Data-Driven Distributionally Preference Robust Optimization Models Based on Random Utility Representation in Multi-Attribute Decision Making

Jian Hu
Department of Industrial and Manufacturing Systems Engineering
University of Michigan - Dearborn, Dearborn, MI 48128
jianhu@umich.edu

Dali Zhang
Sino-US Global Logistics Institute
Antai College of Economics and Management
Shanghai Jiao Tong University
Shanghai 200030, China
zhangdl@sjtu.edu.cn

Huifu Xu
Department of Systems Engineering & Engineering Management
The Chinese University of Hong Kong, Hong Kong
hfxu@se.cuhk.edu.hk

September 8, 2021

Abstract

Preference robust optimization (PRO) has recently been studied to deal with utility based decision making problems under ambiguity in the characterization of the decision maker's (DM) preference. In this paper, we propose a novel PRO modeling paradigm which combines the stochastic utility theory with distributionally robust optimization technique. Based on the stochastic utility theory, our model is applicable to problems with inconsistent and mutable preference representations which are ubiquitous in practice, particularly in group or social decision making. In the framework of distributionally robust optimization, data-driven approaches are discussed to construct two ambiguity sets of the probability distributions of the DM's preference: one is an ellipsoidal moment region with a sample mean and sample covariance matrix, the other is a nonparametric percentile-t bootsrap confidence region. A numerical example of vehicle design demonstrates the effectiveness of the proposed model working with machine learning methods. We first depict the random preference in the structure of piecewise linear additive multi-attribute utility functions, which are either nondecreasing or risk averse, and next develop tractable reformulations and solution algorithms. Finally we extend to general continuous random utility functions and carry out convergence analysis of the proposed piecewise linear approximation as the size of sample data increases.

Key Words: multi-attribute decision making, piecewise linear random utility function, distributionally preference robust optimization, preference ambiguity

1 Introduction

In economics and decision theory, utility functions numerically measure individual or social preferences and tastes over a set of commodities. In decision making under uncertainty, a utility representation characterizes the DM's risk attitude towards the risk arising from systemic randomness (von Neumann and Morgenstern 1947, Sugden 1998). It is traditionally assumed that utility functions are deterministic and choices of the utility functions representing the DM's preferences are consistent. Extensive research, under these assumptions, has addressed utility assessment methods using preference comparison, probability equivalence, value equivalence, and certainty equivalence (Farquhar 1984, Wakker and Deneffe 1996). However, these assumptions are persistently violated in practice (Allais 1953, Simon 1956, Tversky 1969, Tversky and Kahneman 1981). Indeed, it is usually observed that the DM exhibits mutable and ambivalent preference, particularly without sufficient knowledge of complex problems and uncertain environment in the real world (Camerer 1989, Wu 1994, Hey and Orme 1994, Starmer 2000). Stochastic utility theories are subsequently proposed to extend the classical concepts and principles by interpreting the inconsistency of utility choice as the DM's probabilistic and random behavior (see Fishburn (1998), Koppen (2001) and therein). These theories allow one to regard utility as a random function distributed (Thurstone 1927, Daniels 1950). The random utility representation provides the theoretical foundation for the broadly implemented discrete choice models (Cascetta 2009).

In this paper, we consider a generic data-driven multi-attribute decision making problem where either the DM's utility preferences are potentially inconsistent or information is incomplete to assess the true preferences. We propose a novel distributionally robust maximin framework for optimal decision making which incorporates the features of random utility theory, preference learning, and distributionally robust optimization. The purpose of preference learning is to use available information (empirical data) to construct a set of probability distributions of a random multi-attribute utility function. We then base the optimal decision on the worst probability distribution for reducing the impact of the DM's random behavior. An analogy to our framework is the worst-case utility maximization which has been studied over the past few years (Armbruster and Delage 2015, Hu and Mehrotra 2015, Hu and Stepanyan 2017, Hu et al. 2018, Guo and Xu 2020). Literature studies recognize that even a deterministic utility function, assumed in von Neumann-Morgenstern's expected utility theorem, is unlikely to be elicited precisely (Weber and Borcherding 1993, Chajewska and Koller 2000). The worst-case utility maximization model considers the worst scenario from a set of deterministic utility functions for making a stable decision under ambiguity of utility assessment. By considering the distribution of a random utility function rather than the worst case, we are not trying to find the worst, best or optimal utility function to represent the DM's preference, rather we consider the case that the DM may have a range of utility functions (corresponding to his mutable preference) and each has a likelihood (probability) to represent the DM's preference.

This kind of modeling paradigm differs significantly from the worst-case utility models which the current research is focused on. There are at least two advantages to adopt our framework. One is that it is less conservative compared to the worst-case utility maximization models. The other is that it accommodates some inconsistent utility functions (preferences) to be included in the set of plausible utilities and their impact on the overall decision making problem is weighted by their frequencies in the empirical data. This contrasts the worst-case utility maximization model, established according to the traditional utility principles, where the preferences to be elicited must be consistent.

Furthermore, our framework can also be well applied into group or social choice making problems where individual preferences usually disagree with each other. In Harsanyi's social aggregation theorem (Harsanyi 1955), a social welfare function, representing public preference, is a weighted sum of individual utility functions. Regarding the trade-off weights as mass probabilities, individual utility functions are the realizations of an underlying random public preference and the social welfare function is the mean. A practical extension may first collect individual preferences as a random sample and estimate an confidence region of the social welfare function. Our framework next seeks a socially favorite solution over the confidence region of the social welfare, resorting to distributionally robust optimization under moment uncertainty.

1.1 Literature Review

The subjective expected utility maximization hypothesis presents two types of uncertainties that the DM often faces (Anscombe and Aumann 1963): (i) physical uncertainty that a random consequence is yielded regarding the state of nature; (ii) epistemic uncertainty that the DM has an ambiguous belief of the probability measure of the state (see (Hammond 1998a,b) and references therein). The epistemic uncertainty motives Gilboa and Schmeidler (1989) to propose a distributionally robust maximin model with a convex set of distributions to express multiple prior beliefs. The DM's risk preference (taste) toward the physical uncertainty is characterized by a utility function in the von Neumann-Morgenstern expected utility theory (von Neumann and Morgenstern 1947). Responding to the ambiguity in utility assessment, Armbruster and Delage (2015) and Hu and Mehrotra (2015) develop preference robust optimization (PRO for short) models separately.

Armbruster and Delage (2015) consider an ambiguity set of utility functions which meet some criteria such as preferring certain lotteries over other lotteries and being risk averse, S-shaped or prudent. These criteria extend the ones used in the first order or the second order stochastic dominance framework. Instead of trying to identify a single utility function satisfying the criteria, they develop a maximin PRO model where the optimal decision is based on the worst utility function from the ambiguity set and demonstrate how the maximin optimization can be reformulated as a finite dimensional linear programming problem.

Hu and Mehrotra (2015) tackle the issue in a different manner. First, they propose a moment-type framework for constructing the ambiguity set of DM's utility preference which covers a number of important approaches such as the certainty equivalent and pairwise comparison; second, 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; third, they consider lower and upper bound of 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. Qualitative convergence analysis is presented to justify the step-like approximation. A key assumption in Hu and Mehrotra's model is that the true unknown utility function is concave. While this assumption may be satisfied in many practically interesting applications, it excludes some S-shaped utility functions in behavioural economics and psychology (Tversky and Kahneman 1992) and risk management (Brown and Sim 2008, Brown et al. 2012)

Hu et al. (2018) and Guo and Xu (2021) extend the research by considering general non-concave utility functions. Hu et al. (2018) interpret a bounded utility function as a cumulative distribution function (c.d.f) and, analogous to distributionally robust optimization, describe the moment uncertainty of utility functions and the boundary conditions on the derivatives of utility functions. They develop a Lagrangian based Sample Average Approximation (SAA) solution method and present the convergence of the optimal value and the set of optimal solutions of the SAA method to the true counterparts. Guo and Xu (2021) propose a piecewise linear approximation scheme for the elicited utility functions represented via moment-type conditions and develop efficient computational schemes for solving the resulting preference robust optimization problem with the approximated ambiguity set. When the utility functions are concave, they reformulate the approximated maximin problem as a single linear programming problem as in Hu and Mehrotra (2015), and when the utility functions are increasing but not necessarily concave, they propose a derivative free algorithm for solving the PRO where the inner minimization is a linear program. Error bounds for the optimal value and optimal solutions of the outer maximization problem.

Hu and Stepanyan (2017) propose a reference-based robust preference constrained model for not only addressing the ambiguity and inaccuracy in characterizing the DM's individual preference but also reducing the conservativeness of PRO with a large-sized set of utility functions. They specify an $\mathcal{L}_2$-norm neighborhood centered at a given reference utility function on a general space of risk averse utility function. The ambiguity and conservativeness is balanced adjusting the radius of the neighborhood.

1.2 Contributions and Organization of this Paper

The main contributions are as follows:

- We propose a novel utility preference robust optimization modelling paradigm which combines the random utility theory with distributionally robust optimization techniques. The new PRO model is applicable to decision making problems with inconsistent utility preferences which are ubiquitous in practice, particularly in group or social decision making problems, and it complements the existing PRO models which presume consistency of the DM's utility preferences across the decision making. Moreover, the merit of the distributionally robust optimization allows the model to offer a less conservative approach to address the risk arising from the ambiguity of true utility preference.
- We propose two statistical approaches which suit data-driven problems for constructing an ambiguity set of the distributions of piecewise linear random multi-attribute utility function. One is to to construct an ellipsoidal confidence region with sample mean and sample covariance matrix which is widely used in the literature of distributional robust optimization; the other is to specify a nonparametric percentile-t bootstrap confidence region. Tractable reformulations and solution algorithms for solving the resulting PRO problems are developed.
- The proposed PRO model can be better integrated with the features of machine learning. Specifically, machine learning methods can be applied to generate a random sample of the DM's preference, by which we can construct an ambiguity set of distributions in the data-driven problems. This procedure is illustrated in a case study of vehicle design problem where a random sample of consumer's preferences on the passenger vehicles is generated using a generalized logistic regression method and the bootstrap approach is subsequently used to restructure the ambiguity set.
- Finally, we extend to general random utility functions and carry out convergence analysis of the proposed piecewise linear approximation as the size of sample data increases. Specifically, we show convergence of the optimal value obtained from solving a PRO model based on piecewise linear approximation and similar convergence is retained when the the underlying uncertainty data are gained via samples.

The paper is organized as follows. Section 2 structures random multi-attribute utility functions with a piecewise linear additive form. Section 3 proposes the distributional PRO Model and the two data-driven approaches to specify the ambiguity set of the distribution of the random utility function. We consider the cases that utility functions are either generally nondecreasing or risk averse. The tractable reformulations and solution methods of the PRO model are developed in Section 4. Section 4.1 addresses the case of nondecreasing utility function with the ambiguity set configured in either data-driven approach, while Section 4.2 discusses the case of risk averse utility function. In Section 5 we conduct a simulation test to demonstrate the effectiveness of the PRO model and use an problem of vehicle design project selection to illustrate the combination between the data-driven approach and preference learning. Section 6 extends to a general case of random utility function which relaxes the restriction on the piecewise linear structure. We study the convergence properties of the piecewise linear approximation of a random utility function in Section 6.1, the sample average approximation of the expectation maximization in Section 6.2, and the proposed data-driven PRO model in Section 6.3. Section 7 concludes the paper.

2 Random Additive Piecewise Linear Multi-Attribute Utility Functions

We consider multi-attribute utility functions with M attributes. Let x_m be the performance of attribute m, for $m \in \mathfrak{M} := \{1, \ldots, M\}$, and $x := (x_1, \cdots, x_M)' \in \mathbb{R}^m$ be the vector of all attribute performances. Denote by $u_m : \mathbb{R} \to \mathbb{R}$ the marginal utility function for attribute m. As all non-parametric utility assessment methods recommend (Farquhar 1984, Wakker and Deneffe 1996), we consider u_m to be a piecewise linear utility function defined over interval $[a_m, b_m]$ with breakpoints

$$a_m = t_{m,0} < \cdots < t_{m,I_m} = b_m. \tag{1}$$

Let $v_{m,i} := u_m(t_{m,i}) - u_m(t_{m,i-1})$ be the overall increment of u_m over $[t_{m,i-1}, t_{m,i}]$ for $i \in \mathfrak{I}_m := \{1, \ldots, I_m\}$ and $v_m := (v_{m,1}, \cdots v_{m,I_m})' \in \mathbb{R}^{I_m}$. If we regard v_m as a vector of parameters, then we can obtain a class of piecewise linear functions for the m-th marginal utility u_m as follows:

$$u_m(x_m; v_m) := \sum_{i \in \mathfrak{I}_m} \left[\frac{v_{m,i}}{t_{m,i} - t_{m,i-1}} (x_m - t_{m,i-1}) + \sum_{j=1}^{i-1} v_{m,j} \right] \mathbf{1}\{t_{m,i-1} < x_m \leq t_{m,i}\}, \tag{2}$$

where $\mathbf{1}\{t_{m,i-1} < x_m \leq t_{m,i}\}$ is an indicator function.

Let $v := (v_1, \cdots, v_M)'$ and $I := \sum_{m \in \mathfrak{M}} I_m$. We can then define a class of holistic multi-attribute piecewise-linear utility function $u : \mathbb{R}^m \times \mathbb{R}^I \to \mathbb{R}$ with an additive form as

$$u(x; v) := \sum_{m \in \mathfrak{M}} u_m(x_m; v_m). \tag{3}$$

Without loss of generality, we assume that the utility function is nondecreasing and normalized to $[0, 1]$ because the normalization does not affect its representation of the DM's preference. Under this assumption, the utility function u_m is uniquely determined by vector v_m for $m \in \mathfrak{M}$ for a given set of breakpoints. Moreover, the vector v lies in the simplex of $\mathbb{R}^I$ with $v \geq 0$ and $e_I' v = 1$ where e_I is the I-dimensional vector of all ones. Note that a traditionally defined additive multi-attribute utility function is the weighted sum of normalized marginal utility functions. The normalization of u and the definition of the utility function in (2) and (3) ensure that the criterion weight of the m-th attribute is included but hidden in u_m.

