example.

Example 4.1 *(a) Let $\mathscr{G} = \mathscr{G}_M := \left\{g : [a, b] \to \mathbb{R} | \ g \text{ is measurable, } \sup_{t\in[a,b]} |g(t)| \leq 1\right\}$. Then $\mathsf{dl}_{\mathscr{G}}(u, v)$ corresponds to the total variation metric in which case $\mathsf{dl}_{\mathscr{G}}(u, v) \leq 1$.*

(b) Let $\mathscr{G} = \mathscr{G}_L := \{g : [a, b] \to \mathbb{R} | \ g$ is Lipschitz continuous with modulus being bounded by 1$\}$. Then $\mathsf{dl}_{\mathscr{G}}(u, v)$ corresponds to the Kantorovich metric in which case we have

$$\mathsf{dl}_{\mathscr{G}}(u, v) = \int_{[a,b]} \int_{[a,b]} |t - t'| d\pi(t, t') \leq b - a, \tag{4.2}$$

where $\int_{[a,b]} \pi(t, t')dt' = u(t)$ and $\int_{[a,b]} \pi(t, t')dt = v(t')$.

(c) *Let* $\mathscr{G} := \mathscr{G}_L \cap \mathscr{G}_M$. *Then* $\mathsf{dl}_{\mathscr{G}}(u,v)$ *corresponds to the bounded Lipschitz metric in which case* $\mathsf{dl}_{\mathscr{G}}(u,v) \le \min(1, b-a)$.

(d) *Let* $\mathscr{G} = \mathscr{G}_I := \{g : g := \mathbb{1}_{[a,t]}(\cdot)\}$, *where* $\mathbb{1}_{[a,t]}(s) = 1$ *if* $s \in [a,t]$, *otherwise* $\mathbb{1}_{[a,t]}(s) = 0$. *Then* $\mathsf{dl}_{\mathscr{G}}(u,v)$ *corresponds to the uniform Kolmogorov metric in which case* $\mathsf{dl}_{\mathscr{G}}(u,v) \le 1$.

For any two sets $U, V \subset \mathscr{U}$, define $\mathbb{D}_{\mathscr{G}}(U,V) := \sup_{u\in U} \inf_{v \in V} \mathsf{dl}_{\mathscr{G}}(u,v)$, which quantifies the deviation of U from V and

$$\mathbb{H}_{\mathscr{G}}(U,V) := \max\left\{\mathbb{D}_{\mathscr{G}}(U,V), \mathbb{D}_{\mathscr{G}}(V,U)\right\}, \tag{4.3}$$

the Hausdorff distance between the two sets under the pseudo-metric. By convention, when $U = \{u\}$ is a singleton, we write the distance from u to set V as $\mathsf{dl}_{\mathscr{G}}(u,V)$ rather than $\mathbb{D}_{\mathscr{G}}(U,V)$.

4.2 Hoffman's lemma

Using the pseudo-metric, we are able to derive an error bound of any utility function $u \in \mathscr{U}$ deviating from $\mathcal{U}$ in terms of the residual of the linear system defining $\mathcal{U}$. This kind of result is known as Hoffman's lemma and our next result states the Hoffman's lemma in a specific infinite dimensional space of increasing utility functions.

Lemma 4.1 (Hoffman's lemma) *Consider (4.1). Suppose Assumption 4.1 holds and there exist a positive constant* α *and a function* $u_0 \in \mathcal{U}$ *such that*

$$\langle \Psi, u_0\rangle - C + \alpha\mathcal{B} \subset \mathbb{R}^m_-, \tag{4.4}$$

where $\mathcal{B}$ *denotes the unit ball centered at 0 in* $\mathbb{R}^m$ *and* $\mathbb{R}^m_-$ *denotes the negative orthant of* $\mathbb{R}^m$. *Then*

$$\mathsf{dl}_{\mathscr{G}}(u, \mathcal{U}) \le \frac{\mathsf{dl}_{\mathscr{G}}(u,u_0)}{\alpha}\|(\langle \Psi, u\rangle - C)_+\|, \forall u \in \mathscr{U}, \tag{4.5}$$

where $(a)_+ = \max(0,a)$ *and the maximum is taken componentwise.*

Condition (4.4) is known as the Slater's condition, it says that there is at least one utility function u_0 such that $\langle \Psi, u_0\rangle - C$ lies in the interior of $\mathbb{R}^m_-$. The Slater's condition is widely used in the literature of the Hoffman's lemma for linear and convex systems, see Robinson (1975) and references therein.

Proof of Lemma 4.1. Let $\rho := \|(\langle \Psi, u\rangle - C)_+\|$ and $\hat{u} := \left(1 - \frac{\rho}{\rho+\alpha}\right) u + \frac{\rho}{\rho+\alpha} u_0$. Since u and $u_0 \in \mathscr{U}$, then it is easy to verify that $\hat{u} \in \mathscr{U}$. Let $\mathbf{e} \in \mathbb{R}^m$ be a vector with each component being 1. Then $\langle \Psi, u\rangle - C \le \rho\mathbf{e}$ and hence

$$\begin{aligned}
\langle \Psi, \hat{u}\rangle - C &= \left(1 - \frac{\rho}{\rho+\alpha}\right)(\langle \Psi, u\rangle - C) + \frac{\rho}{\rho+\alpha}(\langle \Psi, u_0\rangle - C) \\
&\le \left(1 - \frac{\rho}{\rho+\alpha}\right)\rho\mathbf{e} - \frac{\rho}{\rho+\alpha}\alpha\mathbf{e} = 0.
\end{aligned}$$

This shows $\hat{u} \in \mathcal{U}$. Thus

$$\begin{aligned}
\mathsf{dl}_{\mathscr{G}}(u, \mathcal{U}) &\leq \mathsf{dl}_{\mathscr{G}}(u, \hat{u}) = \sup_{g \in \mathscr{G}} \left\{ \left| \langle g, u \rangle - \left\langle g, \left(1 - \frac{\rho}{\rho + \alpha}\right) u + \frac{\rho}{\rho + \alpha} u_0 \right\rangle \right| \right\} \\
&= \frac{\rho}{\rho + \alpha} \sup_{g \in \mathscr{G}} |\langle g, u \rangle - \langle g, u_0 \rangle| \leq \frac{\mathsf{dl}_{\mathscr{G}}(u, u_0)}{\alpha} \|(\langle \Psi, u \rangle - C)_+\|, \qquad (4.6)
\end{aligned}$$

which is (4.5). $\square$

From the proof of the Hoffman's lemma, we can see that the error bound holds for any $\mathscr{G}$ so long as $\mathsf{dl}_{\mathscr{G}}(u, u_0)$ is bounded. The usefulness of inequality (4.5) lies in the fact that it gives an upper bound for the deviation of the utility function $u \in \mathscr{U}$ from the ambiguity set $\mathcal{U}$ and the bound can be easily computed when $\mathcal{G}$ takes special structures as in Example 4.1. Note also that the residual term $\|(\langle \Psi, u \rangle - C)_+\|$ depends on the product value $\langle \Psi, u \rangle$ rather than u alone. In other words, if u and v give the same product value, then they have the same residual.

Note also that the Hoffman's lemma is established without exploiting conditions that all utility functions are defined over an interval $[a, b]$ and their ranges are scaled to $[0, 1]$. This means (4.5) holds when the utility functions are defined over $\mathbb{R}$ and they are not normalized provided that $\mathsf{dl}_{\mathscr{G}}(u, u_0)$ is bounded. We will need this fact in Section 6 where we extend our discussion to the case when u is defined over $\mathbb{R}$.

4.3 Error bound on the ambiguity set

We now move on to quantify the difference between $\mathcal{U}$ and $\mathcal{U}_N$. We need the following technical result.

Proposition 4.1 *Let $u \in \mathscr{U}$. Assume that $u(\cdot)$ is Lipschitz continuous over an interval $[a, b]$ with modulus L and u_N is its piecewise linear approximation defined as in (3.2). Then*

$$\mathsf{dl}_{\mathscr{G}_I}(u, u_N) = \sup_{t \in [a,b]} |u_N(t) - u(t)| \leq L\beta_N, \qquad (4.7)$$

where $\mathscr{G}_I$ is defined as in Example 4.1 (d), $\beta_N := \max_{i=2,\cdots,N}(t_i - t_{i-1})$ and $L \geq 1/(b - a)$.

The result is easy to establish, we include a proof in the appendix. With Lemma 4.1 and Proposition 4.1, we are ready to quantify the difference between $\mathcal{U}_N$ and $\mathcal{U}$.

Theorem 4.1 (Error bound on $\mathbb{H}_{\mathscr{G}}(\mathcal{U}_N, \mathcal{U})$) *Assume: (a) Assumption 3.3 holds; (b) the Slater's condition (4.4) is fulfilled with some $u_0 \in \mathcal{U}$, (c) $\psi_j, j = 1, \cdots, m$ are continuously differentiable except at a finite number of points. Then there exist positive constants $\hat{\alpha} < \alpha$ and N_0 such that the following assertions hold for two specific structures $\mathscr{G}$ defined in Example 4.1.*

(i) If $\mathscr{G} = \mathscr{G}_L$, then

$$\mathbb{H}_{\mathscr{G}_L}(\mathcal{U}_N, \mathcal{U}) \leq \left(\frac{2}{L} + \frac{b - a}{\hat{\alpha}} \left[\sum_{j=1}^{m} |\psi_j(b) - \psi_j(a)|^2 \right]^{1/2} \right) L\beta_N \qquad (4.8)$$

for all $N \geq N_0$;

(ii) If $\mathscr{G} = \mathscr{G}_I$, then

$$\mathbb{H}_{\mathscr{G}_I}(\mathcal{U}_N, \mathcal{U}) \le \left(1 + \frac{1}{\hat{\alpha}}\left[\sum_{j=1}^{m} |\psi_j(b) - \psi_j(a)|^2\right]^{1/2}\right) L\beta_N \tag{4.9}$$

for all $N \ge N_0$, where β_N is defined as in Proposition 4.1 with $\beta_N \to 0$ as $N \to \infty$.

Proof. Let $\hat{\alpha} < \alpha$ be a positive number. Under the Slater's condition (4.4), there exist a function $u_0^N \in \mathcal{U}_N$ and a positive number N_0 such that

$$\langle \Psi, u_0^N \rangle - C + \hat{\alpha}\mathcal{B} \subset \mathbb{R}^m_- \tag{4.10}$$

for $N \ge N_0$. The existence follows from Proposition 4.1 in that we can always construct a piecewise linear approximate utility function u_0^N to u_0 such that $u_0^N \to u_0$ as $\beta_N \to 0$. By applying Lemma 4.1 to $\mathcal{U}_N$ under the Slater's condition (4.10) for $\tilde{u} \in \mathscr{U}_N$,

$$\mathbb{D}_{\mathscr{G}}(\tilde{u}, \mathcal{U}_N) \le \frac{\mathsf{dl}_{\mathscr{G}}(\tilde{u}, u_0^N)}{\hat{\alpha}} \|(\langle \Psi, \tilde{u}\rangle - C)_+\| \tag{4.11}$$

for all $N \ge N_0$.

Let $u \in \mathcal{U}$ and u_N be defined as in (3.2). Then

$$\begin{aligned}
\|\langle \Psi, u_N\rangle - \langle \Psi, u\rangle\|^2 &= \sum_{j=1}^{m}\left|\int_a^b \psi_j(t)du(t) - \int_a^b \psi_j(t)du_N(t)\right|^2 \\
&= \sum_{j=1}^{m}\left|\int_a^b u(t)d\psi_j(t) - \int_a^b u_N(t)d\psi_j(t)\right|^2 \\
&\le \sum_{j=1}^{m}\left|\int_a^b |u(t) - u_N(t)|d\psi_j(t)\right|^2 \\
&\le L^2\beta_N^2 \sum_{j=1}^{m} |\psi_j(b) - \psi_j(a)|^2, \qquad (4.12)
\end{aligned}$$

where the second equality follows from integration in parts under condition (c). By the triangle inequality for the Pseudo metric and inequality (4.11), we have

$$\begin{aligned}
\mathsf{dl}_{\mathscr{G}}(u, \mathcal{U}_N) &\le \mathsf{dl}_{\mathscr{G}}(u, u_N) + \mathsf{dl}_{\mathscr{G}}(u_N, \mathcal{U}_N) \\
&\le \mathsf{dl}_{\mathscr{G}}(u, u_N) + \frac{\mathsf{dl}_{\mathscr{G}}(u_N, u_0^N)}{\hat{\alpha}}\|(\langle \Psi, u_N\rangle - C)_+\| \\
&= \mathsf{dl}_{\mathscr{G}}(u, u_N) + \frac{\mathsf{dl}_{\mathscr{G}}(u_N, u_0^N)}{\hat{\alpha}}[\|(\langle \Psi, u_N\rangle - C)_+\| - \|(\langle \Psi, u\rangle - C)_+\|] \\
&\le \mathsf{dl}_{\mathscr{G}}(u, u_N) + \frac{\mathsf{dl}_{\mathscr{G}}(u_N, u_0^N)}{\hat{\alpha}}\|\langle \Psi, u_N\rangle - \langle \Psi, u\rangle\| \\
&\le \mathsf{dl}_{\mathscr{G}}(u, u_N) + \frac{\mathsf{dl}_{\mathscr{G}}(u_N, u_0^N)}{\hat{\alpha}}\left(L\beta_N\left[\sum_{j=1}^{m}|\psi_j(b) - \psi_j(a)|^2\right]^{1/2}\right), \qquad (4.13)
\end{aligned}$$

where the equality holds due to $u \in \mathcal{U}$, i.e.$(\langle \Psi, u\rangle - C)_+ = 0$ and the last inequality comes from (4.12). In what follows, we turn to estimate $\mathsf{dl}_{\mathscr{G}}(u, u_N)$ and $\mathsf{dl}_{\mathscr{G}}(u_N, u_0^N)$ when $\mathscr{G}$ takes a specific form.

