Robust Decision Making using a General Utility Set

Jian Hu

Department of Industrial and Manufacturing Systems Engineering

University of Michigan - Dearborn

Dearborn, MI 48128

jianhu@umich.edu

Manish Bansal, Sanjay Mehrotra

Department of Industrial Engineering and Management Sciences

Northwestern University

Evanston, IL 60208

manish.bansal@northwestern.edu, mehrotra@iems.northwestern.edu

September 23, 2015

Abstract

Elicitation of an exact utility function of a decision maker is challenging. In this paper, we address the problem of ambiguity and inconsistency in utility assessments by studying a robust utility-based decision making model where the utility function belongs to a set of general increasing utility functions. We build a robust framework in which the utility function belongs to a set. This set on the utility function is described by boundary and auxiliary conditions. We consider a maximin problem that maximizes the worst-case expected utility of random outcome over the set, thereby hedging the risk arising from uncertainty of the utility function. We study the implications of the uncertain utility on the objective function value of this robust decision model and show that under suitable conditions the Sample Average Approximation (SAA) of a Lagrangian function associated with this model can be solved using a mixed integer linear program. We show that the optimum objective value of the SAA converges to its true counterpart at an exponential rate and we harness this convergence property to present a heuristic which provides a feasible solution for the SAA problem. We illustrate model properties using a portfolio investment problem where investment gains and losses are valued using different uncertain decision (dis)utilities. We also provide computational insights by solving the SAA of this problem as a mixed integer linear program and using our heuristic.

Key Words: Expected Utility Maximization, Utility Function, Distributionally Robust Optimization, Random Target, Portfolio Optimization

1 Introduction

In economics, utility is a measure of satisfaction gained from a good or service and according to the expected utility theory (von Neumann and Morgenstern (1947)), the satisfaction of a decision maker (DM) from a random outcome ξ_1 over a random outcome ξ_2 is given by

$$\mathbb{E}[u(\xi_1)] \geq \mathbb{E}[u(\xi_2)], \tag{1.1}$$

where u is a von Neumann-Morgenstern utility function which characterizes the DM's satisfaction level or risk attitude. Utility functions in decision making can be interpreted as cumulative distribution functions (cdf) (Borch (1968); Berhold (1973)) where a bounded utility function u is normalized to take values in $[0, 1]$ (Castagnoli and LiCalzi (1996); Bordley and LiCalzi (2000)). However, uncertainty of utility function is a major concern in decision science and the lack of accurate description of human behavior is a basic assumption for random utility theory (Thurstone (1927)). Karmarkar (1978) and Weber (1987) ascribed the inaccuracy to cognitive difficulty and incomplete information. Parametric and non-parametric methods have been proposed in the literature for assessing a DM's utility. These include discrete choice models (Train (2009)), parametric estimations using constant absolute risk aversion (CARA) and relative risk aversion (CRRA) utility functions (Pratt (1964) and Arrow (1965)), standard and paired gamble (non-parametric) approaches using preference comparison, probability equivalence, value equivalence, and certainty equivalence (Farquhar (1984), Wakker and Deneffe (1996)). However, despite extensive research in utility assessment, resolving ambiguity that remains after the use of these utility assessment methods is still difficult to solve (Hershey and Schoemaker (1985); Fromberg and Kane (1989); Nord (1992); Chajewska et al. (2000)). Moreover, predetermined forms of utility function in parametric estimations need to be presumed. In view of these difficulties Jacquet-Lagrèze and Siskos (1982) developed a ranking based method, referred to as the UTA method, which considers a disaggregation paradigm. But, unless an exhaustive approach is used, the UTA method is also unable to specify a unique utility function.

In this paper, we study a robust utility-based decision making model which is complementary to the aforementioned utility assessment methods and also handles the ambiguity and inconsistency in utility elicitation by assuming that the utility function u belongs to a set $\mathfrak{U}$ of general increasing utility functions. More specifically, we resolve the ambiguity issue using a robust approach by considering the following decision model (proposed by Armbruster and Delage (2015) and Hu and Mehrotra (2015)):

$$\max_{x \in \mathcal{X}} \min_{u \in \mathfrak{U}} \mathbb{E}[u(\xi(x))] \tag{RVM}$$

where the random outcome ξ is a function of (continuous and/or discrete) decision variables $x \in \mathcal{X} \subseteq \mathbb{R}^n$, and it is assumed that $\xi(x)$ has a finite support. Model (RVM) is a maximin problem that maximizes the worst-case expected utility of random outcome over the utility set $\mathfrak{U}$, thereby hedging the risk arising from the uncertainty of utility function. This setting is also applicable to deterministic decision making where ξ is a deterministic function. It is important to note that in this paper we study model (RVM) where the set $\mathfrak{U}$ consists of general increasing utility functions. In contrast, Armbruster and Delage (2015) and Hu and Mehrotra (2015) were restricted. The set $\mathfrak{U}$ in Armbruster and Delage (2015) consists of increasing concave, S-shaped, or prudent utilities. For individualizing $\mathfrak{U}$ to meet a DM's risk attitude, they use a paired gamble method which designs a number of lottery pairs and requests the DM to compare each pair. As a result, the set $\mathfrak{U}$ consists of the all possible cases describing preference orders given by the DM. Hu and Mehrotra (2015) focused on the problem with increasing concave utilities. The increasing concavity

1

assumption gives a linear programming formulation of the approximation problems. However, it is limited because it is not able to model risk/loss preferences differently. Hu and Mehrotra (2015) specified boundary conditions on utility and the marginal utility, and construct auxiliary conditions using both standard and paired game methods. Likewise, we also specify boundary and auxiliary conditions to combine utility-based robust optimization and parametric and nonparametric utility assessments (See Section 3 for an illustration of these conditions using an example).

This paper makes the following contributions: (i) we study the implications of the uncertain utility (or target) on the objective function value of the general robust utility-based decision making model (RVM) and we define the concept of "cost of target uncertainty"; (ii) we show that under suitable conditions the sample average approximation (SAA) of a Lagrangian function associated with (RVM) can be solved using a mixed integer linear program; (iii) we show that the objective function of this SAA has asymptotic exponential convergence rate; (iv) we harness this convergence property to present a heuristic which provides a feasible solution for the SAA problem; (v) we use a portfolio investment example to discuss the concept of utility (or target) robustness and cost of target uncertainty.