In practice, the vector v of increments are determined through empirical data which are typically random. This motivates us to replace the deterministic vector of parameters v with a random vector $V : (\Omega, \mathcal{F}, P) \to \mathbb{R}^I$ in (3) and subsequently obtain a random piecewise linear utility function mapping from $\mathbb{R}^m$ to $\mathbb{R}$ as

$$u(x; V) = \sum_{m \in \mathfrak{M}} u_m(x_m; V_m). \tag{4}$$

This effectively randomizes the class of parameterized increasing piecewise linear utility functions with a support set of the random parameter V as

$$\mathfrak{V} := \left\{ v \in \mathbb{R}^I \mid e_I' v = 1,\ v \geq 0 \right\}. \tag{5}$$

In some cases we may require that a utility function have properties on its shape. For example, the DM's risk averse preference should be characterised as an increasing concave utility function. The support set in the case of increasing concave utility functions is described as

$$\mathfrak{V}_C := \left\{ v \in \mathbb{R}^I \left| e_I' v = 1,\ Av \geq 0 \right. \right\}, \tag{6}$$

where matrix

$$A := \begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_M \end{bmatrix}_{(I-M)\times I},$$

in which blocks A_m for $m \in \mathfrak{M}$ are

$$A_m := \begin{bmatrix} \frac{1}{t_{m,1}-t_{m,0}} & \frac{-1}{t_{m,2}-t_{m,1}} & & & \\ & \frac{1}{t_{m,2}-t_{m,1}} & \frac{-1}{t_{m,3}-t_{m,2}} & & \\ & & \ddots & \ddots & \\ & & & \frac{1}{t_{m,I_m-1}-t_{m,I_m-1}} & \frac{-1}{t_{m,I_m}-t_{m,I_m-1}} \end{bmatrix}_{(I_m-1)\times I_m}.$$

In the later statement we use $\mathfrak{V}$ for the general description of the support and $\mathfrak{V}_C$ for the specific discussion of increasing concave utility function.

3 Data-Driven Distributional Preference Robust Optimization Model

We consider a decision making problem which aims to maximize the overall utility of attribute performances. Since the DM's utility function is stochastic, we consider the expected utility $\mathbb{E}_P[u(x;V)]$ where the expectation is taken w.r.t the probability distribution of V. Moreover, since the true probability distribution is unknown in some data-driven problems, we may have to rely on partially available information such as empirical data, computation simulations, or subjective judgement to construct an ambiguity set of distributions.

3.1 Models

Denote by $\mathfrak{P}$ the ambiguity set of distributions of V and base the optimal decision on the worst probability distribution from the set as follows:

$$\max_{x\in\mathfrak{X}} \min_{P\in\mathfrak{P}} \mathbb{E}_P[u(x;V)], \tag{DPRO}$$

where $\mathfrak{X} \subseteq \otimes_{m\in\mathfrak{M}}[a_m, b_m] := [a_1, b_1] \times \cdots \times [a_M, b_M]$ is the joint region of possible attribute values. DPRO differs from the existing utility preference robust optimization models in that here the worst probability distribution of the random utility function, instead of the worst case utility function, is chosen.

In practice, x is often a vector-valued function of some action denoted by z, that is, $x = h(z)$. Let $\mathfrak{Z}$ be a feasible action space. It follows that

$$\mathfrak{X} = \{x \mid x = h(z),\ z \in \mathfrak{Z}\}.$$

Consequently we rewrite this variation case of DPRO as

$$\max_{z\in\mathfrak{Z}} \min_{P\in\mathfrak{P}} \mathbb{E}_P[u(h(z); V)]. \tag{DPRO-1}$$

In the case that the relation between x and z is affected by some exogenous uncertainties denoted by a vector of random variables ξ with distribution Q, we may obtain a further variation

$$\max_{z\in\mathfrak{Z}} \min_{P\in\mathfrak{P}} \mathbb{E}_P[\mathbb{E}_Q[u(h(z,\xi); V)]]. \tag{DPRO-2}$$

We next focus on discussing DPRO. All the model configurations and corresponding solution methods presented in this paper can be straightforwardly extended to DPRO-1 and DPRO-2.

3.2 Construction of ambiguity set

A key component of DPRO is the ambiguity set. We now outline two main approaches for constructing the ambiguity set $\mathfrak{P}$ in data-driven environment. Denote by μ and Σ the mean and covariance matrix of V. Here we consider a situation where the true μ and Σ are unknown but it is possible to obtain an approximation with sample data. Let $V^1, \ldots, V^N$ be an independent and identically distributed (iid for short) random sample of V and denote by $\bar{V}$ the sample mean and by S a sample covariance matrix approximating their true counterparts μ and Σ. The generation of a random sample will be illustrated by the case study in Section 5 which, using machine learning, elicits the realizations of the social welfare function representing consumer preferences on the passenger vehicle market in the Unite State.

In what follows we address how to construct a set of moment uncertainty using the random sample. The support set $\mathfrak{V}$ defined in (5) indicates that the components of V are linear dependent, i.e, $e_I' V = 1$ almost surely. Therefore, we can reduce by one dimension to consider the first $I-1$ components of V in the construction of the ambiguity set. Accordingly, V_{M,I_m}, which is the I-th component of V, is equal to 1 less the sum of the other components. Let

$$C := \begin{bmatrix} 1 & & 0 \\ & \ddots & \vdots \\ & 1 & 0 \end{bmatrix}_{(I-1)\times I}. \tag{7}$$

Then CV is the vector consisting of the first $I-1$ components of V. Denote by $\mu_{<-I>}$ and $\Sigma_{<-I>}$ the mean and covariance matrix of CV and by $\bar{V}_{<-I>}$ and $S_{<-I>}$ the sample counterparts of $\mu_{<-I>}$ and $\Sigma_{<-I>}$. Note that

$$\mu_{<-I>} = C\mu,\ \Sigma_{<-I>} = C\Sigma C',\ \bar{V}_{<-I>} = C\bar{V},\ \text{and } S_{<-I>} = CSC'.$$

To facilitate our analysis in the forthcoming discussions, we make a blanket assumption on $\Sigma_{<-I>}$.

Assumption 1 *The covariance matrix $\Sigma_{<-I>}$ is nonsingular.*

Under this assumption, $S_{<-I>}$ is nonsingular for a sufficiently large sample size N, which is needed in the following constructions of the uncertainty set $\mathfrak{P}$. The assumption requires that the components of CV be linearly independent. If the assumption is not satisfied, then we will repeat the procedure of dimension reduction until remaining components are linearly independent.

Ellipsoid approach. A classical approach of distributionally robust optimization is to specify the ambiguity set by moment conditions (Scarf 1958, Bertsimas and Popescu 2005, Yue et al. 2006, Popescu 2007, Delage and Ye 2010). Here we consider an ambiguity with ellipsoidal structure of the first moment:

$$\mathfrak{P}_A(\gamma) := \left\{ P \in \mathscr{P}(\mathbb{R}^I) \;\middle|\; \begin{array}{l} P(V \in \mathfrak{V}) = 1 \\ \left\| S_{<-I>}^{-1/2} \left(C\mathbb{E}_P[V] - \bar{V}_{<-I>} \right) \right\|^2 \leq \gamma \end{array} \right\}, \tag{8}$$

where $\mathscr{P}(\mathbb{R}^I)$ denotes the set of all probability measures on the space $\mathbb{R}^I$, $\|\cdot\|$ denotes the Euclidean norm and $S_{<-I>}^{1/2}$ is a full rank matrix with $S_{<-I>} = \left(S_{<-I>}^{1/2}\right)' S_{<-I>}^{1/2}$ and $S_{<-I>}^{-1/2} = \left(S_{<-I>}^{1/2}\right)^{-1}$. The second condition in the set $\mathfrak{P}_A(\gamma)$ specifies the range of $C\mathbb{E}_P[V]$, that is, we consider only the candidate probability distributions with associated means value of V falling within a specified ellipsoid. Parameter γ determines the size of the ellipsoid. If we are able to recover the true probability distribution with a sample of size N, then we may set $\gamma = 0$. In practice, however, this is very unlikely particularly when the sample size is small. We refer readers to our discussion in Section 6.3 and also Delage and Ye (2010), So (2011) for proper setting of γ.

Bootstrap approach. We now discuss an alternative approach using a percentile-t bootstrap method to specify a confidence region for $\mu_{<-I>}$. Generate K nonparametric bootstrap resamples of the first $I-1$ components, denoted by $V_{<-I>}^{(k,1)}, \ldots, V_{<-I>}^{(k,N)}$, $k = 1, \ldots, K$, based on the empirical distribution. Denote by $\bar{V}_{<-I>}^{*,k}$ and $S_{<-I>}^{*,k}$ the sample mean and the sample covariance matrix of $V_{<-I>}^{(k,1)}, \ldots, V_{<-I>}^{(k,N)}$. Consider the studentized statistics

$$T := \sqrt{N} S_{<-I>}^{-1/2} (\bar{V}_{<-I>} - \mu_{<-I>})$$

and its bootstrap counterparts

$$T_k^* := \sqrt{N} \left(S_{<-I>}^{*,k} \right)^{-1/2} (\bar{V}_{<-I>}^{*,k} - \bar{V}_{<-I>}), \quad k = 1, \cdots, K. \tag{9}$$

The literature addresses multivariate bootstrap confidence regions using data depth and likelihood (Hall 1987, Yeh and Singh 1997, Battista and Gattone 2004). Here we state the procedure given by Yeh and Singh (1997). Recall that Tukey's depth of a point x under some distribution $\mathcal{F}$ is defined as

$$TD(\mathcal{F}, x) = \inf\{\mathcal{F}(H) \mid H \text{ is a half-closed space containing } x\}.$$

Let $\mathcal{F}_K^*$ be the empirical cumulative distribution function built using all of T_k^* and calculate $d_k := TD(\mathcal{F}_K^*, T_k^*)$. Denote by $d_{(1)}, \ldots, d_{(K)}$ the increasing order of $d_1, \ldots, d_K$, and let $T_{(k)}^*$ be the statistics such that $d_{(k)} = TD(\mathcal{F}_K^*, T_{(k)}^*)$. For a given $\alpha \in (0,1)$, let $\hat{T}_{1-\alpha} := \left[T_{(1)}^*, \ldots, T_{(\lceil (1-\alpha)K \rceil)}^* \right]$ be an $M \times \lceil (1-\alpha)K \rceil$ dimensional matrix. Here, $\lceil \cdot \rceil$ is the ceiling function. We construct a convex hull, denoted by $\mathfrak{W}_{1-\alpha}^*$, of $T_{(k)}^*$ for $1 \leq k \leq \lceil (1-\alpha)K \rceil$, i.e.,

$$\mathfrak{W}_{1-\alpha}^* := \left\{ \hat{T}_{1-\alpha} w \;\middle|\; e'_{\lceil (1-\alpha)K \rceil} w = 1, \; w \geq 0 \right\}. \tag{10}$$

Then a $100(1-\alpha)\%$ bootstrap confidence region for $\mu_{<-I>}$ is obtained as

$$\mathfrak{C}^*_{1-\alpha} := \left\{ \bar{V}_{<-I>} - S^{1/2}_{<-I>} \tilde{w}/\sqrt{N} \;\middle|\; \tilde{w} \in \mathfrak{W}^*_{1-\alpha} \right\}. \tag{11}$$

On this basis, we define an ambiguity set of the first moment as

$$\mathfrak{P}_B(\alpha) := \left\{ P \in \mathscr{P}(\mathbb{R}^I) \mid P(V \in \mathfrak{V}) = 1,\ C\mathbb{E}_P[V] \in \mathfrak{C}^*_{1-\alpha} \right\}. \tag{12}$$

In this bootstrap approach, α is the critical value of the confidence region. A larger α means less samples from the bootstrap resampling process are included. Subsequently, $\mathfrak{C}^*_{1-\alpha}$ and $\mathfrak{P}_B(\alpha)$ are smaller.

4 Reformulations and Solution Methods

In this section, we develop computational approaches for solving DPRO with the ambiguity sets $\mathfrak{P}_A(\gamma)$ and $\mathfrak{P}_B(\alpha)$ defined in (8) and (12), respectively. Section 4.1 addresses the case with increasing utility functions, while Section 4.2 discusses the case with risk averse utility functions. We begin by presenting a reformulation of the objective function of DPRO as a linear function of the random vector V of increment in the next proposition.

Proposition 1 *Let*

$$g_{m,i}(x_m) := \mathbf{1}\{t_{m,i-1} < x_m \le t_{m,i}\} \quad \text{and} \quad h_{m,i}(x_m) := \left(\frac{x - t_{m,i-1}}{t_{m,i} - t_{m,i-1}} \right) g_{m,i}(x_m).$$

Let

$$f_m(x_m) := \begin{bmatrix} h_{m,1}(x_m) + \sum_{j \in \mathfrak{I}_m \setminus \{1\}} g_{m,j}(x_m) \\ h_{m,2}(x_m) + \sum_{j \in \mathfrak{I}_m \setminus \{1,2\}} g_{m,j}(x_m) \\ \vdots \\ h_{m,I_m}(x_m) \end{bmatrix}_{I_m \times 1}$$

and $f(x) := (f_1(x_1)', \ldots, f_M(x_M)')'$. *Then*

(i) $u_m(x_m; v_m) = v_m' f_m(x_m)$ *and*

$$u(x; v) = v' f(x). \tag{13}$$

(ii) DPRO is equivalent to

$$\max_{x \in \mathfrak{X}} \min_{P \in \mathfrak{P}} \mathbb{E}_P[V'] f(x) \tag{14}$$

and

$$\max_{x \in \mathfrak{X}} \min_{v \in \mathfrak{F}(\mathfrak{P})} \left\{ u(x; v) = v' f(x) \right\}, \tag{15}$$

where $\mathfrak{F}(\mathfrak{P}) := \{\mathbb{E}_P(V) \mid P \in \mathfrak{P}\}$.