Part (i). If $\mathscr{G} = \mathscr{G}_L$, then

$$\begin{aligned}
\mathsf{dl}_{\mathscr{G}_L}(u, u_N) &= \sup_{g \in \mathscr{G}_L} \left| \int_a^b g(t) du(t) - \int_a^b g(t) du_N(t) \right| \\
&\le \sum_{j=2}^{N} \sup_{g \in \mathscr{G}_L} \left| \int_{t_{i-1}}^{t_i} g(t) du(t) - \int_{t_{i-1}}^{t_i} g(t) du_N(t) \right| \\
&= \sum_{j=2}^{N} \sup_{g \in \mathscr{G}_L} \left| \int_{t_{i-1}}^{t_i} g(t) du(t) - \int_{t_{i-1}}^{t_i} g(t_{i-1}) du(t) + \int_{t_{i-1}}^{t_i} g(t_{i-1}) du_N(t) - \int_{t_{i-1}}^{t_i} g(t) du_N(t) \right| \\
&\le \sum_{j=2}^{N} \sup_{g \in \mathscr{G}_L} \left(\int_{t_{i-1}}^{t_i} |g(t) - g(t_{i-1})| du(t) + \int_{t_{i-1}}^{t_i} |g(t_{i-1}) - g(t)| du_N(t) \right) \\
&\le \beta_N \sum_{j=2}^{N} \left(\int_{t_{i-1}}^{t_i} 1 du(t) + \int_{t_{i-1}}^{t_i} 1 du_N(t) \right) = 2\beta_N, \qquad (4.14)
\end{aligned}$$

where the second equality holds because $\int_{t_{i-1}}^{t_i} g(t_{i-1}) du(t) = \int_{t_{i-1}}^{t_i} g(t_{i-1}) du_N(t) = u(t_i) - u(t_{i-1})$. Together with $\mathsf{dl}_{\mathscr{G}_L}(u_N, u_0^N) \le b - a$ (see (4.2)), we may take supremum w.r.t. u over $\mathcal{U}$ on both sides of the inequality (4.13) to obtain

$$\mathbb{D}_{\mathscr{G}}(\mathcal{U}, \mathcal{U}_N) \le \left(\frac{2}{L} + \frac{b-a}{\hat{\alpha}} \left[\sum_{j=1}^{m} |\psi_j(b) - \psi_j(a)|^2 \right]^{1/2} \right) L\beta_N,$$

and hence the inequality (4.8) holds because $\mathbb{D}_{\mathscr{G}}(\mathcal{U}_N, \mathcal{U}) = 0$.

Part (ii). If $\mathscr{G} = \mathscr{G}_I$, then $\mathsf{dl}_{\mathscr{G}_I}(u_N, u_0^N) \le 1$ (see Example 4.1 (d)) and by Proposition 4.1

$$\mathsf{dl}_{\mathscr{G}_I}(u, u_N) = \sup_{t \in [a,b]} |u(t) - u_N(t)| \le L\beta_N.$$

Following a similar analysis to Part (i), we can obtain (4.9). We omit the details. $\square$

Equation (4.10) is the Slater's condition for the moment-type system when the utility function is restricted to space $\mathscr{U}_N$. We consider two cases as to how $\hat{\alpha}$ may be identified.

Case 1. **α is known.** We can choose any $\hat{\alpha}$ from interval $(0, \alpha)$. However, if $\hat{\alpha}$ is close to α, then we will need a large N in order for (4.10) to hold. On the other hand, if $\hat{\alpha}$ is chosen close to 0, then it will increase the error bound in (4.8) or (4.9). A reasonable choice would be half or two-thirds of α value. To quantify the relationship between $\hat{\alpha}$ and β_N (and hence N) more precisely, let u_0^N be the piecewise linear approximation of u_0 (defined in the Slater's condition (4.4)). Then

$$\begin{aligned}
\|\langle \Psi, u_0\rangle - \langle \Psi, u_0^N\rangle\|_\infty &= \max_{j=1,\cdots,m} \left| \int_a^b \psi_j(t) du_0(t) - \int_a^b \psi_j(t) du_0^N(t) \right| \\
&\le \max_{j=1,\cdots,m} L\beta_N |\psi_j(b) - \psi_j(a)| =: \gamma_N,
\end{aligned}$$

where the inequality follows from integration in parts (similar to the derivation of (4.12)). Let β_N be such that $\gamma_N \leq \alpha - \hat{\alpha}$, or equivalently,

$$\beta_N \leq \frac{\alpha - \hat{\alpha}}{\left[\max_{j=1,\cdots,m} |\psi_j(b) - \psi_j(a)|\right] L}. \tag{4.15}$$

By the Slater's condition (4.4), we have

$$\langle \Psi, u_0^N \rangle - C + \hat{\alpha}\mathcal{B} \subset \langle \Psi, u_0 \rangle - C + \alpha\mathcal{B} \subset \mathbb{R}^m_-. \tag{4.16}$$

Note that in this case $\hat{\alpha} < \alpha$.

Case 2. α **is not known.** We might find the largest $\hat{\alpha}$ satisfying (4.16) for fixed N by solving the following problem

$$\begin{array}{cl} \max\limits_{\alpha', u_N \in \mathscr{U}_N} & \alpha' \\ \text{s.t.} & \langle \psi_j, u_N \rangle + \alpha' \leq c_j, j = 1, \cdots, m. \end{array}$$

This is a linear programming problem given the piecewise linear structure of u_N. In that case, the α satisfying the Slater's condition (4.4) falls within the range $(0, \hat{\alpha} - \gamma_N]$ provided that $\hat{\alpha} > \gamma_N$ because

$$\langle \Psi, u_0 \rangle - C + \alpha\mathcal{B} \subset \langle \Psi, u_0^N \rangle - C + (\alpha + \gamma_N)\mathcal{B} \subset \langle \Psi, u_0^N \rangle - C + \hat{\alpha}\mathcal{B} \subset \mathbb{R}^m_-.$$

Note that in this case $\alpha < \hat{\alpha}$.

Note also that in Theorem 4.1, the error bounds in (4.8) and (4.9) comprise of two terms: the first term bounds the difference between u and u_N under the pseudo-metric $\mathsf{dl}_{\mathscr{G}}$ and the second term bounds the error arising from the moment system when u is approximated by u_N. The second term disappears in the case when ψ_j, $j = 1, \cdots, m$ are step functions. The following corollary states this.

Corollary 4.1 *Let Assumption 3.3 hold. If $\psi_j(t)$ is a step function over $[a, b]$ with jumps at t_i for $i = 1, \cdots, N$, then $\mathbb{H}_{\mathscr{G}_L}(\mathcal{U}_N, \mathcal{U}) \leq 2\beta_N$ and $\mathbb{H}_{\mathscr{G}_I}(\mathcal{U}_N, \mathcal{U}) \leq L\beta_N$.*

Proof. Since $\mathcal{U}_N \subset \mathcal{U}$, then $\mathbb{D}_{\mathscr{G}_L}(\mathcal{U}_N, \mathcal{U}) = 0$. Moreover, it follows from Proposition 3.1 that for any $u \in \mathcal{U}$, its piecewise linear approximation u_N constructed as in (3.2) belongs to $\mathcal{U}_N$. Thus for any $u \in \mathcal{U}$, we obtain from (4.14) that $\mathsf{dl}_{\mathscr{G}_L}(u, \mathcal{U}_N) \leq \mathsf{dl}_{\mathscr{G}_L}(u, u_N) \leq 2\beta_N$, which implies $\mathbb{D}_{\mathscr{G}_L}(\mathcal{U}, \mathcal{U}_N) \leq 2\beta_N$. If $\mathscr{G} = \mathscr{G}_I$, then $\mathsf{dl}_{\mathscr{G}_I}(u, \mathcal{U}_N) \leq \mathsf{dl}_{\mathscr{G}_I}(u, u_N) \leq L\beta_N$, where the last inequality follows from Proposition 4.1, and hence $\mathbb{D}_{\mathscr{G}_I}(\mathcal{U}, \mathcal{U}_N) \leq L\beta_N$. $\square$

The corollary provides us some useful insights: if ψ_j are step-like functions (which is the case in some pairwise comparisons), then we may choose $t_1, \cdots, t_N$ from those points where ψ_j jumps. In this way, we may effectively reduce the modelling error arising from piecewise linear approximation of the utility function. Note also that in this case, the Slater's condition is not needed which means the error bound holds for all N rather than for N sufficiently large.

4.4 Error bound on the optimal value and the optimal solution

We are now ready to quantify the difference between the (PRO-N) model and the (PRO) model. Let ϑ_N and ϑ denote the respective optimal values and X_N^* and X^* the corresponding sets of optimal solutions.

Theorem 4.2 (Error bound on optimal values and optimal solutions) *Let Assumption 3.3 hold and the Slater's condition (4.4) be fulfilled for $u_0 \in \mathcal{U}$. Then the following assertions hold.*

(i)

$$|\vartheta_N - \vartheta| \le \left(1 + \frac{1}{\hat{\alpha}}\left[\sum_{j=1}^m |\psi_j(b) - \psi_j(a)|^2\right]^{1/2}\right) L\beta_N \tag{4.17}$$

for all $N \ge N_0$, where L, $\hat{\alpha}$, β_N and N_0 are defined as in Theorem 4.1.

(ii) Let $v(z) := \min_{u\in\mathcal{U}} \mathbb{E}_P[u(f(x,\xi))]$. Define the growth function

$$\Lambda(\tau) := \min\{v(x) - \vartheta^* : d(x, X^*) \ge \tau, \forall x \in X\} \tag{4.18}$$

and set $\Lambda^{-1}(\eta) := \sup\{\tau : \Lambda(\tau) \le \eta\}$. Then

$$\mathbb{D}(X_N^*, X^*) \le \Lambda^{-1}\left(\left(1 + \frac{1}{\hat{\alpha}}\left[\sum_{j=1}^m |\psi_j(b) - \psi_j(a)|^2\right]^{1/2}\right) L\beta_N\right). \tag{4.19}$$

Proof. Part (i). It is well known that

$$|\vartheta_N - \vartheta| \le \max_{x\in X}\left|\min_{u\in\mathcal{U}_N} \mathbb{E}_P[u(f(x,\xi))] - \min_{u\in\mathcal{U}} \mathbb{E}_P[u(f(x,\xi))]\right|. \tag{4.20}$$

Let δ be a small positive number. For any $x \in X$, we can find $u^x \in \mathcal{U}$ and $u_N^x \in \mathcal{U}_N$ such that

$$\begin{aligned}
\mathbb{E}_P[u^x(f(x,\xi))] &\le \min_{u\in\mathcal{U}} \mathbb{E}_P[u(f(x,\xi))] + \delta,\\
\mathbb{E}_P[u_N^x(f(x,\xi))] &\ge \min_{u\in\mathcal{U}_N} \mathbb{E}_P[u(f(x,\xi))],\\
\sup_{t\in[a,b]} |u_N^x(t) - u^x(t)| &\le \mathbb{H}_{\mathscr{G}_I}(\mathcal{U}_N, \mathcal{U}) + \delta,
\end{aligned}$$

where the last inequality holds because $\mathsf{dl}_{\mathscr{G}_I}(u,v) = \sup_{t\in[a,b]} |u(t) - v(t)|$. Combining the above inequalities

$$\begin{aligned}
\min_{u\in\mathcal{U}_N} \mathbb{E}_P[u(f(x,\xi))] - \min_{u\in\mathcal{U}} \mathbb{E}_P[u(f(x,\xi))] &\le \mathbb{E}_P[u_N^x(f(x,\xi)) - u^x(f(x,\xi))] + \delta\\
&\le \sup_{t\in[a,b]} |u_N^x(t) - u^x(t)| + \delta\\
&\le \mathbb{H}_{\mathscr{G}_I}(\mathcal{U}_N, \mathcal{U}) + 2\delta.
\end{aligned}$$

By exchanging the position of $\mathcal{U}$ and $\mathcal{U}_N$, we have

$$\min_{u\in\mathcal{U}} \ \mathbb{E}_P[u(f(x,\xi))] - \min_{u\in\mathcal{U}_N} \ \mathbb{E}_P[u(f(x,\xi))] \leq \mathbb{H}_{\mathscr{G}_I}(\mathcal{U}_N,\mathcal{U}) + 2\delta.$$

Since $\delta > 0$ can be arbitrarily small, we obtain

$$|\vartheta_N - \vartheta| \ \leq \ \max_{x\in X}\left|\min_{u\in\mathcal{U}} \ \mathbb{E}_P[u(f(x,\xi))] - \min_{u\in\mathcal{U}_N} \ \mathbb{E}_P[u(f(x,\xi))]\right| \leq \mathbb{H}_{\mathscr{G}_I}(\mathcal{U}_N,\mathcal{U}) \tag{4.21}$$

and hence (4.17) follows from (4.9).