The normalized utility function $u \in \mathfrak{U}$ can be interpreted as the cdf of a random variable ζ, i.e., for a given $t \in \mathbb{R}$, $u(t) = \Pr\{\zeta \leq t\}$ (Castagnoli and LiCalzi (1996)), and $\mathfrak{U}$ is regarded as a set of distributions. We now rewrite model (RVM) as

$$\max_{x \in \mathcal{X}} \min_{u \in \mathfrak{U}} \int \Pr\{\xi(x) \geq t\}\, du(t) \qquad \text{(P-RVM)}$$

because

$$\mathbb{E}[u(\omega)] = \mathbb{E}[\Pr\{\omega \geq \zeta | \omega\}] = \mathbb{E}[\mathbb{E}[\mathbf{1}\{\omega \geq \zeta\}|\omega]] = \mathbb{E}[\mathbf{1}\{\omega \geq \zeta\}] = \Pr\{\omega \geq \zeta\}, \qquad (1.2)$$

where $\omega = \xi(x)$.

1.1 Literature Review on Distributionally Robust Optimization

In general, model (P-RVM) with $\mathfrak{U}$ as the set of cdf's can be interpreted as a distributionally robust optimization problem, i.e. a decision problem where the distribution of random parameter ζ is partially specified. The idea of formulating robust optimization problems with unknown parameter distributions originated in Scarf (1958), where in the context of news vendor model the distribution was specified using the first two moments. Dupacová (1987), Prékopa (1995), Bertsimas and Popescu (2005), Bertsimas et al. (2010), and Delage and Ye (2010) use linear or conic constraints to describe the set of distributions with moments. Shapiro and Ahmed (2004) define the distribution uncertainty set with measure bounds and general moment constraints. Calafiore (2007) defines the distribution uncertainty set using the Kullback-Leibler distance from a reference probability measure. By comparison, Pflug and Wozabal (2007), Pflug et al. (2012), and Wozabal (2012) use the Kantorovich or Wasserstein distance to specify the distribution uncertainty set. Bertsimas et al. (2010) use a piecewise linear utility with first and second moment equality constraints and showed that the corresponding problem has semidefinite programming reformulations. Delage and Ye (2010) give general conditions for polynomial time solvability of a distributionally robust model with constraints on first and second moments. Analogous to the price of robustness introduced by Bertsimas and Sim (2004), Li and Kwon (2013) further consider a penalty incurred by the norm-based distance of the first and second moments from references values. In this paper the specification of the distribution set $\mathfrak{U}$ is similar to the definition given by Shapiro and Ahmed (2004). However, the objective function in Model (P-RVM) is a probability measure. The nonconcavity of this objective function (maximization problem) violates the convexity assumption for the minimization

problem considered in Shapiro and Ahmed (2004). A more general model allowing bounds on higher order moments was recently studied in Mehrotra and Papp (2014).

Model (RVM) in our approach also inherits the feature of the first order stochastic dominance in that a class of increasing utility functions is used to address ambiguity and inconsistency in utility assessments (see e.g., Müller and Stoyan (2002) and references therein). Optimization problems with stochastic dominance requirements have been studied in Dentcheva and Ruszczyński (2004, 2008, 2009) for the univariate case, and in Homem-de-Mello and Mehrotra (2009) and Hu et al. (2012, 2013) for the multivariate case. When viewed from the perspective of stochastic dominance, our approach gives a further characterization of the DM's risk preference beyond a general classification of risk attitude by stochastic dominance. This further characterization is done by imposing additional requirements on a DM's utility.

1.2 Organization of this paper

The specification and properties of the utility-based robust decision model (RVM) (or robust target-based model (P-RVM)) are addressed in Section 2. The properties of $\mathfrak{U}$ under boundary and auxiliary conditions are studied in Section 2.1. In Section 2.2, we show that the cost of uncertainty in utility assessment is an increasing concave function in the size of the set $\mathfrak{U}$. In Section 3, we discuss about the constraints defining the uncertainty set in the robust utility-based model (RVM) with the help of an investment decision problem. In Section 4 we study the Lagrangian dual of model (RVM), and reformulate the SAA of the Lagrangian dual problem to be a mixed integer linear program. Section 5 gives results on the asymptotic convergence and the rate of convergence properties of SAA under suitable assumptions. More specifically, it is shown that the optimal value of the SAA problem converges to the optimum value of its true counterpart at an exponential rate. We harness this convergence property to present a heuristic which provides a feasible solution for the problem with large sample size by utilizing the optimal solution of the small sample size problem. We illustrate the properties of the decision model (RVM) with the help of a numerical example in Section 6. We computationally evaluate the effectiveness of our heuristic and the mixed integer linear program in solving the SAA of the portfolio optimization problem in Section 6.1, and make concluding remarks in Section 6.2.

2 Utility Set: Assumptions and Properties

In this section, we discuss the specification of the utility set $\mathfrak{U}$ in model (RVM). Let $\mathfrak{U}$ be a subset of all increasing utility functions u, defined on Θ, which are right-continuous with left limits (RCLL) and satisfy the boundary condition:

$$u(\theta_1) = 0, \quad u(\theta_2) = 1. \tag{2.1}$$

Conditions (2.1) are commonly used for the normalization of utility functions that does not change the preference ranking of any two alternatives (Keeney and Raiffa, 1976, Theorem 4.1). We assume throughout this paper that, for every $x \in \mathcal{X}$, $\xi(x)$ has a finite support in $\Theta = [\theta_1, \theta_2] \subseteq \mathbb{R}$. Let

$$\Pr\{\xi(x) = \xi_k(x)\} = p_k, \quad k = 1, \ldots, K, \tag{2.2}$$

where K is the number of scenarios, and

$$\varphi_x^0(t) := \Pr\{\xi(x) \geq t\} = \sum_{k=1}^{K} p_k \mathbf{1}\{\xi_k(x) \geq t\}. \tag{2.3}$$

3

Let G_1, G_2 be two increasing positive functions which are right-continuous with left limits (RCLL) defined on Θ. The function G_2 is said to be preferred to G_1 (written $G_1 \preceq G_2$) if, for any interval $(t_1, t_2] \subseteq \Theta$,

$$G_2(t_2) - G_2(t_1) \geq G_1(t_2) - G_1(t_1). \tag{2.4}$$

If the left-hand-side of inequality (2.4) is strictly greater than the right-hand-side for all t_1, t_2, we write $G_1 \prec G_2$.