To facilitate the discussions of DPRO, we introduce more notations. Denote two vectors $y = (y_1', \ldots, y_M')' \in \mathbb{R}^I$ and $z = (z_1', \ldots, z_M')' \in \{0,1\}^I$ where $y_m = (y_{m,1}, \ldots y_{m,I_m})' \in \mathbb{R}^{I_m}$ and $z_m = (z_{m,1}, \ldots z_{m,I_m})' \in \{0,1\}^{I_m}$ for $m \in \mathfrak{M}$. Let

$$Y := \begin{bmatrix} Y_1 & & \\ & \ddots & \\ & & Y_M \end{bmatrix}_{I \times I}, \quad Z := \begin{bmatrix} Z_1 & & \\ & \ddots & \\ & & Z_M \end{bmatrix}_{I \times I},$$

where the blocks Y_m and Z_m for $m \in \mathfrak{M}$ are

$$Y_m := \begin{bmatrix} \frac{1}{t_{m,1}-t_{m,0}} & & \\ & \ddots & \\ & & \frac{1}{t_{m,I_m}-t_{m,I_m-1}} \end{bmatrix}_{I_m \times I_m}, \quad Z_m := \begin{bmatrix} \frac{-t_{m,0}}{t_{m,1}-t_{m,0}} & 1 & \cdots & 1 \\ & \frac{-t_{m,1}}{t_{m,2}-t_{m,1}} & \cdots & 1 \\ & & \ddots & \vdots \\ & & & \frac{-t_{m,I_m-1}}{t_{m,I_m}-t_{m,I_m-1}} \end{bmatrix}_{I_m \times I_m}.$$

On this basis, we introduce vector-valued linear functions as

$$\phi(y, z) := Yy + Zz.$$

Denote other five block diagonal matrices

$$B := \begin{bmatrix} B_1 & & \\ & \ddots & \\ & & B_M \end{bmatrix}_{I \times I}, \quad D := \begin{bmatrix} D_1 & & \\ & \ddots & \\ & & D_M \end{bmatrix}_{I \times M}, \quad E := \begin{bmatrix} e_{I_1} & & \\ & \ddots & \\ & & e_{I_M} \end{bmatrix}_{I \times M},$$

$$H^- := \begin{bmatrix} H_1^- & & \\ & \ddots & \\ & & H_M^- \end{bmatrix}_{I \times I}, \quad H^+ := \begin{bmatrix} H_1^+ & & \\ & \ddots & \\ & & H_M^+ \end{bmatrix}_{I \times I},$$

where the blocks B_m, D_m, H_m^-, and H_m^+ for $m \in \mathfrak{M}$ are

$$B_m := \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ \vdots & \vdots & \ddots & \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{I_m \times I_m}, \quad D_m := \begin{bmatrix} t_{m,1} \\ \vdots \\ t_{m,I_m} \end{bmatrix}_{I_m \times 1},$$

$$H_m^- := \begin{bmatrix} t_{m,0} & & \\ & \ddots & \\ & & t_{m,I_m-1} \end{bmatrix}_{I_m \times I_m}, \quad H_m^+ := \begin{bmatrix} t_{m,1} & & \\ & \ddots & \\ & & t_{m,I_m} \end{bmatrix}_{I_m \times I_m}.$$

4.1 General Increasing Utility Function

We first provide a reformulation of DPRO with the ambiguity set $\mathfrak{P}_A(\gamma)$ in the following proposition.

Proposition 2 *DPRO with the ambiguity set $\mathfrak{P}_A(\gamma)$ defined in* (8) *is equivalent to*

$$\max_{x,y,z,\tau,\eta,\pi} \ \gamma^{-1/2}\bar{V}'_{<-I>}\left(S^{-1/2}_{<-I>}\right)'\eta-\pi+\tau \tag{16a}$$

$$s.t.\ \gamma^{-1/2}C'\left(S^{-1/2}_{<-I>}\right)'\eta+e_I\tau\leq Yy+Zz, \tag{16b}$$

$$\|\eta\|\leq\pi, \tag{16c}$$

$$E'z=e_M, \tag{16d}$$

$$E'y=x, \tag{16e}$$

$$H^-z-y\leq 0, \tag{16f}$$

$$H^+z-y\geq 0, \tag{16g}$$

$$z\in\{0,1\}^I, \tag{16h}$$

$$x\in\mathfrak{X}. \tag{16i}$$

Proof: It follows by Proposition 1 that DPRO with $\mathfrak{P}_A(\gamma)$ is equivalent to model (15) with the ambiguity set

$$\mathfrak{F}(\mathfrak{P}_A(\gamma))=\left\{v\in\mathbb{R}^I\ \middle|\ \begin{array}{l} v\geq 0\\ e_I'v=1\\ \|S^{-1/2}_{<-I>}(Cv-\bar{V}_{<-I>})\|^2\leq\gamma \end{array}\right\}.$$

In this case the inner minimization problem of model (15) is

$$\min_{v\in\mathfrak{F}(\mathfrak{P}_A(\gamma))} v'f(x).$$

Let us further represent the condition $\|S^{-1/2}_{<-I>}(Cv-\bar{V}_{<-I>})\|^2\leq\gamma$ in a conic formulation as $\tilde{A}v+\tilde{B}\in\mathfrak{K}$, where

$$\tilde{A}:=\left[\begin{array}{c}\gamma^{-1/2}S^{-1/2}_{<-I>}C\\ \mathbf{0}_{1\times I}\end{array}\right],\qquad \tilde{B}:=\left[\begin{array}{c}-\gamma^{-1/2}S^{-1/2}_{<-I>}\bar{V}_{<-I>}\\ 1\end{array}\right],$$

and $\mathfrak{K}:=\{(s,\zeta)\in\mathbb{R}^I\mid\|s\|\leq\zeta\}$. Let τ be the dual variable of the second condition in $\mathfrak{F}(\mathfrak{P}_A(\gamma))$, and (η,π) be the dual variable of the third condition. The dual problem is represented as

$$\begin{aligned}\max\ & -\tilde{B}'\begin{pmatrix}\eta\\ \pi\end{pmatrix}+\tau\\ \text{s.t. } & \tilde{A}'\begin{pmatrix}\eta\\ \pi\end{pmatrix}+e_I\tau\leq f(x),\\ & \begin{pmatrix}\eta\\ \pi\end{pmatrix}\in\mathfrak{K}^*,\end{aligned}$$

where $\mathfrak{K}^*=\{(\eta,\pi)\mid\|\eta\|\leq\pi\}$ is the dual cone of $\mathfrak{K}$. Expanding $\tilde{A}$ and $\tilde{B}$ in the problem above, we have

$$\begin{aligned}\max_{\tau,\eta,\pi}\ & \gamma^{-1/2}\bar{V}'_{<-I>}\left(S^{-1/2}_{<-I>}\right)'\eta-\pi+\tau\\ \text{s.t. } & \gamma^{-1/2}C'\left(S^{-1/2}_{<-I>}\right)'\eta+e_I\tau\leq f(x),\\ & \|\eta\|\leq\pi.\end{aligned}$$

We substitute for $f(x)$ in the model above, using vectors y, z and function ϕ, i.e, $f(x) = \phi(y, z)$ under the conditions (16d)-(16g). □

Proposition 2 reformulates DPRO as a mixed-integer second order cone program (MISOCP). MISOCP has been well studied in the literature (Benson and Sağlam (2014)). The existing solution algorithms formulate MISOCP to MILP or MINLP.

In our case we can develop a more efficient cutting surface algorithm for solving model (15). We begin by solving a sequence of master problems

$$\max_{x,t}\ t \tag{17a}$$
$$\text{s.t. } t \leq u(x, v^j), \qquad j = 1, \dots, k, \tag{17b}$$
$$x \in \mathfrak{X}. \tag{17c}$$

Let (x^*, t^*) be an optimal solution of model (17). Next we solve the following sub-problem

$$\theta^* := \max_{v \in \mathfrak{F}(\mathfrak{P}_A(\gamma))} v' f(x^*) \tag{18}$$

and obtain the optimal value θ^* and an optimal solution v^*. If $t^* \leq \theta^*$, x^* is an optimal solution of model (15). Otherwise, we add v^* as an additional cut into the constraint (17b).

In general we rewrite model (17) as a MILP problem

$$\max_{x,t,y,z}\ t \tag{19a}$$
$$\text{s.t. } t \leq {v^j}'(Yy + Zz), \qquad j = 1, \dots, k, \tag{19b}$$
$$E'z = e_M, \tag{19c}$$
$$E'y = x, \tag{19d}$$
$$H^- z - y \leq 0, \tag{19e}$$
$$H^+ z - y \geq 0, \tag{19f}$$
$$z \in \{0, 1\}^I, \tag{19g}$$
$$x \in \mathfrak{X}. \tag{19h}$$

Based on the ambiguity set $\mathfrak{P}_A(\gamma)$, the sub-problem (18) is a quadratic constrained problem. This cutting surface approach is summarized in Algorithm 1.

Algorithm 1 Cutting Surface Algorithm for Model DPRO with $\mathfrak{P}_A(\gamma)$

`Step 1`	Choose $\delta > 0$, $k = 1$, and $v^1 = \bar{V}$.
`Step 2`	Find an optimal solution (x^*, t^*) of problem (19).
`Step 3`	Obtain the optimal value θ^* and an optimal solution v^* of problem (18).
`Step 4`	If $t^* \leq \theta^* - \delta$, exit. Otherwise, let $v^{k+1} = v^*$ and $k := k + 1$, go to step 2.

We next discuss the reformulation of DPRO with the ambiguity set $\mathfrak{P}_B(\alpha)$ in the proposition below.

Proposition 3 *DPRO with the ambiguity set $\mathfrak{P}_B(\alpha)$ defined in* (12) *is equivalent to*

$$\max_{x,y,z,\eta,\pi,\tau} \; \bar{V}'_{<-I>}\eta + \pi + \tau \tag{20a}$$

$$s.t. \; C'\eta + e_I\tau \leq Yy + Zz, \tag{20b}$$

$$N^{-1/2}\hat{T}'_{1-\alpha}\left(S^{1/2}_{<-I>}\right)'\eta + e_{\lceil(1-\alpha)K\rceil}\pi \leq 0, \tag{20c}$$

$$E'z = e_M, \tag{20d}$$

$$E'y = x, \tag{20e}$$

$$H^-z - y \leq 0, \tag{20f}$$

$$H^+z - y \geq 0, \tag{20g}$$

$$z \in \{0,1\}^I, \tag{20h}$$

$$x \in \mathfrak{X}. \tag{20i}$$

Proof: By Proposition 1, DPRO with the ambiguity set $\mathfrak{P}_B(\alpha)$ is equivalent to (15) with $\mathfrak{F}(\mathfrak{P}_B(\alpha))$. Also, for any $Cv \in \mathfrak{C}^*_{1-\alpha}$, there exists a $w \in \mathbb{R}^{\lceil(1-\alpha)\rceil}$ such that $Cv = \bar{V}_{<-I>} - N^{-1/2}S^{1/2}_{<-I>}\hat{T}_{1-\alpha}w$, $e'_{\lceil(1-\alpha)K\rceil}w = 1$, and $w \geq 0$. Hence, the inner minimization problem of (15) is written as

$$\begin{aligned}
\min_{v,w} \; & v'f(x) \\
\text{s.t. } & Cv + N^{-1/2}S^{1/2}_{<-I>}\hat{T}_{1-\alpha}w = \bar{V}_{<-I>}, \\
& e'_{\lceil(1-\alpha)K\rceil}w = 1, \\
& e'_I v = 1, \\
& v \geq 0, \; w \geq 0.
\end{aligned}$$

Let η, π, and τ be the dual variables regarding the first three constraints in the problem above. We obtain the dual problem as

$$\begin{aligned}
\max_{\eta,\pi,\tau} \; & \bar{V}'_{<-I>}\eta + \pi + \tau \\
\text{s.t. } & C'\eta + e_I\tau \leq f(x), \\
& N^{-1/2}\hat{T}'_{1-\alpha}\left(S^{1/2}_{<-I>}\right)'\eta + e_{\lceil(1-\alpha)K\rceil}\pi \leq 0.
\end{aligned}$$

We substitute for $f(x)$ in the problem above, using vectors y, z and function ϕ, i.e, $f(x) = \phi(y,z)$ under the conditions (20d)-(20g). □

4.2 Increasing Concave (Risk-Averse) Utility Function

We now assume that all single-attribute marginal utility functions are increasing concave. In this case, $\mathfrak{V}_C$ defined in (6) is the support set of the random vector V of increment. In this section we substitute $\mathfrak{V}_c$ for $\mathfrak{V}$ in the ambiguity sets $\mathfrak{P}_A(\gamma)$ and $\mathfrak{P}_B(\alpha)$ given in (8) and (12). The following proposition reformulates DPRO with the ambiguity set $\mathfrak{P}_A(\gamma)$ to a SOCP problem.

Proposition 4 *DPRO with the ambiguity set $\mathfrak{P}_A(\gamma)$, where $\mathfrak{V}_C$ is the support, is equivalent to*

$$\max_{x,\eta,\tau,\lambda,\zeta,\pi} \ \gamma^{-1/2}\bar{V}'_{<-I>}\left(S^{-1/2}_{<-I>}\right)'\eta - \pi + \tau \tag{21a}$$

$$s.t.\ \gamma^{-1/2}C'\left(S^{-1/2}_{<-I>}\right)'\eta + e_I\tau + A'\zeta - B'\lambda \leq 0, \tag{21b}$$

$$\|\eta\| \leq \pi, \tag{21c}$$

$$D'\lambda \leq x, \tag{21d}$$

$$E'\lambda \leq e_M, \tag{21e}$$

$$\zeta \geq 0,\ \ \lambda \geq 0, \tag{21f}$$

$$x \in \mathfrak{X}. \tag{21g}$$

Proof: We first prove that the inner minimization problem of DPRO can be represented as

$$\min_{y,z}\ x'y + e'_M z \tag{22a}$$

$$\text{s.t. } \|S^{-1/2}_{<-I>}(Cv - \bar{V}_{<-I>})\|^2 \leq \gamma, \tag{22b}$$

$$e'_I v = 1, \tag{22c}$$

$$Av \geq 0, \tag{22d}$$

$$Dy + Ez - Bv \geq 0, \tag{22e}$$

$$v \geq 0,\ y \geq 0,\ z \geq 0. \tag{22f}$$

The $\mathfrak{F}$-mapping of $\mathfrak{P}_A(\gamma)$ with the support $\mathfrak{V}_C$ is written as

$$\mathfrak{F}(\mathfrak{P}_A(\gamma)) = \left\{v \in \mathbb{R}^I_+ \mid v \text{ satisfies conditions (22b) - (22d)}\right\}$$

For any given $v \in \mathfrak{F}(\mathfrak{P}_A(\gamma))$, $u_m(x_m; v_m)$ is a piecewise linear increasing concave function with breakpoints t_i and values $\sum_{j=1}^{i} v_{m_j}$, $i = 1, \ldots, I_m$. It can be written as the minimization of all subgradients at points $\left(t_i, \sum_{j=1}^{i} v_{m_j}\right)$ as

$$\begin{aligned} u_m(x_m; v_m) = \min_{y_m, z_m}\ & x_m y_m + z_m \\ \text{s.t. }\ & t_m y_m + z_m - B_m v_m \geq 0, \\ & y_m \geq 0,\ z_m \geq 0. \end{aligned}$$

Therefore, we combine all u_m to the multiattribute utility function

$$\begin{aligned} u(x; v) = \sum_{m \in \mathfrak{M}} u_m(x_m; v_m) = \min_{y,z}\ & x^T y + e^T_M z \\ \text{s.t. }\ & Dy + Ez - Bv \geq 0, \\ & y \geq 0,\ z \geq 0. \end{aligned}$$

This gives the objective (22a) and constraints (22e) and (22f).