Part (ii). Observe first that $\Lambda(\cdot)$ is a non-decreasing function. Thus its generalized inverse is well defined. For any $x^*_N \in X^*_N$ and $x^* \in X^*$,

$$\begin{aligned}\Lambda(d(x^*_N, X^*)) \ &\leq \ v(X^*_N) - \vartheta^* = v(x^*_N) - v(x^*)\\ &\leq \ |v(x^*_N) - v_N(x^*_N)| + |v_N(x^*_N) - v(x^*)|\\ &\leq \ 2\max_{x\in X}|v(x) - v_M(x)|.\end{aligned}$$

Combining the inequality above with the second inequality of (4.17), we obtain

$$d(x^*_N, X^*) \leq \Lambda^{-1}(2\max_{x\in X}|v(x) - v_N(x)|) \leq \Lambda^{-1}(2\mathbb{H}_{\mathscr{G}_I}(\mathcal{U}_N,\mathcal{U}))$$

and hence (4.19). $\square$

Remark 4.1 *It might be helpful to make a few comments about the theorem.*

(i) In practice, ϑ is not computable whereas ϑ_N is. The error bound established in (4.17) gives the decision maker an interval centered at ϑ_N which contains ϑ. Equivalently, we can say that for a specified precision ϵ, we can use the inequality to estimate β_N such that $|\vartheta_N - \vartheta| \leq \epsilon$ and inequality (4.15) holds. In the case when $t_1, \cdots, t_N$ are evenly spread over $[a, b]$, we know the specified precision is reached when

$$N \geq \frac{L(b-a)}{\epsilon}\left(1 + \frac{1}{\hat{\alpha}}\left[\sum_{j=1}^{m}|\psi_j(b) - \psi_j(a)|^2\right]^{1/2}\right)$$

(ii) The error bound (4.17) is established without requiring the utility functions to be concave and it is derived under the piecewise approximation scheme. We envisage that similar results may be obtained using spline approximation and leave interested readers to investigate. Note that these are mesh-dependent approximation schemes which means that the quality of approximation depends on the number of kinks N. In numerical optimization, there are many mesh-free schemes for approximating a concave function or a quasi-concave function. Indeed, the upper envelope of a concave function by a class of linear functions is a simple example. More recently, Royset (2018) proposes to use DC-functions (difference of convex functions) to approximate upper semi-continuous functions where each convex function in the DC-functions is represented by maximum of a set of affine functions. The number of affine functions in the set affects the quality of approximation. These mesh-free approximation schemes are interesting but it seems difficult to derive an error bound.

(iii) Hu and Mehrotra (2015) consider a utility based PRO model where the ambiguity set is defined by moment-type conditions. Differing from our setting, Hu and Mehrotra (2015) assumes that ψ_j is step-like functions with jumps at grid points $t_j \in T$ and there are two increasing continuous functions which bound the true unknown utility function. In their tractable formulation of the PRO model, they approximate the upper and lower bound functions by piecewise linear functions with kinks at T. Under some moderate conditions, Hu and Mehrotra (2015) show that the optimal values obtained from solving approximated tractable PRO converges as the number of grid points increases ($\beta_N \to 0$). The results established here differ from Hu and Mehrotra's in two-fold. (a) Their convergence result is qualitative whereas our results are quantitative with explicit bounds. (b) Their convergence result is applicable to concave utility PRO models whereas our results are applicable to PRO models where the utility function is not necessarily concave. In Section 6, we will extend the PRO model and the error bound results to the case when the ambiguity set is dependent on decision variables.

Example 4.2 *Consider Example 2.1 where the ambiguity set is defined as in (2.6). Since B is preferred to A, there exist some $u_0 \in \mathscr{U}$ and a small positive number ϵ such that*

$$\int_a^b (F_B(t) - F_A(t))du_0(t) < -\epsilon.$$

Let $\alpha = -\epsilon - \int_a^b (F_B(t) - F_A(t))du_0(t) > 0$. Then the Slater's condition (4.4) is satisfied. Let $\hat{\alpha} \in (0, \alpha)$ be such that

$$\int_a^b (F_B(t) - F_A(t))du_0^N(t) + \hat{\alpha} \subset \mathbb{R}_-.$$

Observe that in this example, $\Psi(t) = F_B(t) - F_A(t)$ satisfying $|\Psi(t)| \leq 1$ for all $t \in [a, b]$. By Theorem 4.1 (ii) and Theorem 4.2,

$$|\vartheta - \vartheta_N| \leq \mathbb{H}_{\mathscr{G}_I}(\mathcal{U}_N, \mathcal{U}) \leq \left(1 + \frac{1}{\hat{\alpha}}\left[|\Psi(b) - \Psi(a)|^2\right]^{1/2}\right) L\beta_N$$

Moreover, if A and B follow discrete distributions, then Ψ is a step function. In that case, we may select the breakpoints in T (in the piecewise linear approximation) from the breakpoints of Ψ and subsequently it follows by Corollary 4.1 that

$$|\vartheta - \vartheta_N| \leq \mathbb{H}_{\mathscr{G}_I}(\mathcal{U}_N, \mathcal{U}) \leq L\beta_N.$$

5 Perturbation/contamination of elicitation data

In the previous section, we focus on quantification of the difference between $\mathcal{U}$ and $\mathcal{U}_N$, and its impact on the optimal values and optimal solutions. Now we change the angle slightly by looking into how perturbation of data in the definition of $\mathcal{U}$ affects the (PRO) model. Such perturbation may arise in the process of preference elicitation.

To explain this, let us re-visit Example 2.1. In practice, the true probability distribution of A and B may be unknown and consequently one may use sample average approximation to replace the expected utility as follows:

$$\frac{1}{N}\sum_{j=1}^{N} u(\xi_A^j) \geq \frac{1}{N}\sum_{j=1}^{N} u(\xi_B^j), \tag{5.1}$$

where $\xi_A^1, \cdots, \xi_A^N$ and $\xi_B^1, \cdots, \xi_B^N$ are iid samples. Two types of errors may be introduced when we use (5.1) to replace (2.6). One is the error arising from sample average approximation and the other is that the samples may be contaminated if they are collected from empirical data. The latter means the sample average does not converge to the true expected utility when the sample size goes to infinity. This kind of situation may occur in data-driven case studies. In the next subsection, we tackle the issues in one go, that is, stability analysis.

5.1 Perturbation of data in ambiguity set

Consider ambiguity set (4.1). Let $\tilde{\Psi}$ and $\tilde{C}$ be a perturbation of Ψ and C respectively. Let

$$\tilde{\mathcal{U}} = \{u \in \mathscr{U} : \langle \tilde{\Psi}, u\rangle \leq \tilde{C}\}. \tag{5.2}$$

In practice, $\tilde{\Psi}$ and $\tilde{C}$ represent perceived data which may contain noise whereas Ψ and C are real data which are not obtainable. They may also represent the data after discretization of continuous distributions of lotteries in pairwise comparisons. The next proposition quantifies the difference between $\tilde{\mathcal{U}}$ and $\mathcal{U}$ under metric $\mathsf{dl}_{\mathscr{G}}$.

Proposition 5.1 (Stability of the ambiguity set) *Let the Slater's condition (4.4) be fulfilled for $u_0 \in \mathcal{U}$ and $\tilde{u}_0 \in \tilde{\mathcal{U}}$ with a positive constant α. Let $\mathscr{G}$ be such that Assumption 4.1 holds. Then*

$$\mathbb{H}_{\mathscr{G}}(\tilde{\mathcal{U}}, \mathcal{U}) \leq \frac{\Delta}{\alpha}\left(\sup_{u\in\tilde{\mathcal{U}}\cup\mathcal{U}} \|\langle \Psi - \tilde{\Psi}, u\rangle\| + \|C - \tilde{C}\|\right), \tag{5.3}$$

where $\Delta := \sup_{u_1,u_2\in\tilde{\mathcal{U}}\cup\mathcal{U}} \mathsf{dl}_{\mathscr{G}}(u_1, u_2)$. In a particular case when $\mathscr{G} = \mathscr{G}_I$, Δ is bounded by 1.

Proof. Let $\tilde{u} \in \tilde{\mathcal{U}}$. By Lemma 4.1,

$$\begin{aligned}
\mathsf{dl}_{\mathscr{G}}(\tilde{u}, \mathcal{U}) &\leq \frac{\mathsf{dl}_{\mathscr{G}}(\tilde{u}, u_0)}{\alpha}\|(\langle\Psi, \tilde{u}\rangle - C)_+\| = \frac{\mathsf{dl}_{\mathscr{G}}(\tilde{u}, u_0)}{\alpha}(\|(\langle\Psi, \tilde{u}\rangle - C)_+ - (\langle\tilde{\Psi}, \tilde{u}\rangle - \tilde{C})_+\|) \\
&\leq \frac{\mathsf{dl}_{\mathscr{G}}(\tilde{u}, u_0)}{\alpha}(\|\langle\Psi - \tilde{\Psi}, \tilde{u}\rangle\| + \|C - \tilde{C}\|),
\end{aligned}$$

which entails

$$\mathbb{D}_{\mathscr{G}}(\tilde{\mathcal{U}}, \mathcal{U}) \leq \frac{\Delta}{\alpha}\left(\sup_{\tilde{u}\in\tilde{\mathcal{U}}} \|\langle\Psi - \tilde{\Psi}, \tilde{u}\rangle\| + \|C - \tilde{C}\|\right).$$

By swapping the positions between $\mathcal{U}$ and $\tilde{\mathcal{U}}$, we obtain (5.3). The bound for Δ has been discussed Example 4.1. $\square$

The term $\sup_{u\in\tilde{\mathcal{U}}\cup\mathcal{U}}\|\langle\Psi-\tilde{\Psi},u\rangle\|$ is not computable, we derive an upper bound for it. In the case when each $u\in\tilde{\mathcal{U}}\cup\mathcal{U}$ is Lipschitz continuous with modulus being bounded by L, the term is bounded by $L(\sum_{i=1}^m[\int_a^b|\psi_i(t)-\tilde{\psi}_i(t)|dt]^2)^{\frac{1}{2}}$.

With Proposition 5.1, we can easily establish stability of the (PRO) model. Let $\tilde{\vartheta}$ denote the optimal value of its perturbation, that is,

$$\text{(PRO-ptb)} \qquad \tilde{\vartheta}:=\max_{x\in X}\min_{u\in\tilde{\mathcal{U}}}\ \mathbb{E}_P[u(f(x,\xi))], \tag{5.4}$$

where $\tilde{\mathcal{U}}$ is defined as in (5.2). The following theorem says that the difference between ϑ and $\tilde{\vartheta}$ is linearly bounded by $\sup_{u\in\tilde{\mathcal{U}}\cup\mathcal{U}}\|\langle\Psi-\tilde{\Psi},u\rangle\|+\|C-\tilde{C}\|$.