Denote by $\mathfrak{B}_\Theta$ the collection of all Borel subsets of Θ, and by $\mathcal{M}$ the set of all finite positive signed measure on $(\Theta, \mathfrak{B}_\Theta)$. In this setting, for a given $x \in \mathcal{X}$, the inner minimization problem of model (P-RVM) is given as

$$\min_{u \in \mathcal{M}} \ \int_\Theta \varphi_x^0(t) du(t) \tag{2.5a}$$

$$\text{s.t.} \ \int_\Theta du(t) = 1, \tag{2.5b}$$

$$G_1 \preceq u \preceq G_2, \tag{2.5c}$$

$$\int_\Theta \varphi^i(t) du(t) \leq c_i, \quad i = 1, \ldots, M. \tag{2.5d}$$

The problem (2.5a)–(2.5d) is specified in a measure space $\mathcal{M}$. Condition (2.5b) ensures that $u \in \mathcal{M}$ is a cdf. Condition (2.5c) constructs a region of the utility (cdf) of ζ bounded by functions G_1 and G_2. Note that the functions G_1 and G_2 in (2.5c) are not necessarily cdf functions. The functions $\varphi^i(\cdot)$, $i = 1, \ldots, M$, are assumed to be Lebesgue-integrable functions defined on Θ, and M is the number of constraints. Note that $\varphi^i(\cdot)$ is not dependent on $\xi(x)$. Also, in a special case, when $\varphi^i(t) = t^k$ is the kth monomial, inequality (2.5d) represents a bound on the kth moment of a distribution function. Inequalities (2.5d), however, are quite general. This will become useful in the example of Section 3. Based on the formulation of problem (2.5), we specify the distribution set $\mathfrak{U}$ in model (P-RVM) as

$$\mathfrak{U} := \{ u \in \mathcal{M} \mid u \text{ satisfies conditions (2.5b), (2.5c), and (2.5d) } \}. \tag{2.6}$$

Note that, by the definition of this preference relationship, (2.5c) is equivalent to

$$G_1 + c_1 \preceq u \preceq G_2 + c_2$$

for any constants $c_1, c_2 \in \mathbb{R}$. Hence, in this paper, we assume that $G_1(\theta_1) = G_2(\theta_1) = 0$. The constraint (2.5c), as a special case, allows us to specify the lower and upper bounds using a reference $u_\#$ on $(\Theta, \mathfrak{B}_\Theta)$ as follows:

$$\rho_1 u_\# \preceq u \preceq \rho_2 u_\#, \tag{2.7}$$

where $\rho_1 \in [0, 1]$ and $\rho_2 \in [1, \infty)$ are two given constants controlling the size of uncertainty region.

2.1 Properties of the Utility Set

We now give some basic continuity properties of the functions satisfying the boundary condition (2.5c). These properties, under suitable assumptions, allow us to specify bounds on the cdf of ζ.

Proposition 2.1 *If G_2 is continuous, then all $u \in \mathcal{M}$ satisfying condition* (2.5c) *are continuous.*

4

Proof: Suppose that there exists $u \in \mathcal{M}$ which satisfies condition (2.5c) and is discontinuous at a point $t \in \Theta$. Since u satisfies RCLL, we have $t > \theta_1$. For any $\epsilon \in (\theta_1, t)$, we can assume $u(t) - u(t - \epsilon) > \delta > 0$. On the other hand, by the continuity of G_2, it follows that there exists $\hat{\epsilon} \in (\theta_1, t)$ such that $G_2(t) - G_2(t - \hat{\epsilon}) < \delta$. It implies that

$$G_2(t) - G_2(t - \hat{\epsilon}) < u(t) - u(t - \hat{\epsilon}),$$

which contradicts $u \preceq G_2$ on the interval $[t - \hat{\epsilon}, t]$. $\qquad\square$

We next discuss the absolute continuity of $u \in \mathcal{M}$. The next proposition shows that to ensure the absolute continuity of all $u \in \mathcal{M}$ satisfying (2.5c), it is sufficient to assume that the upper bounding function G_2 in the boundary conditions (2.5c) is absolute continuous.

Proposition 2.2 *If G_2 is absolute continuous on Θ (with respect to Lebesgue measure), then all $u \in \mathcal{M}$ satisfying condition (2.5c) are absolute continuous on Θ.*

Proof: By the definition of absolute continuity, we know that for any $\epsilon > 0$, there exists $\delta > 0$ such that whenever a finite sequence of pairwise disjoint sub-intervals $(a_k, b_k]$ of Θ satisfies

$$\sum_k |b_k - a_k| < \delta,$$

then

$$\sum_k |u(b_k) - u(a_k)| \le \sum_k |G_2(b_k) - G_2(a_k)| < \epsilon.$$

This implies that u is absolute continuous. $\qquad\square$

The boundary conditions (2.5c) describes a preference (ordering) of the cdf of ζ in comparison with reference functions G_1 and G_2. This preference can also be interpreted using the pdf of ζ as shown in the following proposition.