Let (η, π), τ, ζ, and λ be the dual variables of the constraints (22b) - (22f), respectively. We can write the dual problem of (22) as

$$\begin{aligned}\max_{\eta,\tau,\lambda,\zeta,\pi}\ & \gamma^{-1/2}\bar{V}'_{<-I>}\left(S^{-1/2}_{<-I>}\right)'\eta-\pi+\tau\\ \text{s.t. } & \gamma^{-1/2}C'\left(S^{-1/2}_{<-I>}\right)'\eta+e_I\tau+A'\zeta-B'\lambda\le 0,\\ & \|\eta\|\le\pi,\\ & D'\lambda\le x,\\ & E'\lambda\le e_M,\\ & \zeta\ge 0,\ \ \lambda\ge 0.\end{aligned}$$

□

Next we discuss a reformulation of DPRO with the ambiguity set $\mathfrak{P}_B(\alpha)$ in the proposition below.

Proposition 5 *DPRO with the ambiguity set $\mathfrak{P}_B(\alpha)$, where $\mathfrak{V}_C$ is the support, is equivalent to*

$$\max_{x,\eta,\pi,\tau,\zeta,\lambda}\ \bar{V}'_{<-I>}\eta+\pi+\tau \tag{24a}$$

$$\text{s.t. } C'\eta+e_I\tau+A'\zeta-B'\lambda\le 0, \tag{24b}$$

$$N^{-1/2}\hat{T}'_{1-\alpha}\left(S^{1/2}_{<-I>}\right)'\eta+e_{\lceil(1-\alpha)K\rceil}\pi\le 0, \tag{24c}$$

$$D'\lambda\le x, \tag{24d}$$

$$E'\lambda\le e_M, \tag{24e}$$

$$\zeta\ge 0,\ \ \lambda\ge 0, \tag{24f}$$

$$x\in\mathfrak{X}. \tag{24g}$$

Proof: Analogous to the proof of Proposition 4, the inner minimization problem of DPRO can be represented as

$$\begin{aligned}\min_{y,z}\ & x'y+e'_Mz\\ \text{s.t. } & Cv+N^{-1/2}S^{1/2}_{<-I>}\hat{T}_{1-\alpha}w=\bar{V}_{<-I>},\\ & e'_{\lceil(1-\alpha)K\rceil}w=1,\\ & e'_Iv=1,\\ & Av\ge 0,\\ & Dy+Ez-Bv\ge 0,\\ & v\ge 0,\ \ w\ge 0,\ \ y\ge 0,\ \ z\ge 0.\end{aligned}$$

Letting η, π, τ, ζ, and λ be the dual variables regarding the constraints, respectively, in the problem above, we can derive the Lagrange dual

$$\begin{aligned}\max_{\eta,\pi,\tau,\zeta,\lambda}\ & \bar{V}'_{<-I>}\eta+\pi+\tau\\ \text{s.t. } & C'\eta+e_I\tau+A'\zeta-B'\lambda\le 0,\\ & N^{-1/2}\hat{T}'_{1-\alpha}\left(S^{1/2}_{<-I>}\right)'\eta+e_{\lceil(1-\alpha)K\rceil}\pi\le 0,\\ & D'\lambda\le x,\\ & E'\lambda\le e_M,\\ & \zeta\ge 0,\ \ \lambda\ge 0.\end{aligned}$$

A combination of the above dual program with the outer maximization problem yields (24). $\square$

5 Case studies

In this section, we construct a number of testing cases and undertake numerical experiments to investigate the performance of DPRO with the ambiguity set $\mathfrak{P}_B(\alpha)$ defined in (12). Note that $\mathfrak{P}_B(\alpha)$ is specified by means of bootstrap and Proposition 3 gives the MILP reformulation of this case. In particular, running a simulation test, we look into the the effect of the sample size of the random parameter V, the number of the bootstrap resamples, and the critical value determining the size of $\mathfrak{P}_B$. We next apply DPRO to a real-world project investment problem in the passenger vehicle market.

5.1 Numerical Studies

The test is run with a portfolio optimization problem where the feasible region is $\mathfrak{X} := \{x \in [0,1]^8 \mid e_8^T x = 1\}$. We assume that V follows a true Dirichlet distribution (with a variance τ^2) and use the test procedure below to compare the performances of DPRO and a standard SAA model:

1. Generate an *evaluation* sample $\hat{V}^j$ for $j = 1, \ldots, J$, from the Dirichlet distribution ($J = 10,000$ in this test).
2. Generate *training* samples $V^d := \{V^{d,n},\ n = 1, \cdots, N\}$ with a size N for $d = 1, \ldots, D$ from the same distribution.
3. Use each training sample V^d to find an optimal solution x^d of DPRO with $\mathfrak{P}_B(\alpha)$ (the number of the bootstrap resamples is K) and $\tilde{x}^d$ of the SAA model,
$$\max_{x \in \mathfrak{X}} \frac{1}{N} \sum_{n=1}^{N} u\left(x; V^{d,n}\right); \tag{25}$$
use the evaluation sample to assess the performances of x^d and $\tilde{x}^d$ by calculating
$$\phi^d = \frac{1}{J} \sum_{j=1}^{J} u\left(x^d; \hat{V}^j\right), \quad (\psi^d)^2 = \frac{1}{J-1} \sum_{j=1}^{J} \left[u\left(x^d; \hat{V}^j\right) - \phi^d\right]^2,$$
$$\tilde{\phi}^d = \frac{1}{J} \sum_{j=1}^{J} u\left(\tilde{x}^d; \hat{V}^j\right), \quad (\tilde{\psi}^d)^2 = \frac{1}{J-1} \sum_{j=1}^{J} \left[u\left(\tilde{x}^d; \hat{V}^j\right) - \tilde{\phi}^d\right]^2.$$
4. Report $\bar{x} := 1/D \sum_{d=1}^{D} x^d$, $\tilde{x} := 1/D \sum_{d=1}^{D} \tilde{x}^d$, $\sigma := \sqrt{1/(D-1) \sum_{d=1}^{D} \|x^d - \bar{x}\|^2}$, and $\tilde{\sigma} := \sqrt{1/(D-1) \sum_{d=1}^{D} \|\tilde{x}^d - \tilde{x}\|^2}$.
5. Report $\phi := 1/D \sum_{d=1}^{D} \phi^d$, $\tilde{\phi} := 1/D \sum_{d=1}^{D} \tilde{\phi}^d$, $\psi := \sqrt{1/D \sum_{d=1}^{D} (\psi^d)^2}$, and $\tilde{\psi} := \sqrt{1/D \sum_{d=1}^{D} (\tilde{\psi}^d)^2}$.

Using the evaluation sample, the simulation test assesses the performances — the sample means (ϕ^d, $\tilde{\phi}^d$) and sample variances ($(\psi^d)^2$, $(\tilde{\psi}^d)^2$) of utility values — of the solutions (x^d, $\tilde{x}^d$) obtained from solving DPRO and the SAA model (25) with each training sample V^d. Note that, since the evaluation sample size J is very large, we can regard these estimates as true means and variances. In the numerical results reported at steps 4 and 5, $(\bar{x}, \tilde{x})$ are the averages of these solutions based on all the training samples ($V^d,\ d = 1, \ldots, D$) and $(\sigma, \tilde{\sigma})$ show how large these solutions vary. Also, the means $(\phi, \tilde{\phi})$ of the utility values at these solutions estimate the averages of the performances, while the total standard deviations $(\psi, \tilde{\psi})$ of the utility values indicate the variations of the performances. A more robust model is expected to have smaller variations of its solution and performance.

In this test we analyze the effect of the critical value α, the sample size N, the number of the bootstrap resamples K, and the variance of the underlying distribution τ^2. To this end, we begin by testing a baseline case with these parameters fixed and then adjust each parameter separately to observe its marginal effect.

$\bar{x}$	$\tilde{x}$	σ	$\tilde{\sigma}$	ϕ	$\tilde{\phi}$	ψ	$\tilde{\psi}$
[0.216, 0.115, 0.072, 0.150, 0.103, 0.145, 0.131, 0.068]	[0.206, 0.126, 0.070, 0.138, 0.102, 0.145, 0.116, 0.097]	0.058	0.091	0.313	0.328	0.076	0.112

Table 1: Comparative analysis for the baseline case.

Baseline case. We consider the baseline case with $\alpha = 0.10$, $N = 30$, $K = 100$, and $\tau^2 = 0.010$ and present the test results in Table 1. The averages of the solutions of DPRO and the SAA model are $\bar{x} = [0.216, 0.115, 0.072, 0.150, 0.103, 0.145, 0.131, 0.068]$ and $\tilde{x} = [0.206, 0.126, 0.070, 0.138, 0.102, 0.145, 0.116, 0.0$ The result, $\sigma = 0.058 < \tilde{\sigma} = 0.091$, indicates that the solutions of DPRO based on all of the training samples have a smaller variation than those of the SAA model. Moreover, the solutions of DPRO exhibit more stable performance, since the standard deviation of their utility values is smaller, i.e., $\psi = 0.076 < \tilde{\psi} = 0.112$. On the other hand, we observe that $\phi = 0.313 < \tilde{\phi} = 0.328$, which shows the mean of the utility values at the solutions of DPRO is slightly lower. We compare DPRO and the SAA model using the relative gaps of the pairs $(\phi, \tilde{\phi})$ and $(\psi, \tilde{\psi})$, i.e., $(\tilde{\phi} - \phi)/\tilde{\phi}$ and $(\tilde{\psi} - \psi)/\tilde{\psi}$. The relative gaps of $(\phi, \tilde{\phi})$ and $(\psi, \tilde{\psi})$ are 4.6% and 32.1%. As a robust optimization approach, DPRO enhances the stability of the performance at the cost of sacrificing the average moderately.

<table>
<tr><th>α</th><th>$\bar{x}$</th><th>$\tilde{x}$</th><th>σ^2</th><th>$\tilde{\sigma}^2$</th><th>ϕ</th><th>$\tilde{\phi}$</th><th>ψ</th><th>$\tilde{\psi}$</th></tr>
<tr><td>0.05</td><td>[0.214, 0.116, 0.074, 0.150, 0.101, 0.145, 0.135, 0.065]</td><td rowspan="5">[0.206, 0.126, 0.070, 0.138, 0.102, 0.145, 0.116, 0.097]</td><td>0.053</td><td rowspan="5">0.091</td><td>0.281</td><td rowspan="5">0.328</td><td>0.073</td><td rowspan="5">0.112</td></tr>
<tr><td>0.10</td><td>[0.216, 0.115, 0.072, 0.150, 0.103, 0.145, 0.131, 0.068]</td><td>0.058</td><td>0.313</td><td>0.076</td></tr>
<tr><td>0.15</td><td>[0.216, 0.114, 0.073, 0.151, 0.101, 0.145, 0.134, 0.067]</td><td>0.066</td><td>0.317</td><td>0.080</td></tr>
<tr><td>0.20</td><td>[0.215, 0.112, 0.075, 0.151, 0.101, 0.145, 0.135, 0.067]</td><td>0.075</td><td>0.322</td><td>0.091</td></tr>
<tr><td>0.25</td><td>[0.216, 0.114, 0.070, 0.152, 0.101, 0.145, 0.136, 0.066]</td><td>0.088</td><td>0.326</td><td>0.105</td></tr>
</table>

Table 2: Comparative analysis for α.

Effect of the critical value (α). We vary α in the baseline case to test its marginal effect. As we discussed in Section 3.2, a larger α results in a smaller $\mathfrak{P}_B(\alpha)$. Table 2 reports the results as α increases from 0.05 to 0.25. Note that $\alpha = 0.10$ in the baseline case highlighted in gray and the SAA model is irrelevant to α. Therefore, in the table, adjusting α has no impact on $\tilde{x}$, $\tilde{\sigma}$, $\tilde{\phi}$, and $\tilde{\psi}$. The set $\mathfrak{P}_B(\alpha)$ shrinks with the increment in α such that the conservatism in DPRO is weakened obviously. This is confirmed by the increase in σ from 0.053 to 0.088, ϕ from 0.281 to 0.326, and ψ from 0.073 to 0.105. Hence, α determines the balance between the average and the stability of the performance of DPRO. As α increases, the relative gaps of $(\phi, \tilde{\phi})$ and $(\psi, \tilde{\psi})$ reduce from 14.3% to 0.6% for $(\phi, \tilde{\phi})$ and from 34.5% to 6.6% for $(\psi, \tilde{\psi})$. Indeed, DPRO with $\alpha = 0.25$ achieves the similar average of the performance as the SAA model while keeping somewhat stability.

Effect of the number of the bootstrap resamples (K). We next discuss the marginal effect of K shown in Table 3. Similarly, the SAA model is irrelevant to K. Hence, $\tilde{\sigma}$, $\tilde{\phi}$, and $\tilde{\psi}$ keep unchanged in Table 3. With a larger K, we can construct a confidence region more precisely in the bootstrap approach. As a result, both the average and stability of the performance of DPRO should be boosted. We increase K from 30 to 500 in the test. The baseline case chooses $K = 100$.

K	$\bar{x}$	$\tilde{x}$	σ	$\tilde{\sigma}$	ϕ	$\tilde{\phi}$	ψ	$\tilde{\psi}$
30	[0.219, 0.115, 0.074, 0.144, 0.103, 0.141, 0.135, 0.069]		0.088		0.297		0.098	
50	[0.221, 0.116, 0.074, 0.139, 0.103, 0.141, 0.135, 0.071]		0.081		0.306		0.091	
100	[0.216, 0.115, 0.072, 0.150, 0.103, 0.145, 0.131, 0.068]	[0.206, 0.126, 0.070, 0.138, 0.102, 0.145, 0.116, 0.097]	0.058	0.091	0.313	0.328	0.076	0.112
200	[0.216, 0.114, 0.071, 0.151, 0.103, 0.146, 0.131, 0.067]		0.055		0.317		0.071	
500	[0.216, 0.114, 0.072, 0.152, 0.102, 0.145, 0.131, 0.067]		0.051		0.321		0.068	

Table 3: Comparative analysis for K.

The results σ and ψ, indicating the variations of the solution and performance of DPRO, decrease by 0.037 (= 0.088 - 0.051) and 0.030 (= 0.098 - 0.068), respectively. Meanwhile, the average of the performance, ϕ, increases by 0.024 (= 0.321 - 0.297). For DPRO with $K = 500$ and the SAA model, the relative gaps of $(\phi, \tilde{\phi})$ and $(\psi, \tilde{\psi})$ are 2.1% and 39.5%. It means that, when $K = 500$ in this test, DPRO has much more stable performance than the SAA model while achieving the similar average.