Theorem 5.1 (Stability of the optimal value and the optimal solution) *Let the Slater's condition (4.4) be fulfilled for $u_0\in\mathcal{U}$ and $\tilde{u}_o\in\tilde{\mathcal{U}}$ with a positive constant α. Then*

$$|\tilde{\vartheta}-\vartheta|\le\frac{1}{\alpha}\left(\sup_{u\in\tilde{\mathcal{U}}\cup\mathcal{U}}\|\langle\Psi-\tilde{\Psi},u\rangle\|+\|C-\tilde{C}\|\right) \tag{5.5}$$

and

$$\mathbb{D}(\tilde{X},X^*)\le\Lambda^{-1}\left(\frac{1}{\alpha}\left(\sup_{u\in\tilde{\mathcal{U}}\cup\mathcal{U}}\|\langle\Psi-\tilde{\Psi},u\rangle\|+\|C-\tilde{C}\|\right)\right). \tag{5.6}$$

where $\tilde{X}$ denotes the set of optimal solutions of (PRO-ptb), X^* and Λ^{-1} are defined as in Theorem 4.2. In particular, if each $u\in\tilde{\mathcal{U}}\cup\mathcal{U}$ is Lipschitz continuous with modulus being bounded by L, then (5.5) reduces to

$$|\tilde{\vartheta}-\vartheta|\le\frac{1}{\alpha}\left(L\left(\sum_{i=1}^m\left[\int_a^b\left|\psi_i(t)-\tilde{\psi}_i(t)\right|dt\right]^2\right)^{\frac{1}{2}}+\|C-\tilde{C}\|\right) \tag{5.7}$$

and

$$\mathbb{D}(\tilde{X},X^*)\le\Lambda^{-1}\left(\frac{1}{\alpha}\left(L\left(\sum_{i=1}^m\left[\int_a^b\left|\psi_i(t)-\tilde{\psi}_i(t)\right|dt\right]^2\right)^{\frac{1}{2}}+\|C-\tilde{C}\|\right)\right). \tag{5.8}$$

Proof. Following a similar analysis to the proof of Theorem 4.2, we have

$$\left|\min_{u\in\tilde{\mathcal{U}}}\ \mathbb{E}_P[u(f(x,\xi))]-\min_{u\in\mathcal{U}}\ \mathbb{E}_P[u(f(x,\xi))]\right|\le\mathbb{H}_{\mathscr{G}_I}(\tilde{\mathcal{U}},\mathcal{U}),$$

which leads to (5.5) through a combination with Proposition 5.1. Moreover, under the Lipschitzness condition, we have

$$\sup_{u\in\tilde{\mathcal{U}}\cup\mathcal{U}}\|\langle\Psi-\tilde{\Psi},u\rangle\|\le L\left(\sum_{i=1}^m\left[\int_a^b\left|\psi_i(t)-\tilde{\psi}_i(t)\right|dt\right]^2\right)^{\frac{1}{2}},$$

which entails (5.7). The arguments for the optimal solutions are similar to that of Theorem 4.2, we omit the details. $\square$

If the perturbation arises from discretization of random variables in pairwise comparisons as we commented in Remark 3.1, i.e., sample average approximation, then $\psi_i(t)=F_A(t)-F_B(t)$ in the context of Example 2.1 and $\tilde{\psi}_i(t)$ is a step-like approximation.

5.2 Perturbation of exogenous uncertainty data

It might also be interesting to look into stability of the optimal value against small variation of probability distribution P either because the true probability distribution is not known and we have to approximate it with empirical data or because P is continuously distributed whereas our computational schemes in Section 3 work only for discretely distributed P (see Assumption 3.1). For this purpose, let us write $\vartheta(P)$ for the optimal value of the (PRO) model (2.2) and $\vartheta(Q)$ for the optimal value of the problem when P is perturbed to Q. The following theorem tells us that under some mild conditions, $\vartheta(Q)$ is close to $\vartheta(P)$ if Q is close to P.

Theorem 5.2 *Let $\mathscr{H} := \{h(\cdot) := u(f(x,\cdot)) : \Xi \to \mathbb{R} | u \in \mathcal{U}, x \in X\}$. Assume: (a) $\mathscr{H}$ is equi-continuous, i.e., for every $\epsilon > 0$, there exists a constant $\delta > 0$ such that $\sup_{h\in\mathscr{H}} |h(\xi')-h(\xi'')| < \epsilon$ for any ξ and ξ' with $\|\xi' - \xi''\| \le \delta$; (b) there exists an integrable gauge function $\phi : \Xi \to \mathbb{R}$ such that $|h(\xi)| \le \phi(\xi)$ for all $\xi \in \Xi$ and $h \in \mathscr{H}$. Then the following assertions hold.*

(i) For every sequence of probability distributions $\{Q_N\}$ converging to P under the topology of weak convergence with

$$\int_\Xi \phi(\xi) Q_N(d\xi) \to \int_\Xi \phi(\xi) P(d\xi), \tag{5.9}$$

we have

$$\lim_{N\to\infty} \vartheta(Q_N) = \vartheta(P). \tag{5.10}$$

(ii) Let $\xi^1, \cdots, \xi^N$ be independent random variables having identical distributions to that of ξ and $P_N(\cdot) := \frac{1}{N}\sum_{j=1}^N \mathbb{1}_{\xi^j}(\cdot)$, where $\mathbb{1}_{\xi^j}(\cdot) : \Xi \to \mathbb{R}$ is the indicator function at ξ^j. Then for every $\epsilon > 0$ there exist positive constants $\alpha(\epsilon)$, $\beta(\epsilon)$ and N_0 such that

$$\text{Prob}(|\vartheta(P_N) - \vartheta(P)| \ge \epsilon) \le \alpha(\epsilon) e^{-\beta(\epsilon)N}, \text{ for } N \ge N_0, \tag{5.11}$$

where the probability measure "Prob" is understood as the product of probability measure of P over the measurable space $\Xi \times \Xi \times \dots$ with product Borel sigma-algebra $\mathscr{B} \times \mathscr{B} \times \dots$.

Proof. Part (i). Under conditions (a) and (b), (Rao, 1962, Theorem 3.2) ensures that

$$\lim_{N\to\infty} \sup_{x\in X} \sup_{u\in\mathcal{U}} \left| \int_\Xi u\left(f(x,t)\right) Q_N(dt) - \int_\Xi u\left(f(x,t)\right) P(dt) \right| = 0. \tag{5.12}$$

Since

$$|\vartheta(Q_N) - \vartheta(P)| \le \sup_{x\in X} \sup_{u\in\mathcal{U}} \left| \int_\Xi u\left(f(x,t)\right) Q_N(dt) - \int_\Xi u\left(f(x,t)\right) P(dt) \right|$$

we immediately get (5.10).

Part (ii). It is well known that P_N converges to P under the topology of weak convergence at exponential rate with respect to increase of sample size N. Moreover, since ϕ is integrable,

$\int_\Xi \phi(\xi)P_N(d\xi) \to \int_\Xi \phi(\xi)P(d\xi)$. On the other hand, since Ξ is a compact set in a finite dimensional space, by (Huber and Ronchetti, 2009, Theorem 2.15), the space of probability measures on Ξ, denoted by $\mathscr{P}(\Xi)$, is a Polish space. By Part (i), $\vartheta(\cdot) : \mathscr{P}(\Xi) \to \mathbb{R}$ is continuous under the topology of ϕ-weak convergence [3]. The continuity of $\vartheta(\cdot)$ and the exponential rate of convergence of P_N to P entail (5.11). $\square$

Part (i) of Theorem 5.2 ensures the optimal value $\vartheta(P)$ is stable against a small perturbation of P. Part (ii) of the theorem paves the way for tractable formulations in Section 3 to be applied to the case when P is continuously distributed. Note that Delage et al. (2018a) derive a similar result to Theorem 5.2 (ii) when u is a loss function. Our proof is simpler by exploiting continuity of $\vartheta(\cdot)$ over appropriate subspace of $\mathscr{P}(\Xi)$. In the case when $\mathcal{U}$ is a singleton and the utility function is affine, the result coincides with standard stability result in stochastic programming. It might also be helpful to make some comments on the condition (a) of Theorem 5.2, i.e., the equi-continuity of $\mathscr{H}$. If each function $u \in \mathcal{U}$ is Lipschitz continuous over $[a, b]$ with modulus being bounded by L and $\{f(x, \cdot) : \Xi \to \mathbb{R} | x \in X\}$ is equi-coninuous, then $\mathscr{H}$ is equi-continuous. A sufficient condition for the equi-continuity of $\{f(x, \cdot) : \Xi \to \mathbb{R} | x \in X\}$ is that the function f is Lipschitz continuous over $X \times \Xi$.

6 Extensions

In the preceding discussions, the ambiguity set of utility functions is independent of decision variables and the domain of the utility functions are bounded. While the assumptions are widely used in the literature, they might undesirably limit the scope of application of the models. In this section, we discuss the possibility of lifting these assumptions.

6.1 Ambiguity set depending on the decision variables

In some practical cases, a decision maker's utility function may depend on decision variables. For instance, an investor might adopt different utility functions when he/she invests in different industries, i.e., traditional vs new and strategic. The investor would value more on market share and future prospects in a new high technology industry than short terms gains. In structural engineering, change of the values of some design variables may result in fundamental change of the functionality/reliability of an object and subsequently its utility rather than merely the overall cost. Similar problems exist in network decision where the decision on a link/edge may affect the overall utility of network. In the literature of distributionally robust optimization, there are a lot of discussions on how decisions affect the probability distribution of exogenous uncertainty (decision maker's belief), see for instance Noyan et al. (2019) and references therein. We leave this for exploration in a separate paper as it requires more detailed description of the problem.

[3] Let $\phi : \mathbb{R}^m \to [0, \infty)$ be a gauge function, that is, $\phi \geq 1$ holds outside a compact set. Define $\mathcal{C}_k^\phi$ the linear space of all continuous functions $h : \mathbb{R}^m \to \mathbb{R}$ for which there exists a positive constant c such that $|h(t)| \leq c(\phi(t) + 1), \forall t \in \mathbb{R}^m$. The ϕ-weak topology, denoted by τ_ϕ, is the coarsest topology on $\mathcal{M}_k^\phi$ for which the mapping $g_h : \mathcal{M}_k^\phi \to \mathbb{R}$ defined by $g_h(P') := \int_{\mathbb{R}^m} h(t)P'(dt),\ h \in \mathcal{C}_k^\phi$ is continuous. A sequence $\{P_l\} \subset \mathcal{M}_k^\phi$ is said to converge ϕ-weakly to $P \in \mathcal{M}_k^\phi$ if it converges w.r.t. τ_ϕ.

In some cases, a PRO model with expected utility constraints may be artificially reformulated as an unconstrained PRO model where the ambiguity depends on the decision variable. Consider the following inequality constrained expected utility maximization problem

$$\begin{array}{ll} \max\limits_{x\in X} & \mathbb{E}[u(f(x,\xi))] \\ \text{s.t.} & \mathbb{E}[u(g(x,\xi))] \geq c, \end{array} \tag{6.1}$$

where f and g are continuous functions and c is a constant. We may interpret f as the total return of a portfolio and g is an important part of it or vice versa.

Suppose now the true utility function is not known but it is possible to construct an ambiguity set $\mathcal{U}$ using partial information as we discussed before. Then we may consider the following maximin preference robust optimization problem

$$\begin{array}{lll} \max\limits_{x\in X} & \min\limits_{u\in\mathcal{U}} & \mathbb{E}[u(f(x,\xi))] \\ & \quad\text{s.t.} & \mathbb{E}[u(g(x,\xi))] \geq c. \end{array} \tag{6.2}$$

The robust model (6.2) means that for each fixed implementable decision x, we consider all potential utility functions from set $\mathcal{U}$ whose expected utility value of the current portfolio return outperforms that of the benchmark return and pick up the one which gives the worst expected utility value of $f(x,\xi)$. Let

$$\mathcal{U}(x) := \{u \in \mathcal{U} : \mathbb{E}[u(g(x,\xi))] \geq c\}.$$

We can write the problem (6.2) equivalently as

$$\max_{x\in X} \min_{u\in\mathcal{U}(x)} \quad \mathbb{E}[u(f(x,\xi))]. \tag{6.3}$$

In the latter formulation, the ambiguity set $\mathcal{U}(x)$ depends on x.