Proposition 2.3 *If functions $u_1, u_2 \in \mathcal{M}$ are absolute continuous in Θ, then $u_1 \preceq u_2$ if and only if $u_1' \le u_2'$ a.e., where u_1', u_2' are derivatives of u_1 and u_2.*

Proof: The absolute continuity of u_1 and u_2 implies the existence of their derivatives u_1' and u_2', a.e., i.e., for any $(t_1, t_2] \subseteq \Theta$,

$$u_1(t_2) - u_1(t_1) = \int_{t_1}^{t_2} u_1'(s)ds \quad \text{and} \quad u_2(t_2) - u_2(t_1) = \int_{t_1}^{t_2} u_2'(s)ds.$$

If $u_1' \le u_2'$ a.e. on Θ, then it follows that

$$u_1(t_2) - u_1(t_1) \le u_2(t_2) - u_2(t_1),$$

and consequently $u_1 \preceq u_2$, by definition. On the other hand, if $u_1 \preceq u_2$, at any given interval $(t, t + h] \subset \Theta$, we have

$$u_1(t + h) - u_1(t) \le u_2(t + h) - u_2(t),$$

which implies that, if u_1, u_2 are differentiable at t, then

$$u_1'(t) = \lim_{h \to 0} \frac{u_1(t + h) - u_1(t)}{h} \le \lim_{h \to 0} \frac{u_2(t + h) - u_2(t)}{h} = u_2'(t).$$

$\qquad\square$

Assume that G_1 and G_2 are both absolutely continuous, and let g_1 and g_2 be derivatives of G_1 and G_2. Using Proposition 2.3, we rewrite the boundary conditions (2.5c) as

$$g_1 \le u' \le g_2, \text{ a.e.,} \tag{2.8}$$

which shows an equivalent representation in the form of the pdf u' of ζ. In the context of utility, condition (2.8) specifies the pointwise lower and upper bounds of the marginal utility function u'.

2.2 Cost of Increased Ambiguity and Parametric Properties of the Decision Model

The ambiguity in the utility may result in sub-optimal decision making. Intuitively, the larger the uncertainty set $\mathfrak{U}$, the higher would be the cost of uncertainty. The goal of this section is to provide a parametric construction that can be used to study the effect of uncertainty in our ability to specify ζ. Assume that $\mathfrak{U}_1 \subseteq \mathfrak{U}_2$. Then, we may view the difference in the objective function value of the decision model (P-RVM) as the cost of increased (value of reducing) ambiguity from set $\mathfrak{U}_1$ to $\mathfrak{U}_2$. This is more formally stated in Theorem 2.6. Let

$$C(\mathfrak{U}) := \begin{cases} - \max_{x \in \mathcal{X}} \min_{u \in \mathfrak{U}} \int_{\Theta} \varphi_x^0(t) du(t), & \text{if } \mathfrak{U} \text{ is nonempty,} \\ -\infty & \text{otherwise.} \end{cases} \tag{2.9}$$

In Theorem 2.6 we show that the cost function $C(\mathfrak{U})$ is increasing and concave. The results are proved in Appendix C–E. We first show that the set $\mathfrak{U}$ defined in (2.6) is convex.

Proposition 2.4 $\mathfrak{U}$ *given by* (2.6) *is a convex set.*

Let $\kappa\mathfrak{U} = \{\kappa u \mid u \in \mathfrak{U}\}$ and denote the Minkowski sum of two sets as $\mathfrak{U}_1 \oplus \mathfrak{U}_2 = \{u_1 + u_2 \mid u_1 \in \mathfrak{U}_1, u_2 \in \mathfrak{U}_2\}$. Also for given convex sets $\mathfrak{U}_1 \subseteq \mathfrak{U}_2$, let

$$\mathfrak{U}(\kappa) := (1 - \kappa)\mathfrak{U}_1 \oplus \kappa\mathfrak{U}_2$$

for some $\kappa \in [0, 1]$.

Lemma 2.5 *For* $0 \leq \kappa_1 \leq \kappa_2 \leq 1$, $\mathfrak{U}(\kappa_1) \subseteq \mathfrak{U}(\kappa_2)$.

Theorem 2.6 *The function* $C(\mathfrak{U}(\kappa))$ *is an increasing concave function of* κ.

The increasing property of $C(\mathfrak{U})$ is consistent with the common experience that the decision may become increasingly suboptimal with our inability to specify the target precisely. The risk averse nature of the decision framework (P-RVM) ensures that the incremental inefficiency added to the decision making resulting from an increase in the uncertainty is diminishing.

3 Application in Portfolio Optimization with Ambiguous Utility

In this section we develop an example application to illustrate the usefulness of our model. Specifically, we apply the the robust utility based decision model (RVM) for specifying a robust version of a portfolio optimization problem where the decisions are based on a DM's expected utilities. As mentioned before, the boundary conditions (2.5c) describe an uncertainty region of utility functions and the auxiliary conditions (2.5d) allow us to model additional conditions that narrow the choice of a DM's utility. Such conditions can be used to incorporate information generated from the utility assessment methods mentioned in the introduction. We illustrate the use of these boundary and auxiliary conditions in the following subsections.

3.1 Specifying Boundary Conditions Using Marginal and Reference Utility Functions

We first use a portfolio investment problem to illustrate the construction of the boundary conditions (2.5c) for a DM's utility. John would like to invest \$1 into the index fund market. His best hope is a 200% return, while he understands that it is possible for him to lose all investment. John's

attitude follows the Prospect theory, i.e., John prefers a loss aversion on interval $[0, 1]$ for a possible loss of his investment, while he has risk aversion on $[1, 2]$ for a gain. An example of a normalized smooth S-shaped utility function is

$$\hat{u}_{\#}(\alpha, \beta)(t) := \begin{cases} 0 & t < 0, \\ \frac{-\alpha(1-e^{\pi(\alpha,\beta)(t-1)})+\alpha(1-e^{-\pi(\alpha,\beta)})}{(1+\alpha)(1-e^{-\pi(\alpha,\beta)})} & 0 \le t < 1, \\ \frac{1-e^{-\beta(t-1)}+\alpha(1-e^{-\beta})}{(1+\alpha)(1-e^{-\beta})} & 1 \le t \le 2, \\ 1 & t > 2. \end{cases} \tag{3.1}$$