N	$\bar{x}$	$\tilde{x}$	σ	$\tilde{\sigma}$	ϕ	$\tilde{\phi}$	ψ	$\tilde{\psi}$
10	[0.213, 0.119, 0.071, 0.144, 0.101, 0.149, 0.133, 0.071]	[0.221, 0.121, 0.070, 0.133, 0.099, 0.139, 0.121, 0.096]	0.079	0.126	0.281	0.296	0.107	0.140
30	[0.216, 0.115, 0.072, 0.150, 0.103, 0.145, 0.131, 0.068]	[0.206, 0.126, 0.070, 0.138, 0.102, 0.145, 0.116, 0.097]	0.058	0.091	0.313	0.328	0.076	0.112
50	[0.215, 0.113, 0.073, 0.151, 0.102, 0.145, 0.133, 0.068]	[0.205, 0.129, 0.073, 0.133, 0.099, 0.152, 0.114, 0.096]	0.047	0.072	0.322	0.329	0.060	0.104
100	[0.216, 0.115, 0.072, 0.151, 0.102, 0.145, 0.132, 0.067]	[0.204, 0.123, 0.072, 0.135, 0.101, 0.151, 0.118, 0.096]	0.040	0.068	0.328	0.331	0.053	0.098
1000	[0.216, 0.113, 0.073, 0.151, 0.103, 0.145, 0.133, 0.067]	[0.205, 0.124, 0.073, 0.134, 0.101, 0.151, 0.115, 0.095]	0.037	0.037	0.336	0.334	0.048	0.049

Table 4: Comparative analysis for N.

Effect of the sample size (N). A larger sample size N provides more information on the true distribution and consequently enhances the performances of both DPRO and the SAA model. Table 4 gives the test results when N varies from 10 to 1000 with $N = 30$ for the baseline case. We can see that ϕ and $\tilde{\phi}$ increase by 0.055 (= 0.336 - 0.281) and 0.038 (= 0.334 - 0.296), ψ and $\tilde{\psi}$ decrease by 0.059 (= 0.107 - 0.048) and 0.091 (= 0.140 - 0.049), and σ and $\tilde{\sigma}$ reduce by 0.042 (= 0.079 - 0.037) and 0.089 (= 0.126 - 0.037). We compare DPRO and the SAA model using the two extreme cases with N being 10 or 1000. The relative gaps of $(\phi, \tilde{\phi})$ and $(\psi, \tilde{\psi})$ are 5.1% and 23.8% for $N = 10$ and reduce to -0.5% and 2% when N is creased to 1000. Obviously, DPRO is a better choice in the case when a small size of sample is unavailable to recover the true probability distribution. In contrast, when there is sufficient data for well estimating the distribution, DPRO loses the advantage of robustness. In Section 6.3 we will address the convergence of DPRO as N increases to ∞ from theoretical perspective.

Effect of the variance of the underlying distribution (τ^2). While a small sample size N incurs input uncertainty, variance τ^2 determines the level of stochastic uncertainty, which is inherent to the stochastic nature of the problem (see Corlu et al. (2020) and therein). We now test the impact of stochastic uncertainty on the performances of DPRO and the SAA model by adjusting τ^2 and keeping $N = 30$. The results in Table 5 show that, for both DPRO and the SAA model, the stability and average of the performance are improved as τ^2 varies from 0.02 to 0.005. The reason is that training samples of equal size can better estimate the underlying distribution under lower

Variance	$\bar{x}$	$\tilde{x}$	σ	$\tilde{\sigma}$	ϕ	$\tilde{\phi}$	ψ	$\tilde{\psi}$
0.020	[0.210, 0.125, 0.072, 0.141, 0.103, 0.145, 0.131, 0.073]	[0.215, 0.122, 0.073, 0.133, 0.103, 0.139, 0.119, 0.096]	0.066	0.116	0.274	0.289	0.101	0.133
0.010	[0.216, 0.115, 0.072, 0.150, 0.103, 0.145, 0.131, 0.068]	[0.206, 0.126, 0.070, 0.138, 0.102, 0.145, 0.116, 0.097]	0.058	0.091	0.313	0.328	0.076	0.112
0.005	[0.219, 0.111, 0.072, 0.151, 0.104, 0.144, 0.133, 0.066]	[0.201, 0.129, 0.075, 0.133, 0.099, 0.152, 0.114, 0.097]	0.051	0.084	0.328	0.332	0.066	0.093

Table 5: Comparative analysis for τ^2.

stochastic uncertainty. As τ^2 decreases, ϕ and $\tilde{\phi}$ increase by 0.054 (= 0.328 - 0.274) and 0.058 (= 0.332 - 0.274), ψ and $\tilde{\psi}$ decrease by 0.035 (= 0.101 - 0.066) and 0.060 (= 0.133 - 0.093), and σ and $\tilde{\sigma}$ reduce by 0.015 (= 0.066 - 0.051) and 0.032 (= 0.116 - 0.084). The relative gaps of $(\phi, \tilde{\phi})$ and $(\psi, \tilde{\psi})$ are 5.2% and 24.1% for $\tau^2 = 0.02$, 4.6% and 32.1% for $\tau^2 = 0.01$, and 1.2% and 29% for $\tau^2 = 0.005$. From the results, we may conclude that, even though the stochastic uncertainty is rather low for a small τ^2, the existence of the large input uncertainty makes DPRO more practicable.

5.2 Project Investment

We now illustrate how to combine DPRO with machine learning in a project investment problem. In this problem, an automotive manufacturer needs to learn consumer preference towards passenger vehicles, and on this basis, to pursue the best portfolio investment with a fixed budget among the 10 projects listed in Table 6. These candidate projects include

- *safety promotion:* design a new structure to alleviate the impacts of the collision;
- *new car model development:* create a concept car with the aim of a more fashion style and higher market acceptance before actually producing it;
- *engine upgrade:* upgrade the current engine and its vibration sensor system to achieve more horsepower and less engine noise;
- *e-platform development:* improve the human-vehicle interaction experience and visualize the performances of the car;
- *computational fluid dynamics (CFD) testing system development:* implement related fluid mechanics and numerical analysis into a testing platform to analyze the performance of concept cars;
- *common modular platform (CMP) development:* develop a platform used for subcompact and compact car models with internal combustion engine and battery-electric cars;
- *checking fixture promotion:* enhance the performance of checking fixture to control the dimensions of auto parts (such as trim edge,surface profile, flatness, etc.) in a more convenient way for mass production of parts detection;
- *noise, vibration, harshness (NVH) digitalization:* incorporate the characteristics of the noise, vibration and harshness of vehicles to a digital platform with the target of more efficient computation in evaluating the driver satisfactions;
- *driving assistance system development:* incorporate the latest interface standards and running multiple vision-based algorithms to support real-time multimedia, vision co-processing, and sensor fusion subsystems;

- *digitalization of marketing network:* build a digital platform used for all the dealers across different regions to share the customer resources, the service standards and the marketing strategies.

These projects may lead to enhancement of eight attributes related to vehicle performance, economy, after-sales service, etc, among which the attributes including *sale price, fuel consumption, depreciation rate (yearly), and estimated maintain and repair (M&R) fee (in five years)* are related to consumers' major economic considerations when purchasing new cars, *wheel base, acceleration, comfort rating* are related to consumers' major consideration of vehicle performance, and *dealership*, represented by the number of dealers able to deal with certain services, directly influences consumers' satisfactions on after-sale services. Table 6 displays the estimated consequences and costs of the projects. The column of "Base model" gives the baseline attribute values before the investment, of which the vector is denoted by x^0, while the other columns show the vectors of attribute increment values, denoted by y^d for $d = 1, \ldots, 10$, after the investment on the projects.

	Base model (x^0)	Safety promotion (y^1)	New car model development (y^2)	Engine upgrade (y^3)	E-platform development (y^4)	CFD testing system development (y^5)	CMP development (y^6)	Checking fixture promotion (y^7)	Noise, vibration, harshness (NVH) digitalization (y^8)	Driving assistance system development (y^9)	Digitalization of marketing network (y^{10})
Price ($k)	38	+7	-1	0	0	0	-1.8	0	+1.5	+1.5	0
Fuel consumption (MPG)	30	0	-8	-2	0	0	0	0	0	-2	0
Wheelbase (in)	110	0	-2	0	0	0	-3	0	0	0	0
Acceleration (0-60 miles, sec)	8	0	-1	-1	0	-1.5	0	0	0	-1	0
Comfort rating (0-5)	3.8	+0.02	+0.04	0	+0.08	+0.04	0	+0.02	+0.08	+0.02	+0.03
Dealership (# of dealers)	1050	0	+150	+150	+150	0	0	0	0	0	+150
Depreciation rate (0-1)	0.25	-0.08	-0.03	-0.05	0	0	-0.03	0	-0.03	-0.02	-0.05
M&R fee ($K)	5.5	-0.5	-0.5	+0.3	+0.3	-0.5	0	-0.2	-0.5	-0.2	-0.5
Project costs ($ million)	—	50	100	70	40	20	30	15	80	50	70

Table 6: The list of candidate projects.

Denote by $z \in \{0,1\}^{10}$ the decision vector of which a component $z_d = 1$ means the project d is selected in the investment and otherwise $z_d = 0$. Assume that the improvement on the attributes yielded by each project is independent of the one by any other project. As a result, the attribute values achieved in the investment are obtained as

$$\hat{x}(z) = x^0 + \sum_{d=1}^{10} z_d y^d. \tag{26}$$

Let Φ be the budget limit of the investment and b_d the cost of project d given in Table 6. A budget constraint is thus described as

$$\sum_{d=1}^{10} b_d z_d \le \Phi. \tag{27}$$

Combining (26) and (27), we formulate the feasible region of the attribute value achieved by the investment as

$$\mathfrak{X} = \{\hat{x}(z) \mid z \in \{0,1\}^{10} \text{ satisfies (27)}\}. \tag{28}$$

Consequently, DPRO in this case is specified as

$$\max_{x \in \mathfrak{X}} \min_{P \in \mathfrak{P}_B(\alpha)} \mathbb{E}_P[u(x;V)], \tag{29}$$

where $u(\cdot;V)$ represents the uncertain consumer preference and $\mathfrak{P}_B(\alpha)$ is the ambiguity set of the distribution specified by means of bootstrap given in (12).

Attributes	$t_{m,1}$	$t_{m,2}$	$t_{m,3}$	$t_{m,4}$	$t_{m,5}$	$t_{m,6}$	$t_{m,7}$	$t_{m,8}$
Price ($)	50,000	43,120	38,900	33,750	29,256	25,065	21,970	15,895
Fuel consumption (MPG)	37	35	33	31	29	27	20	-
Wheelbase (in)	97.5	103.5	105.9	107.8	109.3	110.6	-	-
Acceleration (0-60 miles, sec)	9.00	8.30	7.90	7.00	6.30	5.90	5.30	4.80
Comfort (0-5 rating)	3.2	3.5	4.0	4.5	4.6	4.7	5.0	-
Dealership (# of Dealers)	895	1075	1500	1873	2393	2708	2919	3120
Depreciation (Rate:0-1)	0.191	0.184	0.175	0.166	0.157	0.100	-	-
M&R fee ($)	5,190	4,933	4,719	4,353	3,927	3,558	3,219	3,000

Table 7: The control points of attributes*.

* For the attributes such as *Price, Fuel Consumption, Acceleration, Depreciation Rate, and M&R fee*, a higher value means lower utility value. Therefore, we list the breakpoints of these attributes in descending order in the table and use the negative values of them as the breakpoints in the numerical tests.

We next generate a random sample of V using a logistic regression method and construct $\mathfrak{P}_B(\alpha)$. The logistic regression method is developed in Feng (2016) to assess the consumer preference toward passenger vehicles in the US market. The dependent variable of the model is the market share of the car industry in years 2013 and 2014, which is regarded as the probability distribution of consumers' choices among major vehicle brands. The independent variables are the performances of those brands regarding the attributes listed in Table 6. Following the approach in Feng (2016), we select a set of representative car brands and survey the monthly market shares and attribute values for each brand during two years. Let S_j^ℓ be the market share of brand j at the ℓ-th month, x^j be the vector of the attribute values of brand j, and $u(\cdot;v^\ell)$, the utility function with the parameter vector

v^ℓ, represent the consumer preference at the ℓ-th month. The probability that consumers purchase brand j in month ℓ is

$$S_j^\ell = \frac{\exp(\eta u(x^j; v^\ell))}{\sum_{i=1}^J \exp(\eta u(x^i; v^\ell))}, \tag{30}$$

where η is the parameter ensuring the normalization of u. Table 7 gives the control/breakpoints of u defined in (1) and selected in this study. Estimating the parameters η and v^ℓ in the logistic model (30) using the survey data, we generate a random sample of V including 24 month-wise observations, i.e., the estimate of v^ℓ for $\ell = 1, \ldots, 24$. The components of the sample average $\bar{V}$ are given in Table 8.

Attributes	$\bar{V}_{m,1}$	$\bar{V}_{m,2}$	$\bar{V}_{m,3}$	$\bar{V}_{m,4}$	$\bar{V}_{m,5}$	$\bar{V}_{m,6}$	$\bar{V}_{m,7}$
Price	0.477	0.006	0.031	0.093	0.174	0.088	0.131
Fuel consumption	0.292	0.151	0.091	0.098	0.107	0.261	-
Wheelbase	0.792	0.040	0.050	0.056	0.062	-	-
Acceleration	0.317	0.072	0.053	0.041	0.135	0.124	0.123
Comfort	0.005	0.027	0.093	0.522	0.042	0.218	-
Dealership	0.405	0.069	0.025	0.021	0.468	0.008	0.004
Depreciation rate	0.008	0.024	0.132	0.090	0.746	-	-
Maintenance and repair fee	0.012	0.020	0.054	0.730	0.056	0.070	0.058

Table 8: The sample mean $\bar{V}_{m,k}$ at the break points.

In the study we choose the number of bootstrap samples $K = 100$ and the critical value $\alpha = 0.10$ in the setting of $\mathfrak{P}_B(\alpha)$. The budget limit Φ is varied from \$100 million to \$300 million. Table 9 presents the results of model (29) showing the best investment portfolio and the optimal value of consumer preference. When the budget limit is \$100 million, we select the two most valuable projects, "*CFD testing system development*" and "*NVH digitalization*". The optimal results shows that by incorporating these two projects, despite of slightly higher price (by \$ 1.5K), the depreciation rate and the M&R fee both can be significantly reduced by more than 10%, the acceleration performance can be enhanced more than 10% and the comfort rating can be increased from 3.80 to 3.92. These observations may imply that the reliability, car performance, and consumers' satisfaction are the crucial factors influencing the utility. As the budget limit increases, more projects can be selected. When the budget limit increases to \$200 million, we select two additional projects, "*Common modular platform development*" and "*Digitalization of marketing network* ". From the result in Table 9, we can find that these newly included projects alleviate the depreciation rate to 0.19 and the price to \$ 37.7K which are both economic factors. The project "*New model development*" is added into the portfolio with the budget limit of \$300 million. The new projects supported by this additional investment on the one hand keep improving the economic factors (lower price and lower depreciation rate), and also enhancing the car performance (higher acceleration and less fuel consumption). On the other hand, the investment starts to influence the after-sale network with more dealerships, which also explain the the decrease in M&R fees.