Proposition 6.1 *Let x^* denote the optimal solution of problem (6.3) and*

$$\tilde{X} := \inf_{u\in\mathcal{U}} \mathbb{E}[u(g(x,\xi))] - c \geq 0. \tag{6.4}$$

If $x^* \in \tilde{X}$, then problem (6.3) is equivalent to

$$\max_{x\in\tilde{X}} \min_{u\in\mathcal{U}} \mathbb{E}[u(f(x,\xi))]. \tag{6.5}$$

Proof. Let $v(x)$ and $\tilde{v}(x)$ denote respectively the optimal values of the inner minimization problem (6.3) and (6.5), let ϑ^* and $\tilde{v}$ denote respectively the optimal values of the outer maximization problems. Since $\mathcal{U}(x) \subset \mathcal{U}$, then $v(x) \geq \tilde{v}(x)$ for all $x \in X$. Moreover, since $\tilde{X} \subset X$, then

$$\vartheta^* = \max_{x\in X} v(x) \geq \max_{x\in\tilde{X}} \tilde{v}(x) = \tilde{\vartheta}.$$

Conversely for any $x \in \tilde{X}$, $\mathcal{U}(x) = \mathcal{U}$. Thus, the assumption that $x^* \in \tilde{X}$ implies $\mathcal{U}(x^*) = \mathcal{U}$ and subsequently $v(x^*) = \tilde{v}(x^*)$. This shows $\vartheta^* = v(x^*) = \tilde{v}(x^*) \leq \tilde{\vartheta}$ because $x^* \in \tilde{X}$. $\square$

Note that in general x^* may not satisfy (6.4). Based on the discussions above, we may extend the (PRO) model to the following

$$\text{(PRO-x)} \qquad \vartheta := \max_{x\in X} \min_{u\in\mathcal{U}(x)} \mathbb{E}_P[u(f(x,\xi))], \tag{6.6}$$

where

$$\mathcal{U}(x) := \left\{ u \in \mathscr{U} : \int_a^b \psi_j(x,t)du(t) \le c_j, \text{for } j=1,\cdots,m \right\}. \tag{6.7}$$

To solve (PRO-x), we may apply the piecewise linear approximation schemes proposed in Section 3 so that the inner utility minimization problem becomes a linear programming problem. We will resort to existing optimization solvers to solve the maximin problem. The error bounds derived in Section 4 can also be established under some slightly revised conditions. For instance, if there exists a positive constant α such that for each $x \in X$, there exists $u_0 \in \mathcal{U}(x)$ such that

$$\langle \Psi(x,\cdot), u_0\rangle - C + \alpha\mathcal{B} \subset \mathbb{R}^m_-,$$

then it follows from a similar analysis to that of Theorems 4.1 and 4.2, we can show that there exists $\hat{\alpha}$ such that

$$\mathbb{H}_{\mathscr{G}_I}(\mathcal{U}_N(x), \mathcal{U}(x)) \le \left(1 + \frac{1}{\hat{\alpha}}\left[\sum_{j=1}^m |\psi_j(x,b) - \psi_j(x,a)|^2\right]^{1/2}\right) L\beta_N$$

and consequently

$$\begin{aligned} |\vartheta_N - \vartheta| &\le \max_{x\in X}\left| \min_{u\in\mathcal{U}(x)} \mathbb{E}_P[u(f(x,\xi))] - \min_{u\in\mathcal{U}_N(x)} \mathbb{E}_P[u(f(x,\xi))]\right| \\ &\le \max_{x\in X} \mathbb{H}_{\mathscr{G}_I}(\mathcal{U}_N(x), \mathcal{U}(x)) \\ &\le \max_{x\in X}\left(1 + \frac{1}{\hat{\alpha}}\left[\sum_{j=1}^m |\psi_j(x,b) - \psi_j(x,a)|^2\right]^{1/2}\right) L\beta_N. \end{aligned}$$

We omit the details.

6.2 Utility function with unbounded domain

To motivate the discussion, consider a simple case that $f(x,\xi) = x\xi$, the restriction of the range of f to a bounded interval $[a,b]$ implies that the support of ξ is bounded. In some important applications such as finance and economics, the underlying random variables which represent market demand, stock price and rate of return are often described by distributions with unbounded support. This raises a question as to whether our proposed model and computational schemes in the previous sections can be effectively applied to situations where f is unbounded. Here we address the issue.

We start by defining a set of nonconstant increasing functions defined over $\mathbb{R}$ denoted by $\mathscr{U}_\infty$. Differing from the setup in Section 2, we require $u(0) = 0$ which may be reviewed as a normalization condition and 0 may be regarded as a reference point where u changes the shape such as S-shaped utility function. The condition also means $u(t) \leq 0$ for $t < 0$ and $u(t) \geq 0$ for $t > 0$. We no longer restrict the range of u to $[0,1]$. Next, we define the ambiguity set of utility functions

$$\mathcal{U}_\infty := \left\{ u \in \mathscr{U}_\infty : \int_{\mathbb{R}} \psi_j(t) du(t) \leq c_j, \ j = 1, \cdots, m \right\},$$

where $\psi_j : \mathbb{R} \to \mathbb{R}$, $j = 1, \cdots, m$ are a class of integrable functions. The preference robust optimization model based on $\mathcal{U}_\infty$ is defined as

$$(\mathrm{PRO})_\infty \qquad \vartheta_\infty := \max_{x \in X} \min_{u \in \mathcal{U}_\infty} \ \mathbb{E}_P[u(f(x,\xi))]. \tag{6.8}$$

Our aim is to solve the $(\mathrm{PRO})_\infty$ model and our concern is that the numerical schemes proposed in Section 3 cannot be applied to it directly. Let

$$\mathcal{U}_{[a,b]} := \left\{ u \in \mathscr{U}_{[a,b]} : \int_a^b \psi_j(t) du(t) \leq c_j, \ j = 1, \cdots, m \right\}, \tag{6.9}$$

where $\mathscr{U}_{[a,b]}$ denotes the set of nonconstant increasing functions defined over $[a,b]$. Let us compare $\mathcal{U}_\infty$ with $\mathcal{U}_{[a,b]}$. If we view ψ_j as information to be elicited for specifying the true unknown utility function, then the former utilizes all information of ψ_j over $\mathbb{R}$ whereas the latter only uses partial information truncated at a and b. In parallel to the $(\mathrm{PRO})_\infty$ model, we define a preference robust optimization model

$$(\mathrm{PRO})_{[a,b]} \qquad \vartheta_{[a,b]} := \max_{x \in X} \min_{u \in \mathcal{U}_{[a,b]}} \int_{f(x,\xi) \in [a,b]} u(f(x,\xi)) P(d\xi). \tag{6.10}$$

Obviously the $(\mathrm{PRO})_{[a,b]}$ model is identical to the PRO model (2.2) if $[a,b]$ covers the range of f. What we are interested here is the difference between the $(\mathrm{PRO})_\infty$ model and the $(\mathrm{PRO})_{[a,b]}$ model in terms of the optimal value. We will show that under some moderate conditions, the difference between $\vartheta_{[a,b]}$ and ϑ_∞ is small and therefore we may solve the $(\mathrm{PRO})_\infty$ model approximately by solving the $(\mathrm{PRO})_{[a,b]}$ model. The latter can be solved by the piecewise linear approximation scheme detailed in Section 3.

To build the bridge between $\mathcal{U}_\infty$ and $\mathcal{U}_{[a,b]}$, we define the following set:

$$\tilde{\mathcal{U}} := \left\{ u \in \mathscr{U}_\infty : \int_a^b \psi_j(t) du(t) \leq c_j, u(t) = u(a) \text{ for } t < a, u(t) = u(b) \text{ for } t > b, \ j = 1, \cdots, m \right\} \tag{6.11}$$

and establish the difference between $\tilde{\mathcal{U}}$ and $\mathcal{U}_\infty$ in the following proposition.

Proposition 6.2 *Assume that the system of inequalities in (6.11) satisfies the Slater's condition, i.e., there exist $\tilde{u}_0 \in \tilde{\mathcal{U}}$ and a positive constant α such that $\langle \Psi, \tilde{u}_0 \rangle - C + \alpha \mathcal{B} \subset \mathbb{R}^m_-$. Assume also that there exists a positive number θ such that*

$$\sup_{g \in \mathscr{G}, u \in \mathcal{U}_\infty} \int_{\mathbb{R}} |g(t)| du(t) \leq \theta. \tag{6.12}$$

Then for any $\epsilon > 0$ there exist constants $a < 0$ and $b > 0$ such that

$$\mathbb{H}_{\mathscr{G}}(\tilde{\mathcal{U}}, \mathcal{U}_\infty) \leq \frac{2\theta}{\alpha} \sup_{u \in \mathcal{U}_\infty} \left\| \int_{\mathbb{R} \setminus [a,b]} \Psi(t) du(t) \right\| + \epsilon. \tag{6.13}$$

Proof. Observe that $\tilde{\mathcal{U}} \subset \mathcal{U}_\infty$, i.e., $\mathbb{D}_{\mathscr{G}}(\tilde{\mathcal{U}}, \mathcal{U}_\infty) = 0$. It follows from the condition (6.12) that for any $\epsilon > 0$ there exist constants $a < 0$ and $b > 0$ such that

$$\sup_{g \in \mathscr{G}, u \in \mathcal{U}_\infty} \int_{\mathbb{R}\setminus[a,b]} |g(t)| du(t) \leq \epsilon.$$

For any fixed $u \in \mathcal{U}_\infty$, let $\tilde{u} = u(t)$ for $t \in [a,b]$ and $\tilde{u}(t) = u(a)$ for $t < a$ and $\tilde{u}(t) = u(b)$ for $t > b$. Then

$$\mathsf{dl}_{\mathscr{G}}(u, \tilde{u}) = \sup_{g \in \mathscr{G}} \left| \int_{\mathbb{R}\setminus[a,b]} g(t) du(t) \right| \leq \epsilon. \tag{6.14}$$

Let $\rho := \left\| \left(\int_a^b \Psi(t) du(t) - C \right)_+ \right\|$. Since u coincides with $\tilde{u}$ over $[a,b]$, then $\rho = \left\| \left(\int_a^b \Psi(t) d\tilde{u}(t) - C \right)_+ \right\|$. Let $\hat{u} := \left(1 - \frac{\rho}{\rho+\alpha}\right) \tilde{u} + \frac{\rho}{\rho+\alpha} u_0$. Similar to the proof of Lemma 4.1, we can show that $\hat{u} \in \tilde{\mathcal{U}}$ and

$$\begin{aligned}
\mathbb{D}_{\mathscr{G}}(\tilde{u}, \tilde{\mathcal{U}}) \leq \mathsf{dl}_{\mathscr{G}}(\tilde{u}, \hat{u}) &\leq \frac{\Delta}{\alpha} \left\| \left(\int_a^b \Psi(t) d\tilde{u}(t) - C \right)_+ \right\| \\
&= \frac{\Delta}{\alpha} \left\| \left(\int_a^b \Psi(t) du(t) - C \right)_+ - \left(\int_{\mathbb{R}} \Psi(t) du(t) - C \right)_+ \right\| \\
&\leq \frac{\Delta}{\alpha} \left\| \int_{\mathbb{R}\setminus[a,b]} \Psi(t) du(t) \right\|,
\end{aligned} \tag{6.15}$$

where $\Delta := \sup_{u \in \mathcal{U}_\infty, \tilde{u} \in \tilde{\mathcal{U}}} \sup_{g \in \mathscr{G}} |\langle g, u \rangle - \langle g, \tilde{u} \rangle| \leq 2\theta$. Combining (6.14) and (6.15), we obtain

$$\mathbb{D}_{\mathscr{G}}(u, \tilde{\mathcal{U}}) \leq \mathsf{dl}_{\mathscr{G}}(u, \tilde{u}) + \mathbb{D}_{\mathscr{G}}(\tilde{u}, \tilde{\mathcal{U}}) \leq \frac{\Delta}{\alpha} \left\| \int_{\mathbb{R}\setminus[a,b]} \Psi(t) du(t) \right\| + \epsilon.$$

By taking supremum w.r.t u over $\mathcal{U}_\infty$ on both sides of the inequality above, we obtain (6.13). $\square$

From Proposition 6.2, if $\sup_{u \in \mathcal{U}_\infty} \left\| \int_{\mathbb{R}\setminus[a,b]} \Psi(t) du(t) \right\|$ is small, then the difference between $\tilde{\mathcal{U}}$ and $\mathcal{U}_\infty$ will not be significant. To see when the former quantity will be small, we consider a simple case with $m = 1$, that is, $\Psi(t) = \psi_1(t)$, and u is differentiable. In that case we have

$$\int_{\mathbb{R}\setminus[a,b]} \Psi(t) du(t) = \int_{-\infty}^a \psi_1(t) u'(t) dt + \int_b^\infty \psi_1(t) u'(t) dt.$$

Thus if there is a non-negative function $g(t)$ such that $|\psi_1(t) u'(t)| \leq g(t)$ for all $t \in \mathbb{R}$ and $\int_{\mathbb{R}} g(t) dt < \infty$, then $\sup_{u \in \mathcal{U}_\infty} \left\| \int_{\mathbb{R}\setminus[a,b]} \Psi(t) du(t) \right\|$ will be small for $-a$ and b sufficiently large. Moreover, if we choose

$$\mathscr{G} = \mathscr{G}_I := \{g : g = \mathbb{1}_{[t,0]}(\cdot) \text{ for } t \in [a,0] \text{ and } g = \mathbb{1}_{[0,t]}(\cdot) \text{ for } t \in [0,b]\},$$

then for any given $u_1, u_2 \in \mathcal{U}_\infty$, by $u_1(0) = u_2(0) = 0$, we have

$$d_{\mathscr{G}_I}(u_1, u_2) = \sup_{g \in \mathscr{G}_I} |\langle g, u_1 \rangle - \langle g, u_2 \rangle| = \sup_{t \in [a,b]} |u_1(t) - u_2(t)|. \tag{6.16}$$

Moreover, condition (6.12) means

$$\sup_{t\in[a,b],u\in\mathcal{U}_\infty} |u(t)| \leq \theta. \tag{6.17}$$

Condition (6.17) is satisfied when $u \in \mathcal{U}_\infty$ is uniformly Lipschitz continuous over $[a, b]$.