The constant α is the ratio of the slope on the loss side over that on the gains side. The risk seeking part of the S-shaped utility function is completely decided for the given α and β to satisfy the property of smoothness. We denote this risk seeking coefficient by $\pi(\alpha, \beta)$. The continuity of $\hat{u}'_{\#}(\alpha, \beta)(t)$ at $t = 1$ implies

$$\alpha(1 - e^{-\beta})\pi(\alpha, \beta) + \beta e^{-\pi(\alpha,\beta)} = \beta. \tag{3.2}$$

A reasonable range of α suggested by Tversky and Kahneman (1992) is between 2.0 and 2.25, and John chooses $\alpha = 2$. The constant β is the constant-absolute-risk-averse (CARA) coefficient of the exponential utility function on $[1, 2]$. Further analysis shows that $\beta = 3$ best describes John's risk preference. Solving equation (3.2), we obtain $\pi(2, 3) = 0.9949$. Observe that $\hat{u}_{\#}(2, \beta)(1) = 2/3$ for all values of β, i.e., the utility of no investment is $2/3$. The reference marginal utility is given as

$$\hat{u}'_{\#}(\alpha, \beta)(t) = \begin{cases} \frac{\alpha\pi(\alpha,\beta)e^{\pi(\alpha,\beta)(t-1)}}{(1+\alpha)(1-e^{-\pi(\alpha,\beta)})} & 0 \le t < 1, \\ \frac{\beta e^{-\beta(t-1)}}{(1+\alpha)(1-e^{-\beta})} & 1 \le t \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

These reference utility and marginal utility functions are drawn in Figure 1a and Figure 1b respectively.

(a) $\hat{u}_{\#}(\alpha, \beta)(t)$
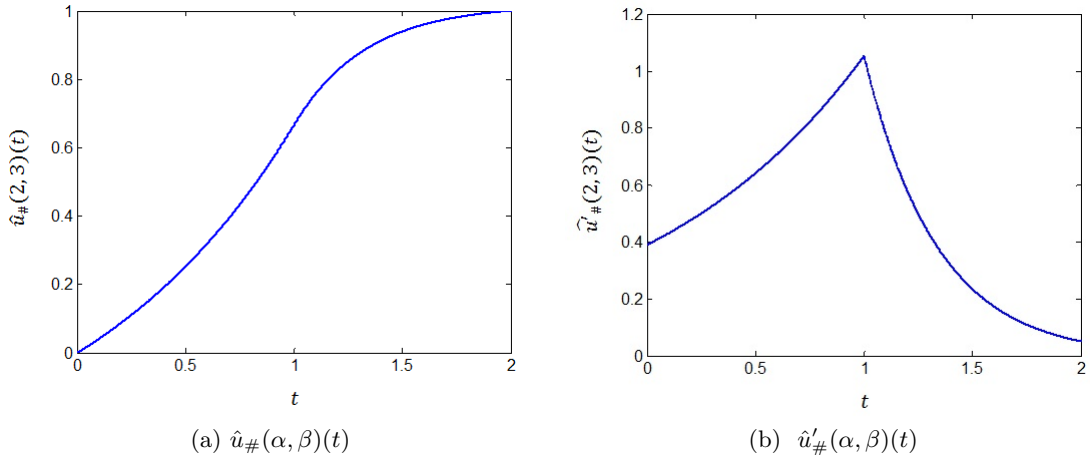
(b) $\hat{u}'_{\#}(\alpha, \beta)(t)$

Figure 1: Reference Utility and Marginal Utility Functions for $\alpha = 2, \beta = 3$

Similar to the use of a reference cdf of a random target to specify the boundary conditions (2.7), with respect to this reference utility function $u_{\#}(2, 3)$, John constructs the boundary conditions as

$$\rho_1 \hat{u}_{\#}(2, 3) \preceq u \preceq \rho_2 \hat{u}_{\#}(2, 3). \tag{3.4}$$

Furthermore, the reference utility $\hat{u}_{\#}(\alpha, \beta)(t)$ is an absolute continuous function. Hence, the boundary condition (3.4) is equivalent to

$$\rho_1 \hat{u}'_{\#}(2,3) \le u' \le \rho_2 \hat{u}'_{\#}(2,3), \tag{3.5}$$

where u', the derivative of u, is the marginal utility function. We will adjust ρ_1 and ρ_2 in our case study in Section 6. We will then discuss the effect of these parameters on the optimal value and solution of (P-RVM).

3.2   Using Utility Assessment Methods to Specify Auxiliary Conditions

A DM's utility can be assessed by various methods such as the standard or paired gamble methods, the probability equivalence method, the value equivalence method, and the certainty equivalence method.

3.2.1   Utility Assessment using Gamble and Certainty Equivalent Methods.

We briefly discuss about the standard or paired gamble method and the certainty equivalence method for the utility assessment and the construction of auxiliary condition (2.5d) (for details refer to Hu and Mehrotra (2015), Farquhar (1984), Wakker and Deneffe (1996) and references therein). The preference comparison in the paired gamble method Farquhar (1984) uses the pairwise comparison of two random outcomes to assess the utility. On the other hand, instead of point estimation, the certainty equivalent method uses a range for each certainty equivalent of a random event. Hu and Mehrotra (2015) provide numerical examples to illustrate the constructions of the auxiliary condition (2.5d) using the preference comparison method and the certainty equivalent method.

3.2.2   Utility Assessment using a Moment based Approach.

We now discuss a moment based approach to construct the auxiliary condition (2.5d). Our primary motivation is to illustrate the use of moments in modeling random targets. However, we point out that Abbas (2007) discusses properties and applications of the moments of the random target, and provides a method to identify a utility function satisfying moment assessments. The DM suggests a range, denoted by $[\underline{\gamma}_k, \overline{\gamma}_k]$, of the kth moment of the random target ζ to allow for uncertainty in parameter estimate, i.e.,

$$\underline{\gamma}_k \le \mathbb{E}[\zeta^k] \le \overline{\gamma}_k.$$

By letting

$$\begin{aligned}
\varphi^i(t) &:= t^k, & c_i &:= \overline{\gamma}_k, \\
\varphi^{i+1}(t) &:= -t^k, & c_{i+1} &:= -\underline{\gamma}_k,
\end{aligned} \tag{3.6}$$

the condition (3.6) now describes bounds on the moments of ζ. In the following proposition we give an approach for the estimation of the moments of ζ.