6 Piecewise Linear Approximation of General Random Utility Functions

Our discussions in the preceding sections are based on the assumption that the DM's true utility function of each attribute has a piecewise linear structure with fixed breakpoints. A random utility

Budget Limit ($ million)	Safety promotion	New car model development	Engine upgrade	E-platform development	CFD testing system development	CMP development	Checking fixture promotion	NVH digitalization	Driving assistance system development	Digitalization of marketing network
100 (Optimal solution z)	$z_1 = 0$	$z_2 = 0$	$z_3 = 0$	$z_4 = 0$	$z_5 = 1$	$z_6 = 0$	$z_7 = 0$	$z_8 = 1$	$z_9 = 0$	$z_{10} = 0$
100 (Attribute values $\hat{x}(z)$)	Price \$ 39.5K, Fuel Consumption 30MPG, Wheelbase 110in, Acceleration 6.5 (0-60miles, sec) Comfort rating 3.92, Dealership 1050, Depreciation rate 0.22, M&R fee \$ 4.5K									
100 (Optimal value)	0.5991									
200 (Optimal solution z)	$z_1 = 0$	$z_2 = 0$	$z_3 = 0$	$z_4 = 0$	$z_5 = 1$	$z_6 = 1$	$z_7 = 0$	$z_8 = 1$	$z_9 = 0$	$z_{10} = 1$
200 (Attribute values $\hat{x}(z)$)	Price \$ 37.7K, Fuel Consumption 30MPG, Wheelbase 107in, Acceleration 6.5 (0-60miles, sec) Comfort rating 3.95, Dealership 1200, Depreciation rate 0.14, M&R fee \$ 4.0K									
200 (Optimal value)	0.6318									
300 (Optimal solution z)	$z_1 = 0$	$z_2 = 1$	$z_3 = 0$	$z_4 = 0$	$z_5 = 1$	$z_6 = 1$	$z_7 = 0$	$z_8 = 1$	$z_9 = 0$	$z_{10} = 1$
300 (Attribute values $\hat{x}(z)$)	Price \$ 36.7K, Fuel Consumption 22MPG, Wheelbase 105in, Acceleration 5.5 (0-60miles, sec) Comfort rating 3.99, Dealership 1350, Depreciation rate 0.11, M&R fee \$ 3.5K									
300 (Optimal value)	0.6519									
x^0	Price \$ 38.0K, Fuel Consumption 30MPG, Wheelbase 110in, Acceleration 8 (0-60miles, sec) Comfort rating 3.80, Dealership 1050, Depreciation rate 0.25, M&R fee \$ 5.5K									

Table 9: The optimal solutions and values of model (29) with different budget limits.

function can then be formulated as $u(\cdot; V)$. In this section, we relax the assumption by considering a general random utility function which is merely continuous and investigate its approximation by random piecewise linear utility functions. Specifically we consider the following expected utility maximization problem

$$\max_{x \in \mathfrak{X}} \mathbb{E}[U(x)], \tag{31}$$

where $U(x) = \sum_{m \in \mathfrak{M}} U_m(x)$, U_m is a continuous nondecreasing random marginal utility function defined on the domain $[a_m, b_m]$ for attribute $m \in \mathfrak{M}$ and $\mathfrak{X} \subseteq \otimes_{m \in \mathfrak{M}} [a_m, b_m] = [a_1, b_1] \times \cdots \times [a_m, b_m]$. To ease the exposition, we assume that U is normalized with $U(a_1, \ldots, a_M) = 0$ and $U(b_1, \ldots, b_M) = 1$ almost surely.

6.1 Static Piecewise Linear Approximation

We begin by discussing piecewise linear approximation of $U(x)$ and its impact on the optimal value and optimal solutions. For each fixed $m \in \mathfrak{M}$, let $t_{m,i}$, for $i \in \mathfrak{I}_m$, be defined as in (1). Let

$$V_{m,i} := U_m(t_{m,i}) - U_m(t_{m,i-1}), \qquad i \in \mathfrak{I}_m,$$

be the increment of U_m over interval $[t_{m,i-1}, t_{m,i}]$ and $V_m = (V_{m,i}, \cdots, V_{m,I_m})'$. We construct a piecewise utility function over $[a_m, b_m]$, denoted by $u_m(\cdot; V_m)$, with breakpoints $t_{m,i}$, $i \in \mathfrak{I}_m$, and

$$u_m(t_{m,i}; V_m) = U_m(t_{m,i}), \quad \text{for } i \in \mathfrak{I}_m \tag{32}$$

and use $u_m(\cdot; V_m)$ to approximate U_m for $m = 1, \cdots, M$. Let $V := (V_1', \cdots, V_M')'$ and

$$u(x; V) := \sum_{m \in \mathfrak{M}} u_m(x_m; V_m).$$

Note that the dimension of vector V is $I := \sum_{m=1}^M I_m$. We propose to obtain an approximate optimal value and optimal solution of problem (31) by solving the following piecewise linear approximated expected utility maximization problem:

$$\max_{x \in \mathfrak{X}} \mathbb{E}[u(x; V)]. \tag{33}$$

Let ϑ denote the optimal value of problem (31) and ϑ_I the optimal value of problem (33). Let X^* and X_I be the respective sets of optimal solutions. The subscript I indicates the optimal value depends on the total number of breakpoints I in the piecewise linear approximation. Of course, it also depends on the location of these points. Let

$$\Delta := \max_{m\in\mathfrak{M}} \max_{i\in\mathfrak{I}_m}(t_{m,i} - t_{m,i-1}). \tag{34}$$

Obviously in order to secure a good approximation of ϑ by ϑ_I, we need Δ to be sufficiently small. In the case when the breakpoints are evenly spread, this is equivalent to setting I a large value. Unless specified otherwise, we assume in the rest of discussions that $I \to \infty$ ensures $\Delta \to 0$. The next proposition addresses the approximation of problem (31) by problem (33) in terms of the optimal value and optimal solutions.

Proposition 6 *Suppose that for $m \in \mathfrak{M}$, U_m is Lipschitz continuous over $[a_m, b_m]$ with random Lipschitz modulus L_m almost surely, where $\mathbb{E}[L_m^p] < \infty$ for some $p > 0$ and the expectation is taken w.r.t the probability distribution of the random factors underlying $U(x)$. Then the following assertions hold.*

(i) For any $\epsilon > 0$,

$$\text{Prob}\left(|\vartheta - \vartheta_I| \geq \epsilon\right) \leq \text{Prob}\left(\sup_{x\in\mathfrak{X}} |u(x;V) - U(x)| \geq \epsilon\right) \leq \left(\frac{\Delta}{\epsilon}\right)^p \mathbb{E}[L^p], \tag{35}$$

where $L := \sum_{m\in\mathfrak{M}} L_m$. Prob stands for probability.

(ii) Let $\{x_I\}$ be a sequence of optimal solutions obtained from solving problem (33). Then every cluster point of the sequence is an optimal solution of problem (31), that is,

$$\lim_{I\to\infty} \text{Prob}(\mathbb{D}(X_I, X^*) \geq \epsilon) = 0, \tag{36}$$

where $\mathbb{D}(A, B)$ denotes the access distance of set A over set B.

Proof: Part (ii) follows directly from Part (i) and the well-known stability results in parametric programming, see e.g. (Liu and Xu 2013, Lemma 3.8). We only prove Part (i). The first inequality follows from the fact that

$$|\vartheta - \vartheta_I| \leq \sup_{x\in\mathfrak{X}} |u(x;V) - U(x)|.$$

Thus we only prove the second inequality. By the monotonicity, Lipschtz continuity of the utility function, and equality (32),

$$|u_m(x_m;V_m) - U_m(x_m)| \leq |U_m(t_{m,i}) - U_m(t_{m,i-1}))| \leq L_m|t_{m,i} - t_{m,i-1}| \leq L_m\Delta,$$

for any $x_m \in [t_{m,i-1}, t_{m,i}]$, which ensures that

$$\sup_{x_m\in[a_m,b_m]} |u_m(x_m;V_m) - U_m(x_m)| \leq L_m\Delta,$$

for $m \in \mathfrak{M}$ and hence,

$$\sup_{x\in\mathfrak{X}} |u(x;V) - U(x)| \leq \sup_{x\in\otimes_{m\in\mathfrak{M}}[a_m,b_m]} |u(x;V) - U(x)| \leq \sum_{m\in\mathfrak{M}} L_m\Delta = L\Delta.$$

Consequently by the Markov inequality

$$\text{Prob}\left(\sup_{x\in\mathfrak{X}}|u(x;V)-U(x)|\geq\epsilon\right)\leq\text{Prob}\left(L\Delta\geq\epsilon\right)\leq\left(\frac{\Delta}{\epsilon}\right)^{p}\mathbb{E}[L^p].$$

The proof is complete. □

Inequality (35) may be used to estimate Δ or I for a specified probability δ, i.e., when

$$\Delta<\epsilon\left(\frac{\delta}{\mathbb{E}[L^p]}\right)^{\frac{1}{p}}, \tag{37}$$

we will have

$$\text{Prob}\left(\sup_{x\in\mathfrak{X}}|u(x;V)-U(x)|\geq\epsilon\right)\leq\delta.$$

The result means that when Δ satisfies condition (37), there is a probability of $1-\delta$ such that the error arising from the piecewise linear approximation of the random utility function is less or equal to ϵ. The result provides a theoretical guarantee for using the piecewise linear utility function model in absence of complete information on the structure of the true random utility function.

6.2 Sample Average Approximation

In practice, the true probability distribution of V is often unknown, but it is possible to obtain some observations from empirical data or computer simulation. Let $V^1,\cdots,V^N$ denote an iid random sample of V and $\bar{V}$ the sample mean. By Proposition 1, we propose to approximate $\mathbb{E}[U(\cdot)]$ using the sample average

$$\frac{1}{N}\sum_{n=1}^{N}u(\cdot;V^n)=\frac{1}{N}\sum_{n=1}^{N}f(\cdot)^TV^n=f(\cdot)^T\bar{V}=u(\cdot;\bar{V}).$$

The next proposition gives a qualitative description of such approximation.

Proposition 7 *Under the settings and conditions of Proposition 6, for any $\epsilon>0$ and $\delta>0$, there exists $N_0>0$ such that*

$$\text{Prob}\left(\sup_{x\in\mathfrak{X}}|u(x;\bar{V})-\mathbb{E}[U(x)]|\geq\epsilon\right)\leq\delta, \tag{38}$$

for all $N\geq N_0$ and $\Delta\leq\frac{\epsilon}{2\mathbb{E}[L]}$, where Δ is defined as in Proposition 6.

Proof: We write $\bar{V}=(\bar{V}_1',\ldots,\bar{V}_M')'$, where $\bar{V}_m$ is the associated sample average for the vector of increments regarding the marginal utility function of attribute m. By the triangle inequality, we have

$$|u_m(x;\bar{V}_m)-\mathbb{E}[U_m(x)]|\leq|u_m(x;\bar{V}_m)-\mathbb{E}[u_m(x;V_m)]|+|\mathbb{E}[u_m(x;V_m)]-\mathbb{E}[U_m(x)]|,\ \text{for } m=1,\cdots,M.$$

By the monotonicity, Lipschtz continuity of the utility function, and equality (32),

$$|u_m(x_m;V_m)-U_m(x_m)|\leq|U_m(t_{m,i})-U_m(t_{m,i-1}))|\leq L_m|t_{m,i}-t_{m,i-1}|\leq L_m\Delta,\forall x_m\in[t_{m,i-1},t_{m,i}],$$

which implies that

$$\sup_{x_m\in[a_m,b_m]}|\mathbb{E}[u_m(x;V_m)]-\mathbb{E}[U_m(x)]|\leq\sup_{x_m\in[a_m,b_m]}\mathbb{E}|u_m(x;V_m)-U_m(x)|\leq\mathbb{E}[L_m]\Delta.$$

Thus

$$\sup_{x\in\mathfrak{X}} |\mathbb{E}[u(x;V)] - \mathbb{E}[U(x)]| \leq \mathbb{E}\left[\sum_{m\in\mathfrak{M}} \sup_{x_m\in[a_m,b_m]} |u_m(x;V_m) - U_m(x)|\right] \leq \mathbb{E}[L]\Delta. \tag{39}$$

Let Δ be sufficiently small such that $\mathbb{E}[L]\Delta \leq \epsilon/2$. By Proposition 1, we have

$$\begin{aligned}\text{Prob}\left(\sup_{x\in\mathfrak{X}} |u(x;\bar{V}) - \mathbb{E}[U(x)]| \geq \epsilon\right) &\leq \text{Prob}\left(\sup_{x\in\mathfrak{X}} |u(x;\bar{V}) - \mathbb{E}[u(x;V)]| \geq \frac{\epsilon}{2}\right) \\ &\leq \text{Prob}\left(\|\bar{V} - \mathbb{E}[V]\| \sup_{x\in\mathfrak{X}} \|f(x)\| \geq \frac{\epsilon}{2}\right).\end{aligned}$$

By the definition of f in Proposition 1, we know that $\sup_{x\in\mathfrak{X}} \|f(x)\| < \infty$. Thus by Cramér's large deviation theorem, there exists a positive integer N_0 and positive constant $\Upsilon(\epsilon)$ (depending on ϵ with $\Upsilon(0) = 0$) such that for all $N \geq N_0$

$$\text{Prob}\left(\|\bar{V} - \mathbb{E}[V]\| \sup_{x\in\mathfrak{X}} \|f(x)\| \geq \frac{\epsilon}{2}\right) \leq e^{-\Upsilon(\epsilon)N}. \tag{40}$$

For fixed $\delta \in (0,1)$, we can set $N_0(\epsilon,\delta) := -\frac{\ln\delta}{\Upsilon(\epsilon)}$ and subsequently obtain

$$\text{Prob}\left(\|\bar{V} - \mathbb{E}[V]\| \sup_{x\in\mathfrak{X}} \|f(x)\| \geq \frac{\epsilon}{2}\right) \leq \delta \tag{41}$$

for all $N \geq N_0(\epsilon,\delta)$. The proof is complete. □

With the theoretical justification of $u(x;\bar{V})$ to $\mathbb{E}[U(x)]$, we consider the sample data based utility maximization problem

$$\max_{x\in\mathfrak{X}} u(x;\bar{V}). \tag{42}$$

Let $\vartheta_{N,I}$ denote the optimal value of problem (42) and $X_{N,I}$ the set of the optimal solutions. Here the subscripts N and I are used to indicate that these quantities depend on both the sample size N and the number of breakpoints I. By combining Propositions 6 and 7, we can establish the convergence of $u(x;\bar{V})$ to $\mathbb{E}[U(x)]$ and associated optimal solutions as both N and I go to infinity. The next proposition addresses this.