We now turn to compare $\tilde{\mathcal{U}}$ with $\mathcal{U}_{[a,b]}$. Since for each $u \in \tilde{\mathcal{U}}$, $\int_{\mathbb{R}} \psi_j(t)du(t) = \int_a^b \psi_j(t)du(t)$, then the projection of u on $\mathscr{U}_{[a,b]}$ lies in $\mathcal{U}_{[a,b]}$. On the other hand, every u in $\mathcal{U}_{[a,b]}$ can be effectively extended to $\mathbb{R}$ so that the extended utility function falls in $\tilde{\mathcal{U}}$. The discussion means $\mathcal{U}_{[a,b]}$ is indeed the projection $\tilde{\mathcal{U}}$ over $\mathscr{U}_{[a,b]}$. By exploiting the relationship, we can quantify the difference between ϑ_∞ and $\vartheta_{[a,b]}$ in the following theorem.

Theorem 6.1 *Suppose that*

$$\sup_{u\in\mathcal{U}_\infty,x\in X} \int_{\mathbb{R}^m} |u(f(x,\xi))|P(d\xi) < \infty, \tag{6.18}$$

and the conditions in Proposition 6.2 are fulfilled. Then for any $\epsilon > 0$, there exist constants $a < 0$ and $b > 0$ such that

$$\left|\vartheta_\infty - \vartheta_{[a,b]}\right| \leq \frac{2\theta}{\alpha} \sup_{u\in\mathcal{U}_\infty} \left\| \int_{\mathbb{R}\setminus[a,b]} \Psi(t)du(t) \right\| + 2\epsilon, \tag{6.19}$$

where θ is a positive constant satisfying (6.17).

Proof. It follows from the condition (6.18) that for any $\epsilon > 0$ there exist constants $a < 0$ and $b > 0$ such that

$$\sup_{u\in\mathcal{U}_\infty,x\in X} \int_{f(x,\xi)\in\mathbb{R}\setminus[a,b]} |u(f(x,\xi))|P(d\xi) \leq \frac{\epsilon}{3}. \tag{6.20}$$

Since $\tilde{\mathcal{U}} \subset \mathcal{U}_\infty$, inequality (6.20) implies

$$\sup_{u\in\tilde{\mathcal{U}}} (|u(a)|P((-\infty, a)) + |u(b)|P((b, +\infty))) \leq \frac{\epsilon}{3}. \tag{6.21}$$

By the definition of ϑ_∞ and $\vartheta_{[a,b]}$

$$\begin{aligned} |\vartheta_\infty - \vartheta_{[a,b]}| &\leq \sup_{x\in X} \left| \inf_{u\in\mathcal{U}_\infty} \int_{\mathbb{R}^m} u(f(x,\xi))P(d\xi) - \inf_{\hat{u}\in\mathcal{U}_{[a,b]}} \int_{f(x,\xi)\in[a,b]} \hat{u}(f(x,\xi))P(d\xi) \right| \\ &\leq \sup_{x\in X} \left| \inf_{u\in\mathcal{U}_\infty} \int_{\mathbb{R}^m} u(f(x,\xi))P(d\xi) - \inf_{\tilde{u}\in\tilde{\mathcal{U}}} \int_{\mathbb{R}^m} \tilde{u}(f(x,\xi))P(d\xi) \right| \\ &\quad + \sup_{x\in X} \left| \inf_{\tilde{u}\in\tilde{\mathcal{U}}} \int_{\mathbb{R}^m} \tilde{u}(f(x,\xi))P(d\xi) - \inf_{\hat{u}\in\mathcal{U}_{[a,b]}} \int_{f(x,\xi)\in[a,b]} \hat{u}(f(x,\xi))P(d\xi) \right| \end{aligned} \tag{6.22}$$

Let us estimate the first term at the right hand side of the last inequality of (6.22). Observe that

$$
\begin{aligned}
& \inf_{u\in\mathcal{U}_\infty}\int_{\mathbb{R}^m} u(f(x,\xi))P(d\xi) - \inf_{\tilde{u}\in\tilde{\mathcal{U}}}\int_{\mathbb{R}^m} \tilde{u}(f(x,\xi))P(d\xi) \\
\leq\ & \inf_{u\in\mathcal{U}_\infty}\sup_{\tilde{u}\in\tilde{\mathcal{U}}}\left|\int_{\mathbb{R}^m} u(f(x,\xi))P(d\xi) - \int_{\mathbb{R}^m} \tilde{u}(f(x,\xi))P(d\xi)\right| \\
\leq\ & \inf_{u\in\mathcal{U}_\infty}\sup_{\tilde{u}\in\tilde{\mathcal{U}}}\left|\int_{f(x,\xi)\in[a,b]} u(f(x,\xi))P(d\xi) - \int_{f(x,\xi)\in[a,b]} \tilde{u}(f(x,\xi))P(d\xi)\right| + \frac{2\epsilon}{3} \\
\leq\ & \inf_{u\in\mathcal{U}_\infty}\sup_{\tilde{u}\in\tilde{\mathcal{U}}}\sup_{t\in[a,b]}|u(t)-\tilde{u}(t)| + \frac{2\epsilon}{3} \\
=\ & \inf_{u\in\mathcal{U}_\infty}\sup_{\tilde{u}\in\tilde{\mathcal{U}}} d_{\mathscr{G}_I}(u,\tilde{u}) + \frac{2\epsilon}{3}, \qquad (6.23)
\end{aligned}
$$

where the last inequality follows from (6.16). Thus

$$
\sup_{x\in X}\left|\inf_{u\in\mathcal{U}_\infty}\int_{\mathbb{R}^m} u(f(x,\xi))P(d\xi) - \inf_{\tilde{u}\in\tilde{\mathcal{U}}}\int_{\mathbb{R}^m} \tilde{u}(f(x,\xi))P(d\xi)\right| = \mathbb{H}_{\mathscr{G}_I}(u,\tilde{u}) + \frac{2\epsilon}{3}. \qquad (6.24)
$$

Let us now look into the second term. Let δ be a small positive number. For any $x\in X$, we can find $\hat{u}_x\in\mathcal{U}_{[a,b]}$ and its extended function $\tilde{u}_x\in\tilde{\mathcal{U}}$ such that

$$
\begin{aligned}
\int_{f(x,\xi)\in[a,b]} \hat{u}_x(f(x,\xi))P(d\xi) &\leq \inf_{\hat{u}\in\mathcal{U}_{[a,b]}}\int_{f(x,\xi)\in[a,b]} \hat{u}(f(x,\xi))P(d\xi) + \delta, \\
\int_{\mathbb{R}^m} \tilde{u}_x(f(x,\xi))P(d\xi) &\geq \inf_{\tilde{u}\in\tilde{\mathcal{U}}}\int_{\mathbb{R}^m} \tilde{u}(f(x,\xi))P(d\xi).
\end{aligned}
$$

Consequently we have

$$
\begin{aligned}
& \inf_{\tilde{u}\in\tilde{\mathcal{U}}}\int_{\mathbb{R}^m} \tilde{u}(f(x,\xi))P(d\xi) - \inf_{\hat{u}\in\mathcal{U}_{[a,b]}}\int_{f(x,\xi)\in[a,b]} \hat{u}(f(x,\xi))P(d\xi) \\
\leq\ & \int_{\mathbb{R}^m} \tilde{u}_x(f(x,\xi))P(d\xi) - \int_{f(x,\xi)\in[a,b]} \hat{u}_x(f(x,\xi))P(d\xi) + \delta \\
=\ & \int_{f\in\mathbb{R}\setminus[a,b]} \tilde{u}_x(f(x,\xi))P(d\xi) + \delta \\
\leq\ & \sup_{\tilde{u}\in\tilde{\mathcal{U}}}(|\tilde{u}(a)|P((-\infty,a)) + |\tilde{u}(b)|P((b,+\infty))) + \delta.
\end{aligned}
$$

By exchanging the position of $\tilde{\mathcal{U}}$ and $\mathcal{U}_{[a,b]}$, we have

$$
\begin{aligned}
& \inf_{\hat{u}\in\mathcal{U}_{[a,b]}}\int_{f(x,\xi)\in[a,b]} \hat{u}(f(x,\xi))P(d\xi) - \inf_{\tilde{u}\in\tilde{\mathcal{U}}}\int_{\mathbb{R}^m} \tilde{u}(f(x,\xi))P(d\xi) \\
\leq\ & \sup_{\tilde{u}\in\tilde{\mathcal{U}}}(|\tilde{u}(a)|P((-\infty,a)) + |\tilde{u}(b)|P((b,+\infty))) + \delta.
\end{aligned}
$$

Since $\delta>0$ can be arbitrarily small, we obtain

$$
\begin{aligned}
& \sup_{x\in X}\left|\inf_{\tilde{u}\in\tilde{\mathcal{U}}}\int_{\mathbb{R}^m} \tilde{u}(f(x,\xi))P(d\xi) - \inf_{\hat{u}\in\mathcal{U}_{[a,b]}}\int_{f(x,\xi)\in[a,b]} \hat{u}(f(x,\xi))P(d\xi)\right| \\
\leq\ & \sup_{u\in\tilde{\mathcal{U}}}(|u(a)|P((-\infty,a)) + |u(b)|P((b,+\infty))) \leq \frac{\epsilon}{3}. \qquad (6.25)
\end{aligned}
$$

Combining (6.22)-(6.25), we obtain (6.19) from (6.13). $\square$

Condition (6.18) plays an important role in the established theorem. It is a kind of uniform integrability condition which captures the relationship between tails of u and P. To see how the condition could be possibly fulfilled, we consider a situation where f satisfies some kind of growth condition, i.e., there exist an exponent $\gamma > 0$ and some locally bounded function $\eta : \mathbb{R}^n \to \mathbb{R}$ such that

$$|f(x,t)| \leq \eta(x)(\|t\|^\gamma + 1), \forall (x,t) \in \mathbb{R}^n \times \mathbb{R}^m. \tag{6.26}$$

Here $\eta(\cdot)$ is said to be a locally bounded function if the convergence of $\{x^N\}$ implies the boundedness of $\{\eta(x^N)\}$. It is easy to observe that the right hand side of (6.26) controls the growth of $|f(x,t)|$ w.r.t. variation of the second argument. Let $C := \sup_{x\in X} \eta(x)$. Since X is assumed to be a compact set throughout this paper, then $C < \infty$. We assume without loss of generality that $C \geq 1$. Let

$$\phi(t) := \max\left(\sup_{u\in\mathcal{U}_\infty} u(C(\|t\|^\gamma + 1)), \sup_{u\in\mathcal{U}_\infty} -u(-C(\|t\|^\gamma + 1))\right). \tag{6.27}$$

Then the growth condition and monotonic increasing property of $u(\cdot)$ imply that for all $(x,t) \in X \times \mathbb{R}^m$ and $u \in \mathcal{U}_\infty$,

$$\begin{aligned} u(f(x,t)) &\leq u\left(C(\|t\|^\gamma + 1)\right) \leq \phi(t) \text{ for } f(x,t) \geq 0, \\ -u(f(x,t)) &\leq -u\left(-C(\|t\|^\gamma + 1)\right) \leq \phi(t) \text{ for } f(x,t) < 0, \end{aligned}$$

which means

$$|u(f(x,t))| \leq \phi(t), \forall (x,t) \in X \times \mathbb{R}^m, u \in \mathcal{U}_\infty.$$

Thus condition (6.18) is subsumed by the following condition

$$\int_{\mathbb{R}^m} \phi(\xi)P(d\xi) < \infty. \tag{6.28}$$

Condition (6.28) means that ϕ is integrable, or the moment of ϕ is finite. Moreover, existence of a and b in (6.20) is equivalent to existence of some compact set Ξ_ϵ such that $\int_{\mathbb{R}^m\setminus\Xi_\epsilon} \phi(\xi)P(d\xi) \leq \frac{\epsilon}{3}$, which is related to the tightness, or roughly speaking the tail behaviour, of probability distribution P.