Proposition 3.1 *Let ϑ_k be a random variable with p.d.f. $k(\theta_2 - \theta_1)^{-k}(t - \theta_1)^{k-1}\mathbf{1}\{t \in \Theta\}$. Then, the kth moment of ζ with c.d.f. $u \in \mathfrak{U}$ is*

$$\mathbb{E}[\zeta^k] = (\theta_2 - \theta_1)^k(1 - \mathbb{E}[u(\vartheta_k)]). \tag{3.7}$$

8

We assume that John provides estimates of expected utilities of an outcome that is uniformly distributed on $[0, 2]$ and an outcome that follows a triangle distribution with the lower limit 0, the upper limit 2, and the mode 2. Following Proposition 3.1, this gives auxiliary constraints

$$2(1 - \overline{\tau}_1) =: \underline{\gamma}_1 \le \int_0^2 t du = 2(1 - \mathbb{E}[u(\vartheta_1)]) \le \overline{\gamma}_1 := 2(1 - \underline{\tau}_1), \tag{3.8}$$

$$4(1 - \overline{\tau}_2) =: \underline{\gamma}_2 \le \int_0^2 t^2 du = 4(1 - \mathbb{E}[u(\vartheta_2)]) \le \overline{\gamma}_2 := 4(1 - \underline{\tau}_2), \tag{3.9}$$

where $[\underline{\tau}_j, \overline{\tau}_j]$ ($j = 1$ and 2) quantifies the lower and upper bounds of the expected utilities of ϑ_j. We take $\mathbb{E}[u(\vartheta_1)] \in [0.5, 0.55]$ and $\mathbb{E}[u(\vartheta_2)] \in [0.75, 0.8]$. Correspondingly, $\underline{\gamma}_1 = 0.9$, $\overline{\gamma}_1 = 1$, $\underline{\gamma}_2 = 0.8$, and $\overline{\gamma}_2 = 1$.

4   Lagrangian Dual Problem and its Sample Average Approximation

We now develop an approach for solving (P-RVM) with the inner minimization problem reformulated as (2.5). Note that the objective function in our problem is non-convex, due to the indicator function φ_x^0 in the objective function (2.3). However, recall from Proposition 2.4 that the set $\mathfrak{U}$ is convex, and for a fixed x the inner problem (2.5) is a convex optimization problem in $u \in \mathfrak{U}$. We take advantage of this convexity to develop our solution approach for (P-RVM). Specifically, we study the Lagrangian dual problem of the inner problem (2.5) and its sample average approximation. The following assumptions are selectively used for the duality and convergence results proved in this paper.

(A1). The set $\mathfrak{U}$ has an interior point, i.e., there is a $u \in \mathcal{U}$ such that $G_1 \prec u \prec G_2$ and $\int_\Theta \varphi^i(t) du(t) < c_i$ for all $i = 1, \ldots, M$.
(A2). The functions G_1, G_2 are absolute continuous.
(A3). The set $\mathcal{X}$ is a nonempty compact set.
(A4). The function $\xi_k(x)$ is continuous in $\mathcal{X}$ for $k = 1, \ldots, K$.
(A5). The functions G_1, G_2 are Lipschitz continuous on Θ, and φ^i, $i = 1, \ldots, M$, are bounded on Θ.

Assumption (A1) is the Slater condition needed to ensure that (see Theorem 4.1 below) there is no duality gap between the primal problem (2.5) and its Lagrangian dual problem. Under Assumption (A2), we provide a SAA of the Lagrangian dual problem. The asymptotic convergence of this SAA is ensured in Theorem 5.2 under Assumptions (A1) - (A4). Furthermore, using Assumptions (A1)–(A5), we will show an exponential convergence rate in Theorem 5.3.

For $x \in \mathcal{X}$ and $\lambda \in \Re^{M+1}$, we denote

$$\mathcal{L}(x, \lambda) := \varphi_x^0 + \lambda_0 + \sum_{i=1}^M \lambda_i \varphi^i, \; and \tag{4.1}$$

$$h(\lambda) := -\lambda_0 - \sum_{i=1}^M c_i \lambda_i. \tag{4.2}$$

Note that, for given x and λ, $\mathcal{L}(x, \lambda)$ is a function mapping Θ to $\mathbb{R}$. In the later statement we denote by $\mathcal{L}(x, \lambda)(t)$ the value of the function $\mathcal{L}(x, \lambda)$ at some t. The Lagrangian of problem (2.5)

is

$$L(x, u, \lambda) := \int_{\Theta} \mathcal{L}(x, \lambda)(t) du(t) + h(\lambda), \tag{4.3}$$

and the Lagrangian dual problem of (2.5) is given as

$$\max_{\lambda} \left\{ \psi(x, \lambda) := \inf_{u \in \mathcal{D}} L(x, u, \lambda) \right\} \tag{4.4}$$
$$\text{s.t. } \lambda_i \geq 0, \quad i = 1, \ldots, M,$$

where the set

$$\mathcal{D} := \{ u \in \mathcal{M} \mid G_1 \preceq u \preceq G_2 \} \tag{4.5}$$

consists of the boundary condition (2.5c). For a given x and λ, on a measurable region $\{t \in \Theta \mid \mathcal{L}(x, \lambda)(t) > 0\}$ the minimum of $L(x, u, \lambda)$ over the set $\mathcal{D}$ is obtained at G_1. Similarly, on a measurable region $\{t \in \Theta \mid \mathcal{L}(x, \lambda)(t) < 0\}$ the minimum of $L(x, u, \lambda)$ is obtained at G_2. Hence, we have

$$\psi(x, \lambda) = \int_{\Theta} [\mathcal{L}(x, \lambda)(t)]_+ dG_1(t) - \int_{\Theta} [\mathcal{L}(x, \lambda)(t)]_- dG_2(t) + h(\lambda), \tag{4.6}$$

where $[a]_+ := \max\{a, 0\}$ and $[a]_- := \max\{-a, 0\}$.