Proposition 8 *Assume the setting and conditions of Propositions 6 and 7. Then for any positive number ϵ,*

$$\lim_{N,I\to\infty} \text{Prob}\left(|\vartheta_{N,I} - \vartheta| \geq \epsilon\right) = \lim_{N,I\to\infty} \text{Prob}\left(\sup_{x\in\mathfrak{X}} |u(x;\bar{V}) - \mathbb{E}[U(x)]| \geq \epsilon\right) = 0 \tag{43}$$

and

$$\lim_{N,I\to\infty} \text{Prob}(\mathbb{D}(X_{N,I}, X^*) \geq \epsilon) = 0, \tag{44}$$

Proof. Equality (43) follows from (35) and (38) whereas (44) follows from (43) and classical stability results (e.g. (Liu and Xu 2013, Lemma 3.8)). □

6.3 Convergence of DPRO

The convergence results established in Proposition 8 is based on the assumption that the sample size N can be arbitrarily large. In some practical data-driven problems, this assumption may not be fulfilled. This motivates us to consider DPRO. In Section 3.2, we construct either an ellipsoid moment region $\mathfrak{P}_A$ in (8) or bootstrap confidence region $\mathfrak{P}_B$ in (12). We first discuss DPRO with $\mathfrak{P}_A$ and next address $\mathfrak{P}_B$.

For the convenience of reading, we repeat the definition of $\mathfrak{P}_A(\gamma_N)$ below:

$$\mathfrak{P}_A(\gamma_N) := \left\{ P \in \mathscr{P}(\mathbb{R}^I) \;\middle|\; \begin{array}{l} P(V \in \mathfrak{V}) = 1 \\ \left\| S_{<-I>}^{-1/2} \left(C\mathbb{E}_P[V] - \bar{V}_{<-I>} \right) \right\|^2 \leq \gamma_N \end{array} \right\}, \tag{45}$$

We follow Corollary 1 in Delage and Ye (2010) to set γ_N which guarantees that with probability $1-\alpha$, the true mean value lies with in the specified region. The next proposition states this.

Proposition 9 *Under Assumption 1,*

$$\text{Prob}\left(\left\| \Sigma_{<-I>}^{-1/2} (C\mathbb{E}[V] - \bar{V}_{<-I>}) \right\|^2 \leq \gamma_N^{(1)} \right) \geq 1 - \alpha,$$

where $\gamma_N^{(1)} := (R(I-1))^2(2 + \sqrt{2\ln(1/\alpha)})^2/N$ *and* $R := \left\| \Sigma_{<-I>}^{-1/2} \right\|$.

Proof: It suffices to verify the condition of Corollary 1 in Delage and Ye (2010). Let $\mu_{<-I>} := C\mathbb{E}[V]$. Since the support of V is the simplex $\mathfrak{V}$, then $\|\bar{V}_{<-I>} - \mu_{<-I>}\| \leq I - 1$ almost surely. Moreover, Assumption 1 guarantees $R < \infty$. Thus

$$\left\| \Sigma_{<-I>}^{-1/2} (\bar{V}_{<-I>} - \mu_{<-I>}) \right\| \leq \left\| \Sigma_{<-I>}^{-1/2} \right\| \, \left\| \bar{V}_{<-I>} - \mu_{<-I>} \right\| \leq (I-1)R, \text{ a.s.}$$

which verifies the condition required in Corollary 1 in Delage and Ye (2010). □

Our next proposition states conditions under which the true probability distribution of V falls in the ambiguity set $\mathfrak{P}_A(\gamma_N)$ by adjusting γ_N according to the sample size N and confidence level $1-\alpha$.

Proposition 10 *For a given* $\alpha \in (0, e^{-2}(2 - e^{-2}))$*, let*

$$\gamma_N^{(2)} = \frac{8(I-1)^2 R^2 e^2 \ln^2(1/(1 - \sqrt{1-\alpha}))}{N}, \tag{46}$$

where R is defined in Proposition 9. Then there exists a positive integer N_0 depending on α such that, for the true probability distribution P of the random vector V,

$$\text{Prob}\left(P \in \mathfrak{P}_A(\gamma_N^{(2)}) \right) \geq 1 - \alpha$$

for all $N \geq N_0$.

Proof. Let $\beta := 1 - \sqrt{1-\alpha}$ and $\hat{\Sigma}_{<-I>} = \Sigma_{<-I>}/2$. Under Assumption 1, we have

$$\left\| \hat{\Sigma}_{<-I>}^{-1/2} (C\mathbb{E}[V] - \bar{V}_{<-I>}) \right\| \leq \sqrt{2}(I-1)R < \infty, \text{ a.s.}$$

Hence, the bounded support guarantees the "Condition (G)" in So (2011), i.e., for any $p \geq 1$,

$$\mathbb{E}\left[\left\| \hat{\Sigma}_{<-I>}^{-1/2} (C\mathbb{E}[V] - \bar{V}_{<-I>}) \right\|^p \right] \leq [2(I-1)^2 R^2 p]^{p/2}.$$

It follows by Proposition 4 in So (2011) that

$$\text{Prob}\left(\left\|\hat{\Sigma}_{<-I>}^{-1/2}(C\mathbb{E}[V] - \bar{V}_{<-I>})\right\| \le \gamma_N^{(2)}\right) \ge 1-\beta.$$

In addition, by the weak law of large number, there exists N_0 depending on α such that

$$\text{Prob}\left(S_{<-I>} \succ \hat{\Sigma}_{<-I>}\right) \ge 1-\beta,$$

for all $N \ge N_0$. Consequently,

$$\begin{aligned}
& \text{Prob}\left(\left\|S_{<-I>}^{-1/2}(C\mathbb{E}[V] - \bar{V}_{<-I>})\right\| \le \gamma_N^{(2)}\right) \\
\ge\ & \text{Prob}\left(\left\|S_{<-I>}^{-1/2}(C\mathbb{E}[V] - \bar{V}_{<-I>})\right\| \le \gamma_N^{(2)} \;\middle|\; S_{<-I>} \succ \hat{\Sigma}_{<-I>}\right) \text{Prob}\left(S_{<-I>} \succ \hat{\Sigma}_{<-I>}\right) \\
\ge\ & \text{Prob}\left(\left\|\hat{\Sigma}_{<-I>}^{-1/2}(C\mathbb{E}[V] - \bar{V}_{<-I>})\right\| \le \gamma_N^{(2)}\right) \text{Prob}\left(S_{<-I>} \succ \hat{\Sigma}_{<-I>}\right) \\
\ge\ & (1-\beta)^2 = 1-\alpha,
\end{aligned}$$

which completes the proof. □

By Proposition 1, we can rewrite DPRO with $\mathfrak{P}_A(\gamma_N)$ equivalently as

$$\max_{x\in\mathfrak{X}} \min_{v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N))} u(x;v). \tag{47}$$

where

$$\mathfrak{F}(\mathfrak{P}_A(\gamma_N)) := \left\{v \in \mathfrak{V} \mid (Cv - \bar{V}_{<-I>})^T S_{<-I>}^{-1}(Cv - \bar{V}_{<-I>}) \le \gamma_N\right\}. \tag{48}$$

Let γ_N decrease to 0 as N increases to ∞ like $\gamma_N^{(2)}$ defined in (46). Then the ambiguity set $\mathfrak{F}(\mathfrak{P}_A(\gamma_N))$ converges to the singleton $\{\mu = \mathbb{E}[V]\}$. Let ϑ_{DPRO} denote the optimal value of the maximin problem (47). We investigate the convergence of ϑ_{DPRO} to ϑ, the optimal value of problem (31), as N and I increase to ∞ and γ_N goes to 0. The following theorem addresses this.

Theorem 1 (DPRO with Ellipsoid ambiguity (48)) *Let γ_N in (45) monotonically decrease to 0 as $N \to \infty$. For any $\alpha \in (0,1)$ and small $\epsilon > 0$, there exist positive numbers I_0, N_0 such that*

$$\text{Prob}\left(\|\vartheta_{\text{DPRO}} - \vartheta\| \ge \epsilon\right) \le \alpha \tag{49}$$

for all $N \ge N_0, I \ge I_0$.

Proof. For fixed $v \in \mathfrak{F}(\mathfrak{P}_A(\gamma_N))$, it follows by (13) that

$$\begin{aligned}
|u(x,v) - \mathbb{E}[U(x)]| &\le |u(x,v) - u(x,\bar{V})| + |u(x,\bar{V}) - \mathbb{E}[U(x)]| \\
&= |f(x)'(v-\bar{V})| + |u(x,\bar{V}) - \mathbb{E}[U(x)]|.
\end{aligned} \tag{50}$$

Let $\tilde{S}_{<-I>} := 2\Sigma_{<-I>}$, E denote the event that $\tilde{S}_{<-I>} \succ S_{<-I>}$ and F denote the event that $|u(x,\bar{V}) - \mathbb{E}[U(x)]| \le \epsilon/2$. Since $\hat{S}_{<-I>} \succ \Sigma_{<-I>}$, by law of large numbers and Proposition 7, there exists a positive integer N_0 such that for all $N \geq N_0$, $\text{Prob}(E \cap F) \geq 1-\alpha$. By (50)

$$
\begin{aligned}
& \text{Prob}\left(|\vartheta_{\text{DPRO}} - \vartheta| \geq \epsilon\right) \\
\leq\ & \text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N))} |u(x,v) - \mathbb{E}[U(x)| \geq \epsilon\right) \\
=\ & \text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N))} |u(x,v) - \mathbb{E}[U(x)]| \geq \epsilon \bigcap (E\cap F)\right) \\
& +\text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N))} |u(x,v) - \mathbb{E}[U(x)]| \geq \epsilon \bigcap \overline{E\cap F}\right) \\
\leq\ & \text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N))} |u(x,v) - u(x,\bar{V})| + |u(x,\bar{V}) - \mathbb{E}[U(x)]| \geq \epsilon \,\middle|\, E\cap F\right) \text{Prob}(E\cap F) \\
& +\text{Prob}(\overline{E\cap F}) \\
\leq\ & \text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N))} |u(x,v) - u(x,\bar{V})| \geq \frac{\epsilon}{2}\right) \text{Prob}(E\cap F) + \text{Prob}(\overline{E\cap F}) \\
\leq\ & \text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N))} |f(x)'(v-\bar{V})| \geq \frac{\epsilon}{2}\right) + \alpha. \qquad (51)
\end{aligned}
$$

Let $f_{<-I>}(x)$, $v_{<-I>}$ and $\bar{V}_{<-I>}$ denote respectively the vectors which consist of the first $I-1$ components of $f(x), v$ and $\bar{V}$, let $f_I(x), v_I, \bar{V}_I$ denote the I-th component of $f(x)$, v and $\bar{V}$, and e_{I-1} denotes an $(I-1)$-dimensional vector with all ones. Then

$$
v_I - \bar{V}_I = 1 - e'_{I-1} v_{<-I>} - (1 - e'_{I-1}\bar{V}_{<-I>}) = e'_{I-1}(\bar{V}_{<-I>} - v_{<-I>})
$$

and subsequently

$$
f(x)'(v-\bar{V}) = f_{<-I>}(x)'(v_{<-I>} - \bar{V}_{<-I>}) + f_I(x) e'_{I-1}(\bar{V}_{<-I>} - v_{<-I>}).
$$

By the Hölder inequality

$$
\begin{aligned}
|f(x)'(v-\bar{V})| \ \leq\ & \|\tilde{S}^{1/2}_{<-I>} f_{<-I>}(x)\| \|\tilde{S}^{-1/2}_{<-I>}(v_{<-I>} - \bar{V}_{<-I>})\| \\
& +|f_I(x)| \|\tilde{S}^{1/2}_{<-I>} e_{I-1}\| \|\tilde{S}^{-1/2}_{<-I>}(v_{<-I>} - \bar{V}_{<-I>})\| \\
\leq\ & (I-1)\|\tilde{S}^{1/2}_{<-I>}\| \|f(x)\| \|\tilde{S}^{-1/2}_{<-I>}(v_{<-I>} - \bar{V}_{<-I>})\|. \qquad (52)
\end{aligned}
$$

Let N_0 be sufficiently large such that

$$
\frac{\epsilon}{2(I-1)\|\tilde{S}^{1/2}_{<-I>}\| \sup_{x\in\mathfrak{X}} \|f(x)\|} > \sqrt{\gamma_N} \qquad (53)
$$

Combining (51)-(53), we have

$$
\begin{aligned}
& \text{Prob}\left(|\vartheta_{\text{DPRO}} - \vartheta| \geq \epsilon\right) \\
\leq\ & \text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N))} |f(x)'(v-\bar{V})| \geq \frac{\epsilon}{2}\right) + \alpha \\
\leq\ & \text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N))} (I-1)\|\tilde{S}^{1/2}_{<-I>}\|\|f(x)\|\|\tilde{S}^{-1/2}_{<-I>}(v_{<-I>} - \bar{V}_{<-I>})\| \geq \frac{\epsilon}{2}\right) + \alpha \\
\leq\ & \text{Prob}\left(\sup_{v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N)} \|\tilde{S}^{-1/2}_{<-I>}(v_{<-I>} - \bar{V}_{<-I>})\| \geq \frac{\epsilon}{2\sup_{x\in\mathfrak{X})}(I-1)\|\tilde{S}^{1/2}_{<-I>}\|\|f(x)\|}\right) + \alpha \\
\leq\ & \text{Prob}\left(\sup_{v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N))} \|\tilde{S}^{-1/2}_{<-I>}(v_{<-I>} - \bar{V}_{<-I>})\| > \sqrt{\gamma_N}\right) + \alpha \\
\leq\ & \text{Prob}\left(\sup_{v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N))} (Cv - \bar{V}_{<-I>})^T S^{-1}_{<-I>}(Cv - \bar{V}_{<-I>}) > \gamma_N\right) + \alpha \\
=\ & 0 + \alpha = \alpha
\end{aligned}
$$

and hence (49), where the last inequality is due to the fact that $\tilde{S}_{<-I>} \succ S_{<-I>}$ and the last equality is because of (48). □

We now move on to investigate the convergence of ϑ_{DPRO} to ϑ when the ambiguity set is constructed by the bootstrap method. Recall that T^*_k defined in (9), for $k = 1, 2, \ldots$, are bootstrap samples (note that here the number of the bootstrap resamples $K = \infty$). Denote by $\mathfrak{W}_{1-\alpha}$ a region obtained forming the convex hull of the $100(1-\alpha)\%$ interior bootstrap points among all possible T^*_k based on Turkey's depth. Note that $\mathfrak{W}^*_{1-\alpha}$ in (10), constructed using a finite number of T^*_k, is a Monte Carlo approximation of $\mathfrak{W}_{1-\alpha}$. Consider DPRO with

$$
\mathfrak{P}_{\hat{B}}(\alpha) = \left\{P \in \mathscr{P}(\mathbb{R}^I) \mid P(V \in \mathfrak{V}) = 1,\ C\mathbb{E}_P[V] = \bar{V}_{<-I>} - S^{1/2}_{<-I>}\tilde{w}/\sqrt{N}, \tilde{w} \in \mathfrak{W}_{1-\alpha}\right\}. \quad (54)
$$

The difference between $\mathfrak{P}_{\hat{B}}(\alpha)$ and $\mathfrak{P}_B(\alpha)$ in (12) is that $\mathfrak{P}_B(\alpha)$ is denoted using $\mathfrak{W}^*_{1-\alpha}$ while $\mathfrak{P}_{\hat{B}}(\alpha)$ is based on $\mathfrak{W}_{1-\alpha}$. The next lemma refers to Theorem 1 in Yeh and Singh (1997) and the comments following the theorem.