To illustrate the discussions above, let us consider the case that $f(x,\xi) = x\xi$ where $x \in X := [1,2]$ and $\xi : \mathbb{R} \to \mathbb{R}$ is a random variable, $\mathcal{U}_\infty = \{u\}$ is a singleton and u is a S-shaped utility function defined as in (3.7). Then

$$u(f(x,\xi)) = \begin{cases} (1 - e^{-\alpha x\xi})/\alpha & \text{if } \xi \geq 0, \\ \lambda * (e^{\beta x\xi} - 1)/\beta & \text{otherwise.} \end{cases}$$

Consequently we can set

$$\phi(t) = \begin{cases} (1 - e^{-2\alpha t})/\alpha & \text{if } t \geq 0, \\ \lambda * (e^{\beta t} - 1)/\beta & \text{otherwise.} \end{cases}$$

From (6.28), we can see that when α and β becomes larger, the requirement on the tails of P to ensure condition (6.28) becomes more demanding.

We can also relate condition (6.28) to condition (5.9) in the stability analysis (Theorem 5.2). That is, the ϕ function in (5.9) can be chosen as in (6.28). From (6.28), we can see that when α and β becomes larger, there could be more probability measures satisfying (5.9) and hence (5.10). This means the optimal value of the PRO model is more stable against perturbation of P when the decision maker is more conservative. We will give a full discussion on the issue and statistical analysis in a separate paper.

Finally, we note that despite the results are established for the utility functions defined over $\mathbb{R}$, they can easily be applied to the case when u is defined over $[0, \infty)$ or $(-\infty, 0)$ simply by setting $a = 0$ and $b = 0$ respectively.

7 Numerical tests

We have carried out some numerical experiments on the PRO models discussed in the proceeding sections with application to a portfolio optimization problem. In this section, we report the test results.

7.1 Setup

We gathered $K = 37$ monthly returns of $n = 8$ assets (CBOE 10 Year Treasury Note Yield Index (TNX), CBOE 30-Year Treasury Bond Yield Index (TYX), CBOE Gold Index (GOX), Dow Jones Industrial Average Index (DJI), iShares MSCI EAFE Index (EFA), NASDAQ Composite Index (IXIC), S&P 500 Index (GSPC), and Wilshire 5000 Total Market Index (W5000)) from 1/2009 to 1/2012. Let $\xi^k = (\xi_1^k, \cdots, \xi_n^k)$ denote the samples of the random vector of monthly returns with equal probabilities $p_k := 1/K$ for $k = 1, \cdots, K$. The investor plans to purchase the eight index funds and then his profit will be $x^T\xi$ in a month with decision vector $x \in X := \{x \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = 1\}$. The PRO model for the portfolio optimization problem is

$$\max_{x\in X} \min_{u\in\mathcal{U}} \quad \sum_{k=1}^{K} p_k u(x^T\xi^k). \tag{7.1}$$

We propose to solve problem (7.1) via (PRO-N) where $\mathcal{U}$ is replaced by $\mathcal{U}_N$. To examine the performance of (PRO-N), we carry out the tests with a specified true utility function and investigate how the optimal value and the worst case utility function converge as the information of the investor's utility preference increases. For the increasing concave case, we set $u^*(t) = 1-e^{-10t}$, and for the increasing case, we set $u^*(t) = (1-e^{-3t})/3$ for $t \geq 0$, $u^*(t) = (2*(e^{8t}-1))/8$ for $t < 0$. Although the investor is unaware that his preferences can be characterized as this function, we assume that his decision never contradicts with results suggested by such a function, see similar assumption in Armbruster and Delage (2015). Note that the monthly returns are located in the interval $[-0.5, 0.5]$, we may restrict $\mathscr{U}$ to be a set of all increasing (concave) utility functions mapping from $[-0.5, 0.5]$ to $[u^*(-0.5), u^*(0.5)]$, and $\mathscr{U}_N$ the corresponding set of piecewise linear approximation functions.

7.2 Design of the ambiguity set

As we discussed in Section 2, the ambiguity set of utility functions $\mathcal{U}$ is characterized by the information about the investor's preferences. Following the elicitation strategy proposed in Armbruster and Delage (2015), we ask the investor questionnaires through comparing a risky gamble with two outcomes and a certain outcome, denoted respectively by

$$Z_1 = \begin{cases} r_1, & \text{with probability} \quad 1-p, \\ r_3, & \text{with probability} \quad p, \end{cases} \qquad \text{and} \quad Z_2 = r_2.$$

In this way, each questionnaire can be described by four values $r_1 \leq r_2 \leq r_3$ and a probability p. If the utility function is normalized to satisfy $u(r_1) = 0$ and $u(r_3) = 1$, then this query will become whether $u(r_2) > p$ or not. Now the task turns to how to choose these four values. Here, we follow the *random relative utility split scheme* proposed in Armbruster and Delage (2015).

Step 0. Set $m := 0$.

Step 1. Choose r_1 and r_3 uniformly from $[-0.5, 0.5]$ and set $r_2 := \frac{r_1+r_3}{2}$.

Step 2. Let $I_1 := \min_{u\in\mathcal{U}_N} u(r_2)$ and $I_2 := \max_{u\in\mathcal{U}_N} u(r_2)$, let $p := \frac{I_1+I_2}{2}$. Take the increasing concave case for an example, I_1 is obtained by solving the following problem:

$$\begin{aligned} \min_{(a_j,b_j),j=1,\cdots,N-1} \quad & \sum_{k=1}^{K} p_k \left\{ (a_1 r_2 + b_1)\mathbb{1}_{[t_1,t_2]}(r_2) + \sum_{j=2}^{N-1} (a_j r_2 + b_j)\mathbb{1}_{(t_j,t_{j+1}]}(r_2) \right\} \\ \text{s.t.} \quad & a_{j-1}t_j + b_{j-1} = a_j t_j + b_j, j = 2, \cdots, N-1, && (7.2a) \\ & a_1 r_1 + b_1 = 0, && (7.2b) \\ & a_{N-1} r_3 + b_{N-1} = 1, && (7.2c) \\ & a_{j+1} - a_j \leq 0, j = 1, \cdots, N-2, && (7.2d) \\ & \sum_{j=1}^{N-1} a_j \int_{t_j}^{t_{j+1}} \psi_i(t)dt \leq 0, \ i = 1, \cdots, m. && (7.2e) \end{aligned}$$

Note that we drop constraint (3.10d) here because the utility function is concave and ψ_i is step-like in which case PLA does not incur any error, see Remark 3.1. They are needed with utility function is merely increasing.

Step 3. Let $m := m + 1$ and

$$\psi_m(t) := (1-p)I_{t\geq r_1}(t) + pI_{t\geq r_3}(t) - I_{t\geq r_2}(t).$$

If $(1-p)u^*(r_1) + pu^*(r_3) \leq u^*(r_2)$, or equivalently $\int_a^b -\psi_m(t)du^*(t) \leq 0$, then we regard that the decision maker prefers Z_2 to Z_1, in which case only those u whose value at r_2 falls in the interval $[p, I_2]$ will be considered and subsequently we add the constraint $\int_a^b -\psi_m(t)du_N(t) \leq 0$ to the ambiguity set $\mathcal{U}_N$. Otherwise, add $\int_a^b \psi_m(t)du_N(t) \leq 0$. Go to Step 1.

7.3 Numerical experiments

We divide the numerical tests into five parts depending on the nature of true utility function and the ambiguity set: (i) the utility function is increasing and concave, (ii) the utility function is increasing but not necessarily concave, (iii) the ambiguity set depends on the decision variables, (iv) perturbation of data in the ambiguity set and (v) perturbation of exogenous uncertainty. All of the tests are carried out in MATLAB R2015a and JuliaPro 1.0.3.2 installed on a PC (16.00 GB RAM, CPU 3.6 GHz) with Intel Core i7 processor.

From the computational schemes in Section 3, piecewise linear approximation scheme enables us to evaluate the inner utility minimization problem regardless of whether the utility functions in the ambiguity set are concave, or merely increasing or dependent on decision variables. Throughout the numerical tests, we use the derivative-free algorithm COBYLA (Powell (1994)) in the open-source library NLopt for nonlinear optimization (Johnson (2007)) to solve the maximin problems because it allows us to take the advantage of LP nature of the inner minimization problem.

(i) Increasing concave utility function. We look into convergence of the worst case utility function and the optimal value as the number of queries increases, see Figures 1 and 2. It can be seen that the worst case utility function gets closer and closer to the true utility function as more queries are asked. The objective values increase as the number of queries increases. This is due to the fact that the ambiguity set $\mathcal{U}_N$ becomes smaller as the number of queries increases, and consequently the set of feasible solutions of the inner minimization problem in the (PRO-N) model becomes smaller, and hence the objective values increase. It should be mentioned that in this set of tests, the grid points $\{t_1, \cdots, t_N\}$ are taken from the outcomes in the queries and endpoints, that is, r_1, r_2, r_3, -0.5 and 0.5. If there are m queries, then $N = 3m + 2$. Hence, as the number of queries increases, the number of discretization points also increases. This implies that ψ_i is a step function with jumps $t_1, \cdots, t_N$ for $i = 1, \cdots, m$.

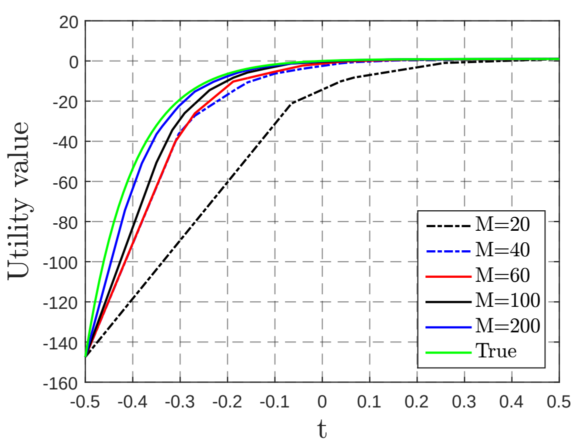

Figure 1: Increasing concave utility case: worst case utility functions

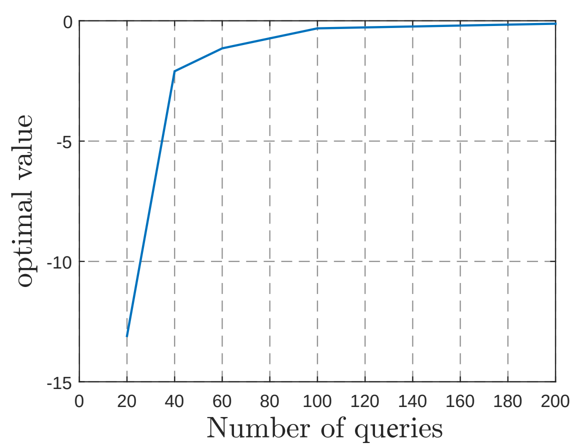

Figure 2: Increasing concave utility case: optimal values

(ii) Increasing utility function without concavity. As in the concave utility case, we also look into change of the worst case utility function and the optimal value as the number of queries increases. The results are displayed in Figures 3 and 4.

The worst case utility function converges to the true utility function as the number of queries increases. Since in this case the optimal value of (PRO-N) differs from that of (PRO), i.e., $\vartheta_N \neq \vartheta$, by Corollary 4.1 and Theorem 4.2, $\vartheta \in [\vartheta_N - L\beta_N, \vartheta_N + L\beta_N]$ with $L = 2$. Figure 4 depicts the intervals as N increases.