We now present the weak and strong duality relationship of the primal problem (2.5) and the dual problem (4.4) in the following theorem. Note that our problem is in a functional space, hence a formal proof (which we could not find in the literature) of weak and strong duality is required. The proof of Theorem 4.1 follows the steps in the proof of Theorem 6.2.4 in Bazaraa et al. (2006), where the finite dimensional case is considered. We provide the proof in Appendix F for completeness.

Theorem 4.1 *At any given $x \in \mathcal{X}$, we have that*

1. *(the weak duality) the optimal value of the primal problem (2.5) is greater than or equal to that of the dual problem (4.4),*
2. *(the strong duality) under Assumption (A1), there is no duality gap between (2.5) and (4.4).*

Theorem 4.1 shows that, under Assumption (A1), there is no duality gap between (2.5) and (4.4). Hence, we can rewrite problem (P-RVM) as

$$\max_{x, \lambda} \psi(x, \lambda)$$
$$\text{s.t. } \lambda_i \geq 0, \quad i = 1, \ldots, M, \tag{4.7}$$
$$x \in \mathcal{X}.$$

We now study the SAA of problem (4.7). Assumption (A2) ensures the existence of the derivatives of G_1, and G_2. These derivatives are denoted by g_1 and g_2, a.e. Hence, by (4.6) we have

$$\psi(x, \lambda) = \int_{\Theta} [\mathcal{L}(x, \lambda)(t)]_+ g_1(t) dt - \int_{\Theta} [\mathcal{L}(x, \lambda)(t)]_- g_2(t) dt + h(\lambda). \tag{4.8}$$

Let $\zeta_1, \ldots, \zeta_N$ be iid samples following the uniform distribution on Θ. We approximate problem (4.7) by

$$\max_{x, \lambda} \left\{ \psi_N(x, \lambda) := \frac{1}{N} \sum_{j=1}^{N} g_1(\zeta_j)[\mathcal{L}(x, \lambda)(\zeta_j)]_+ - g_2(\zeta_j)[\mathcal{L}(x, \lambda)(\zeta_j)]_- + h(\lambda) \right\}$$
$$\text{s.t. } \lambda_i \geq 0, \quad i = 1, \ldots, M, \tag{4.9}$$
$$x \in \mathcal{X}.$$

The SAA problem (4.9) is non-convex because the objective ψ_N has the indicator function (2.3). We now give an equivalent mixed-integer reformulation of (4.9) by introducing intermediate $2N$ continuous variables s_j^1 and s_j^2; and KN binary variables $r_{k,j}$.

Theorem 4.2 *Suppose $\zeta_1 > \cdots > \zeta_N$. Let $\zeta_{N+1} := \theta_1$ and g_1, g_2 be the derivatives of G_1, G_2. Then problem (4.9) is equivalent to a mixed-integer program*

$$\max_{x,\lambda,s^1,s^2,r} \frac{1}{N} \sum_{j=1}^{N} g_1(\zeta_j)s_j^1 - g_2(\zeta_j)s_j^2 + h(\lambda) \tag{4.10a}$$

$$s.t. \; \xi_k(x) + \sum_{j=1}^{N}(\zeta_j - \zeta_{j+1})r_{k,j} \geq \zeta_1, \quad k = 1, \ldots, K, \tag{4.10b}$$

$$r_{k,j} \geq r_{k,j+1}, \quad k = 1, \ldots, K, \; j = 1, \ldots, N-1, \tag{4.10c}$$

$$s_j^1 - s_j^2 \leq \sum_{k=1}^{K} p_k(1 - r_{k,j}) + \lambda_0 + \sum_{i=1}^{M} \lambda_i \varphi^i(\zeta_j), \tag{4.10d}$$

$$r_{k,j} \in \{0, 1\}, \quad k = 1, \ldots, K, \; j = 1, \ldots, N, \tag{4.10e}$$

$$s_j^1, \; s_j^2 \geq 0, \quad j = 1, \ldots, N, \tag{4.10f}$$

$$\lambda_i \geq 0, \quad i = 1, \ldots, M, \tag{4.10g}$$

$$x \in \mathcal{X}. \tag{4.10h}$$

The following technical lemma needed in the proof of Theorem 4.2 gives an intermediate mixed-integer formulation of problem (4.9).

Lemma 4.3 *Problem (4.9) can be equivalently represented as*

$$\max_{x,\lambda,s^1,s^2,r} \frac{1}{N} \sum_{j=1}^{N} g_1(\zeta_j)s_j^1 - g_2(\zeta_j)s_j^2 + h(\lambda) \tag{4.11a}$$

$$s.t. \; \xi_k(x) + (\zeta_j - \theta_1)r_{k,j} \geq \zeta_j, \quad k = 1, \ldots, K, \; j = 1, \ldots, N, \tag{4.11b}$$

$$(x, \lambda, s^1, s^2, r) \; satisfies \; conditions \; (4.10d) - (4.10h). \tag{4.11c}$$

Proof: We first claim that problem (4.9) is equivalent to

$$\max_{x,\lambda,s^1,s^2} \frac{1}{N} \sum_{j=1}^{N} g_1(\zeta_j)s_j^1 - g_2(\zeta_j)s_j^2 + h(\lambda) \; \text{where} \tag{4.12a}$$

$$s.t. \; s_j^1 - s_j^2 \leq \mathcal{L}(x, \lambda)(\zeta_j), \tag{4.12b}$$

$$(x, \lambda, s^1, s^2) \; satisfies \; conditions \; (4.10f) - (4.10h). \tag{4.12c}$$