Lemma 1 *If P, the true probability measure of V, is absolutely continuous in $\mathbb{R}^I$, then there exists $\delta_{1-\alpha} > 0$ depending on α such that*

$$
\lim_{N\to\infty} \mathfrak{W}_{1-\alpha} \subseteq \mathfrak{B}(\delta_{1-\alpha}),\ \ a.s.,
$$

where $\mathfrak{B}(\delta_{1-\alpha})$ is an I-dimensional sphere centered at 0 with radius $\delta_{1-\alpha}$.

We next address the convergence property of DPRO with $\mathfrak{P}_{\hat{B}}(\alpha)$ in the following theorem.

Theorem 2 (DPRO with Bootstrap ambiguity (54)**)** *Let $\tau \in (0,1)$ and suppose that P, the true probability measure induced by V, is absolutely continuous in $\mathbb{R}^I$. For any small positive number ϵ, there exist positive numbers I_0, N_0 such that*

$$
\text{Prob}\left(|\vartheta_{\text{DPRO}} - \vartheta| \geq \epsilon\right) \leq \tau \quad (55)
$$

for all $N \geq N_0, I \geq I_0$.

Proof. Choose $\eta \in (0, \tau)$ and $\beta = 1 - (\tau - \eta)$. It follows by Lemma 1 that there exist $\delta_{1-\alpha} > 0$ and $N_0 > 0$ such that

$$\text{Prob}(\mathfrak{W}_{1-\alpha} \subseteq \mathfrak{B}(\delta_{1-\alpha})) \geq \beta,$$

for all $N > N_0$. Let $\gamma_N^{(3)} := \delta_{1-\alpha}/\sqrt{N}$. Note that

$$\mathfrak{F}(\mathfrak{P}_{\hat{B}}(\alpha)) = \{v \in \mathfrak{V} \mid S_{<-I>}^{-1/2}(Cv - \bar{V}_{<-I>}) = \tilde{w}/\sqrt{N}, \tilde{w} \in \mathfrak{W}_{1-\alpha}\}$$

and

$$\mathfrak{F}(\mathfrak{P}_A(\gamma_N^{(3)})) = \{v \in \mathfrak{V} \mid S_{<-I>}^{-1/2}(Cv - \bar{V}_{<-I>}) = \tilde{w}/\sqrt{N}, \tilde{w} \in \mathfrak{B}(\delta_{1-\alpha})\}.$$

Let G denote the event that $\mathfrak{F}(\mathfrak{P}_{\hat{B}}(\alpha)) \subseteq \mathfrak{F}(\mathfrak{P}_A(\gamma_N^{(3)}))$. We then obtain $\text{Prob}(G) \geq \beta$ for $N > N_0$. Consequently, for $N > N_0$, we have

$$\begin{aligned}
\text{Prob}\left(|\vartheta_{\text{DPRO}} - \vartheta| \geq \epsilon\right) &\leq \text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_{\hat{B}}(\alpha))} |u(x,v) - \mathbb{E}[U(x)]| \geq \epsilon\right) \\
&\leq \text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_{\hat{B}}(\alpha))} |u(x,v) - \mathbb{E}[U(x)]| \geq \epsilon \,\middle|\, G\right)\text{Prob}(G) \\
&\quad + \text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_{\hat{B}}(\alpha))} |u(x,v) - \mathbb{E}[U(x)]| \geq \epsilon \,\middle|\, \overline{G}\right)\text{Prob}(\overline{G}) \\
&\leq \text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N^{(3)}))} |u(x,v) - \mathbb{E}[U(x)]| \geq \epsilon\right) + \text{Prob}(\overline{G}).
\end{aligned}$$

It follows from a similar analysis to the proof of Theorem 1 that, if N_0 and I_0 are sufficiently large, then

$$\text{Prob}\left(\sup_{x\in\mathfrak{X}, v\in\mathfrak{F}(\mathfrak{P}_A(\gamma_N^{(3)}))} |u(x,v) - \mathbb{E}[U(x)]| \geq \epsilon\right) \leq \eta,$$

for $N > N_0$ and $I > I_0$. Hence,

$$\text{Prob}\left(|\vartheta_{\text{DPRO}} - \vartheta| \geq \epsilon\right) \leq \eta + (1 - \beta) = \tau.$$

□

7 Conclusions

The past decade has witnessed a surge of research interest in PRO which handles decision making problems under ambiguity of DM's utility preferences. The traditional utility theory, which existing PRO models rely on, assumes preference representations to be deterministic and consistent. Here we propose a novel PRO model which, combining the stochastic utility theory and distributionally robust optimization techniques, is capable of dealing with decision making problems with inconsistent and mutable utility preferences.

We propose two data-driven statistical approaches to construct an ambiguity set of the probability distributions of additive random multi-attribute piecewise linear utility functions. One is to construct an ellipsoidal confidence region with sample mean and sample covariance matrix which is widely used in the literature of distributional robust optimization; the other is to specify a percentile-t bootstrap confidence region. We develop the reformulations and solution algorithms for the two cases in which random utility functions are either generally nondecreasing or concave.

We conduct a numerical test to analyze the effect of the crucial parameters in the bootstrap approach including the size of empirical data, the number of the bootstrap resamples, the critical value of the confidence region, and the variance of the underlying true distribution. The results exhibit the merits of our approach compared to the classical SAA method. Another case study for vehicle design demonstrates the effectiveness of the proposed PRO model working with machine learning. This case is to generate a random sample of consumer preference using a logistic regression method, specify a bootstrap based ambiguity set of the probability distribution of consumer preference, and determine an investment policy among many vehicle development projects for maximizing the expectation of consumer preference over the ambiguity set.

Finally, to extend the scope of applicability of the proposed model and computational schemes, we consider the PRO model with general random utility functions and discuss approximation of general random utility functions by piecewise linear random utility functions. Specifically, we quantify propagation of the error of the approximation to the optimal value and optimal solutions. In the case when the data of the piecewise linear random utility functions are obtained from samples, we demonstrate the convergence of the optimal value obtained from solving the sample-based piecewise linear approximated PRO model to its true counterpart as sample size increases.

References

Allais, M. 1953. Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica* **21**(4) 503–546. doi:10.2307/1907921.

Anscombe, F. J., R. J. Aumann. 1963. A Definition of Subjective Probability. *The Annals of Mathematical Statistics* **34**(1) 199–205. doi:10.1214/aoms/1177704255.

Armbruster, Benjamin, Erick Delage. 2015. Decision Making under Uncertainty when Preference Information is Incomplete. *Management Science* **61**(1) 111–128.

Battista, Tonio Di, Stefano A. Gattone. 2004. Multivariate bootstrap confidence regions for abundance vector using. *Environmental and Ecological Statistics* **11**(4) 355–365. doi:10.1007/s10651-004-4183-z.

Benson, H. Y., Ümit Sağlam. 2014. Mixed-Integer Second-Order Cone Programming: A Survey | Theory Driven by Influential Applications | Tutorials in OR. *INFORMS Tutorials in Operations Research* doi:10.1287/educ.2013.0115.

Bertsimas, D., I. Popescu. 2005. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* **15**(3) 780–804.

Brown, David B., Enrico De Giorgi, Melvyn Sim. 2012. Aspirational Preferences and Their Representation by Risk Measures. *Management Science* **58**(11) 2095–2113. doi:10.1287/mnsc.1120.1537.

Brown, David B., Melvyn Sim. 2008. Satisficing Measures for Analysis of Risky Positions. *Management Science* **55**(1) 71–84. doi:10.1287/mnsc.1080.0929.

Camerer, Colin F. 1989. An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty* **2**(1) 61–104. doi:10.1007/BF00055711.

Cascetta, Ennio. 2009. *Transportation Systems Analysis Models and Applications*. 2nd ed. Springer, New York.

Chajewska, U., D. Koller. 2000. Utilities as Random Variables: Density Estimation and Structure Discovery. Stanford, California, 63–71.

Corlu, Canan G., Alp Akcay, Wei Xie. 2020. Stochastic simulation under input uncertainty: A Review. *Operations Research Perspectives* **7** 100162. doi:10.1016/j.orp.2020.100162.

Daniels, H. E. 1950. Rank Correlation and Population Models. *Journal of the Royal Statistical Society. Series B (Methodological)* **12**(2) 171–191.

Delage, Erick, Yinyu Ye. 2010. Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems. *Operations Research* **58**(3) 595–612. doi:10.1287/opre.1090.0741.

Farquhar, Peter H. 1984. Utility Assessment Methods. *Management Science* **30**(11) 1283–1300.

Feng, Wen. 2016. Assessment of Social Preference in Automotive Market using Generalized Multinomial Logistic Regression. Master Thesis. Dept of IMSE, University of Michigan - Dearborn.

Fishburn, P. C. 1998. Stochastic Utility. S. Barbera, P. J. Hammond, C. Seidl, eds., *Handbook of Utility Theory*, vol. 1. Kluwer Academic Publishers, Boston, MA, 273–320.

Gilboa, I., D. Schmeidler. 1989. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economic* **18**(2) 141–153.

Guo, Shaoyan, Huifu Xu. 2020. Statistical robustness in utility preference robust optimization models. *Mathematical Programming* doi:10.1007/s10107-020-01555-5.

Guo, Shaoyan, Huifu Xu. 2021. Utility Preference Robust Optimization with Moment-Type Information Structure. URL `http://www.optimization-online.org/DB_HTML/2021/01/8233.html`.

Hall, Peter. 1987. On the Bootstrap and Likelihood-Based Confidence Regions. *Biometrika* **74**(3) 481–493. doi:10.2307/2336687.

Hammond, P. J. 1998a. Objective Expected Utility. S. Barbera, P. J. Hammond, C. Seidl, eds., *Handbook of Utility Theory*, vol. 1. Kluwer Academic Publishers, Boston, MA, 143–156.

Hammond, P. J. 1998b. Subjective Expected Utility. S. Barbera, P. J. Hammond, C. Seidl, eds., *Handbook of Utility Theory*, vol. 1. Kluwer Academic Publishers, Boston, MA, 213–271.

Harsanyi, J. C. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* **63**(4) 309–321.

Hey, John, Chris Orme. 1994. Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica* **62**(6) 1291–1326.

Hu, J., M. Bansal, S. Mehrotra. 2018. Robust Decision Making using a General Utility Set. *European Journal of Operational Research* **269**(2) 699–714.

Hu, J., S. Mehrotra. 2015. Robust Decision Making Using a Risk Averse Utility Set with an Application to Portfolio Optimization. *IIE Transactions* **47**(4) 358–372.

Hu, Jian, Gevorg Stepanyan. 2017. Optimization with Reference-Based Robust Preference Constraints. *SIAM Journal on Optimization* **27**(4) 2230–2257. doi:10.1137/16M1105050.

Koppen, M. 2001. Characterization Theorems in Random Utility Theory. Neil J. Smelser, Paul B. Baltes, eds., *International Encyclopedia of the Social & Behavioral Sciences*. Pergamon, Oxford, 1646–1651. doi:10.1016/B0-08-043076-7/00600-8. URL `http://www.sciencedirect.com/science/article/pii/B0080430767006008`.

Liu, Yongchao, Huifu Xu. 2013. Stability analysis of stochastic programs with second order dominance constraints. *Mathematical Programming* **142**(1) 435–460. doi:10.1007/s10107-012-0585-0.

Popescu, Ioana. 2007. Robust Mean-Covariance Solutions for Stochastic Optimization. *Operations Research* **55**(1) 98–112. doi:10.1287/opre.1060.0353.

Scarf, H. 1958. A min-max solution of an inventory problem. Studies in the Mathematical Theory of Inventory and Production, 201–209.

Simon, H. A. 1956. Rational choice and the structure of the environment. *Psychological Review* **63**(2) 129–138. doi:10.1037/h0042769.

So, Anthony Man-Cho. 2011. Moment inequalities for sums of random matrices and their applications in optimization. *Mathematical Programming* **130**(1) 125–151. doi:10.1007/s10107-009-0330-5.

Starmer, Chris. 2000. Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk. *Journal of Economic Literature* **38**(2) 332–382.

Sugden, R. 1998. Alternatives to Expected Utility: Foundations. S. Barbera, P. J. Hammond, C. Seidl, eds., *Handbook of Utility Theory*, vol. 2. Kluwer Academic Publishers, Boston, MA, 273–320.

Thurstone, L. L. 1927. A law of comparative judgment. *Psychological Review* **34**(4) 273–286. doi:10.1037/h0070288.

Tversky, A., D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* **211**(4481) 453–458. doi:10.1126/science.7455683.

Tversky, A., D. Kahneman. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* **5** 297–323.

Tversky, Amos. 1969. Intransitivity of preferences. *Psychological Review* **76**(1) 31–48. doi:10.1037/h0026750.

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

Wakker, Peter, Daniel Deneffe. 1996. Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown. *Management Science* **42**(8) 1131–1150.

Weber, M., K. Borcherding. 1993. Behavioral influences on weight judgments in multiattribute decision making. *European Journal of Operational Research* **67**(1) 1–12.

Wu, George. 1994. An empirical test of ordinal independence. *Journal of Risk and Uncertainty* **9**(1) 39–60. doi:10.1007/BF01073402.

Yeh, Arthur B., Kesar Singh. 1997. Balanced Confidence Regions Based on Tukey's Depth and the Bootstrap. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**(3) 639–652. doi:https://doi.org/10.1111/1467-9868.00088.

Yue, Jinfeng, Bintong Chen, Min-Chiang Wang. 2006. Expected Value of Distribution Information for the Newsvendor Problem. *Operations Research* **54**(6) 1128–1136. doi:10.1287/opre.1060.0318.