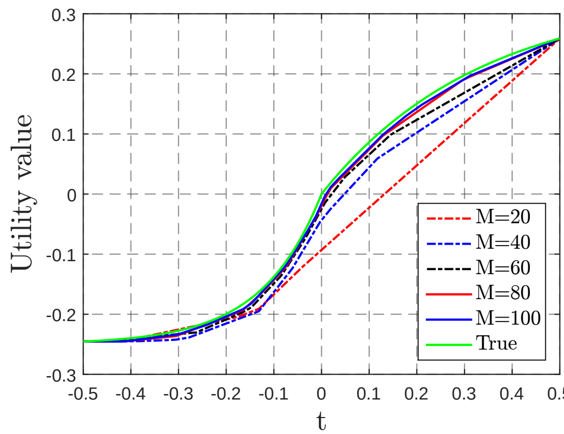

Figure 3: Increasing utility case: worst case utility functions

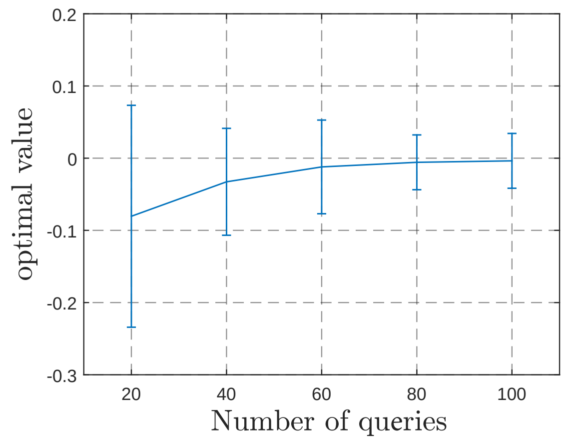

Figure 4: Increasing utility case: optimal values

(iii) Perturbation of data in the ambiguity set. In this set of experiments, we look into the change of the optimal value as the data in the ambiguity set is perturbed. Specifically we consider a situation where the underlying functions $\psi_j(t)$, $j = 1, \cdots, m$ in the ambiguity set is perturbed by small positive number, that is, $\tilde{\psi}_j(t) := \psi_j(t) - \delta_j$ where $\delta_j > 0$ for $j = 1, \cdots, m$. We use the same true utility function as that in Part (ii) of the experiments where the utility function is increasing but not necessarily concave. During the tests, the number of queries is fixed at $m = 80$. Let $\delta := (\sum_{j=1}^{m} \delta_j^2)^{1/2}$. Figure 5 depicts the change of the optimal values as δ increases from 0.01 to 1 (whereby we also require each δ_i to increase). We can see an decreasing tendency with the only exception at $\delta = 0.1$. This is because as δ_i increases, the feasible set of the inner minimization problem becomes larger and hence the optimal value of the minimization problem becomes smaller and the exception occurs when the maximin is not solved to optimality because this is a blackbox solver which goes beyond our control.

(iv) Perturbation of exogenous uncertainty. In this set of experiments, we examine change of the optimal value against small variation of samples Ξ. The setting of true utility function and the queries are the same as Part (iv). We consider a situation where Ξ is perturbed to $\tilde{\Xi} := \Xi + \delta B$ where the elements of matrix B are generated from uniform distribution over $[-1, 1]$ independently. For fixed B, we investigate change of the optimal value as δ increases. From Figure 6, we can see that as δ decreases from 0.1 to 0, the optimal value of the perturbed PRO model converges to the one without perturbation. The monotonic increasing trend may be interpreted by the fact that the derivative of the optimal value function in δ is positive near 0 and it may not be the case when Ξ changes along with δ.

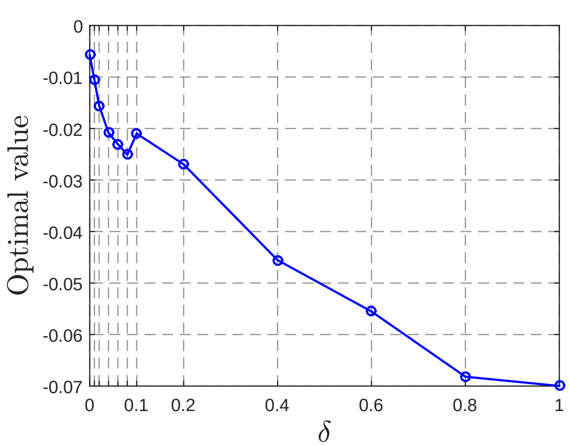

Figure 5: Perturbation of data in the ambiguity set

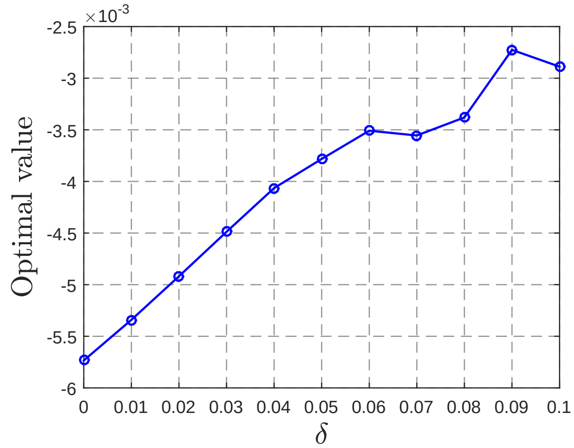

Figure 6: Perturbation of exogenous uncertainty

8 Conclusion

In this paper we re-visit utility preference robust optimization models with three novel extensions: (a) the utility function is merely increasing, (b) the domain of the utility function is unbouded and (c) the ambiguity set involves decision variables. Extension (a) allows one to accommodate a broader class of utility functions including S-shaped utility functions and more general quasiconcave utility functions, extension (b) effectively lifts implicit restriction on the boundedness of $f(x, \xi)$ that the existing PRO models impose and extension (c) enables us to encompass expected utility maximization problems with expected utility constraints.

We provide a unified approach to solve the subsequent PRO models, that is, piecewse linear approximation. We suggest a decision maker to solve (PRO-N) rather than (PRO) when the latter is difficult to solve. The error bound provides theoretical grounding for this, i.e., the difference between the optimal values of the two models can be controlled under some moderate conditions. In Figure 4, we show how the error bound may be effectively used to provide a guidance (in interval) for the optimal value of (PRO). Likewise, the stability results give decision makers a theoretical guarantee that the optimal value of (PRO) based on perceived data is close to the one with real data and this is particularly relevant in a data-driven environment.

In a particular case when the true utility function is concave and the ambiguity set elicited through pairwise questionnaires with finite scenario of outcomes, we show that PLA does not introduce any error, that is, it can produce the same results as those support function based methods in the literature for concave PRO models. The numerical test results confirm effectiveness of the extended models and PLA approach.

The PRO model in this paper is built on Von Neumann-Morgenstein expected utility theory that a decision maker's preference can be represented by his expected utility when the preference satisfies four axiomatic properties. However, there are extensive experimental evidences that some of these properties such as independence may fail to hold. Various alternative models are subsequently proposed including Yaari's dual theory of choice and Kahneman and Tversky's prospect theory. It might be interesting to investigate robust formulation of the latter models when the information of probaility weighting is incomplete. It will also be interesting to establish links between PRO models in utility to those in risk management, we leave these for future research.

References

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

D. B. Brown, E. De Giorgi and M. Sim. 2012. Aspirational preferences and their representation by risk measures. *Management Science*, 58(11): 2095–2113.

D. B. Brown and M. Sim. 2009. Satisficing measures for analysis of risk positions. *Management Science*, 55(1): 71–84.

R. T. Clemen and T. Reilly. 2001. *Making Hard Decisions with Decision Tools Suite*. Duxbury, Pacific Grove, CA, 2nd edition.

E. Delage, S. Guo and H. Xu. 2018a. Shortfall risk models when information of loss function is incomplete. Optimization-online, `http://www.optimization-online.org/DB_HTML/2018/04/6593.html`.

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

P. H. Farquhar. 1984. Utility assessment methods. *Management Science*, 30: 1283–1300.

H. Föllmer and A. Schied. 2002. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6: 429–447.

T. Galaabaatar and E. Karni. 2013. Subjective expected utility with incomplete preferences. *Ecomometrica*, 81: 155–284.

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

W. Haskell, W. Huang and H. Xu. 2018. Preference elicitation and robust optimization with multi-attribute Quasi-Concave choice functions. arXiv:1805.06632.

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

J. Hu and G. Stepanyan 2017. Optimization with reference-based robust preference constraints. *SIAM Journal on Optimization*, 27: 2230–2257.

J. Hu, M. Bansal and S. Mehrotra. 2018. Robust decision making using a general utility set. *European Journal of Operational Research*, 269(2): 699–714.

P. J. Huber and E. M. Ronchetti. 2009. *Robust Statistics*, John Wiley & Sons, New Jersey.

Steven G. Johnson. 2007. The NLopt nonlinear-optimization package, http://github.com/stevengj/nlopt.

U. S. Karmarkar. 1978. Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance*, 21: 61–72.

N. Noyan, G. Rudolf and M. Lejeune. 2019. Distributionally robust optimization under decision dependent ambiguity set with an application to machine scheduling. Optimization online, `http://www.optimization-online.org/DB_FILE/2020/01/7591.pdf`.

M. J. D. Powell. 1994. A direct search optimization method that models the objective and constraint functions by linear interpolation, in Advances in Optimization and Numerical Analysis, eds. S. Gomez and J.-P. Hennart (Kluwer Academic: Dordrecht), 51–67.

R. Ranga Rao. 1962. Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, 33: 659–680.

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

W. Römisch. 2003. Stability of stochastic programming problems. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, Handbooks in Operations Research and Management Science, volume 10, chapter 8. Elsevier, Amsterdam.

J. O. Royset. 2018. Approximations of semicontinuous functions with applications to stochastic optimization and statistical estimation. Arxiv: 1709.06730.

P. Vayanos, D. McElfresh, Y. Ye, J. Dickerson and E. Rice. 2020. Active preference elicitation via adjustable robust optimization. arXiv:2003.01899.

J. Von Neumann and O. Morgenstern. 1947. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton.

W. Wang and H. Xu. 2020. Robust spectral risk optimization when information on risk spectrum is incomplete. *SIAM Journal on Optimization*, 30: 3198–3229.

M. Weber. 1987. Decision making with incomplete information. *European Journal of Operational Research*, 28: 44–57.

Appendix A: Proofs.

Proof of Proposition 3.3. Let $z_i := u(t_i)$ and $I := \{1, \cdots, N\}$ with $z_1 = 0, z_N = 1$. Then

Problem (3.5) can be reformulated as

$$
\begin{aligned}
\max_{x\in X}\ \min_{a_k,b_k,z,\beta}\quad & \sum_{k=1}^{K} p_k(a_k f(x,\xi^k)+b_k) && (8.1a)\\
\text{s.t.}\quad & z_{i+1}-z_i=\beta_i(t_{i+1}-t_i), \forall i\in I\backslash\{N\}, && (8.1b)\\
& z_{i+1}-z_i\geq \beta_{i+1}(t_{i+1}-t_i), \forall i\in I\backslash\{N-1,N\}, && (8.1c)\\
& 0\leq \beta_i\leq L, \forall i\in I\backslash\{N\}, && (8.1d)\\
& \sum_{i=1}^{N-1}\beta_i\int_{t_i}^{t_{i+1}}\psi_j(t)dt\leq c_j,\ \forall j=1,\cdots,m, && (8.1e)\\
& a_k t_i+b_k\geq z_i, \forall i\in I, k=1,\cdots,K, && (8.1f)\\
& a_k\geq 0, k=1,\cdots,K, && (8.1g)\\
& z_1=0, z_N=1, && (8.1h)
\end{aligned}
$$

where conditions (8.1b), (8.1c) and (8.1d) characterize piecewise linearity, concavity, Lipschitz continuity and increasing property of the utility function, note that conditions (8.1b) and (8.1c) imply $\beta_i\geq\beta_{i+1}$. (8.1d) represents Lipschitz continuity of utility function where L is the positive constant defined as in Assumption 3.3. Condition (8.1e) represents the moment type conditions of the ambiguity. The representation of the objective function (8.1a) and conditions (8.1f) and (8.1g) are obtained based on the support function of increasing concave functions by virtue of Lemma 3.1. Using Lagrangian duality of the inner minimization problem, we obtain the equivalence of problem (8.1) and problem (3.8), where θ, v, η, λ and μ are the dual variables corresponding to the constraints (8.1b), (8.1c), (8.1d), (8.1e) and (8.1f), respectively. $\square$

Proof of Proposition 4.1. By the Lipschitz continuity and non-decreasing property of u over $[a,b]$,

$$
0\leq \frac{u(t_i)-u(t_{i-1})}{t_i-t_{i-1}}\leq L. \tag{8.2}
$$

Thus for any $t\in[t_{i-1},t_i]$,

$$
\begin{aligned}
|u_N(t)-u(t)| &= \left|u(t_{i-1})+\frac{u(t_i)-u(t_{i-1})}{t_i-t_{i-1}}(t-t_{i-1})-u(t)\right|\\
&= \left|\frac{t_i-t}{t_i-t_{i-1}}(u(t_{i-1})-u(t))+\frac{t-t_{i-1}}{t_i-t_{i-1}}(u(t_i)-u(t))\right|\\
&\leq \frac{t_i-t}{t_i-t_{i-1}}L|t_i-t_{i-1}|+\frac{t-t_{i-1}}{t_i-t_{i-1}}L|t_i-t_{i-1}|\\
&= L|t_i-t_{i-1}|\leq L\beta_N,
\end{aligned}
$$

which gives rise to (4.7). The conclusion that $L\geq 1/(b-a)$ follows from (8.2) in that $u(a)=0$ and $u(b)=1$. $\square$