Problems (4.9) and (4.12) are equivalent if and only if the optimal solution $(\hat{s}^1, \hat{s}^2, \hat{x}, \hat{\lambda})$ of (4.12) satisfies

$$\hat{s}_j^1 = [\mathcal{L}(\hat{x}, \hat{\lambda})(\zeta_j)]_+, \quad \hat{s}_j^2 = [\mathcal{L}(\hat{x}, \hat{\lambda})(\zeta_j)]_-, \quad j = 1, \ldots, N,$$

where $\hat{s}^i = (\hat{s}_1^i, \ldots, \hat{s}_N^i)$ for $i = 1$, 2. Suppose that there exists $j \in \{1, \ldots, N\}$ such that $\hat{s}_j^1 - \hat{s}_j^2 \leq \mathcal{L}(\hat{x}, \hat{\lambda})(\zeta_j) - \delta_j$ for some $\delta_j > 0$. Let e_j denote the vector whose jth element is 1 and other elements

are 0. The solution $(\hat{s}^1 + \delta_j e_j, \hat{s}^2, \hat{x}, \hat{\lambda})$ is a feasible solution of problem (4.12) at which the objective value is larger than the optimal value. Therefore, $\hat{s}_j^1 - \hat{s}_j^2 = \mathcal{L}(\hat{x}, \hat{\lambda})(\zeta_j)$, which implies

$$\hat{s}_j^1 \geq [\mathcal{L}(\hat{x}, \hat{\lambda})(\zeta_j)]_+, \quad \hat{s}_j^2 \geq [\mathcal{L}(\hat{x}, \hat{\lambda})(\zeta_j)]_-.$$

Now assume $\hat{s}_j^1 = [\mathcal{L}(\hat{x}, \hat{\lambda})(\zeta_j)]_+ + \sigma_j$ for some $\sigma_j > 0$. It then follows that $\hat{s}_j^2 = [\mathcal{L}(\hat{x}, \hat{\lambda})(\zeta_j)]_- + \sigma_j$. $(\hat{s}^1 - \sigma_j e_j, \hat{s}^2 - \sigma_j e_j, \hat{x}, \hat{\lambda})$ is a feasible solution with a larger objective value.

Introducing the intermediate binary variables $r_{k,j}$, we finally replace condition (4.12b) by conditions (4.11b) and (4.10d). $\qquad \square$

Proof: (Theorem 4.2) By condition (4.11b), $r_{k,j}$ is forced to be 1 if $\gamma := \xi_k(x) < \zeta_j$ for given $x \in \mathcal{X}$ and $k \in \{1, \ldots, K\}$. Since $\zeta_1 > \cdots > \zeta_N$ is a decreasing sequence, $r_{k,1} \geq \cdots \geq r_{k,N}$ described in condition (4.10c) can be generated as cuts in problem (4.11) without changing its optimal value.

We complete the proof by using Lemma 4.3 and showing that the following two sets $\mathcal{G}$ and $\mathcal{G}'$, described by conditions (4.10b), (4.10c), and (4.11b), are equivalent. Let

$$\mathcal{G} := \left\{ r \in [0,1]^N \;\middle|\; \gamma + \sum_{j=1}^N (\zeta_j - \zeta_{j+1}) r_j \geq \zeta_1, \; r_1 \geq \cdots \geq r_N \right\}$$

and

$$\mathcal{G}' := \left\{ r \in [0,1]^N \;|\; \gamma + (\zeta_j - \theta_1) r_j \geq \zeta_j, \; j = 1, \ldots, N, \; r_1 \geq \cdots \geq r_N \right\}.$$

Note that, by definition, we have $\theta_1 \leq \gamma \leq \theta_2$. Choose $\zeta_0 > \theta_2$. There exists some $j \in \{1, \ldots, N+1\}$ such that $\zeta_{j-1} > \gamma \geq \zeta_j$. If $j = 1$, $\mathcal{G} = \mathcal{G}' = \left\{ r \in [0,1]^N \;|\; r_1 \geq \cdots \geq r_N \right\}$. We next assume $j \geq 2$. If $r \in \mathcal{G}'$, then $r_i = 1$ for $i \geq j - 1$, and $r_j \geq \cdots \geq r_N$. It is easy to check that $r \in \mathcal{G}$, and thus $\mathcal{G}' \subseteq \mathcal{G}$.

On the other hand, for $r \in \mathcal{G}$, we also have $r_i = 1$ for all $i \leq j - 1$ and $r_j \geq \cdots \geq r_N$. Otherwise, there exists $s \in \{1, \ldots, j-1\}$ such as $r_i = 0$ for $i \geq s$ and $r_i = 1$ for $i < s$. At this r, the first condition $\gamma + \sum_{j=1}^N (\zeta_j - \zeta_{j+1}) r_j \geq \zeta_1$ in $\mathcal{G}$ implies that $\gamma \geq \zeta_s > \zeta_{j-1}$. It is contradictory to the previous assumption that $\zeta_{j-1} > \gamma \geq \zeta_j$. It follows that $r \in \mathcal{G}'$, and hence, $\mathcal{G} \subseteq \mathcal{G}'$. $\qquad \square$

Remark 4.4 *If we construct the perturbation region (2.7) around a preferred (normalized) reference cdf $u_\#$, N iid samples $\zeta^1, \ldots, \zeta^N$ can be taken under $u_\#$. In this case the reformulation (4.10) of problem (4.9) is given by*

$$\max_{x, \lambda, s^1, s^2, r} \frac{1}{N} \sum_{j=1}^N \rho_1 s_j^1 - \rho_2 s_j^2 + h(\lambda) \tag{4.13a}$$

$$\text{s.t. } (x, \lambda, s^1, s^2, r) \text{ satisfies conditions } (4.10b) \text{ - } (4.10h). \tag{4.13b}$$

Remark 4.5 *Luedtke et al. (2010) present the following valid-inequalities for $\mathcal{G}'$:*

$$\gamma + \sum_{l=1}^n \left(\zeta_{t_l} - \zeta_{t_{l+1}} \right) r_{t_l} \geq \zeta_{t_1} \quad \forall T = \{t_1, \ldots, t_n\} \subseteq \{1, \ldots, N\}, \tag{4.14}$$

where $t_1 < \ldots < t_n$ and $\zeta_{t_{n+1}} = \theta_1$, and prove that these inequalities are facet-defining for $\mathrm{conv}(\mathcal{G}')$ if and only if $t_1 = 1$. Notice that for $T = \{1, \ldots, N\}$, inequality (4.14) is same as a defining inequality of $\mathcal{G}$. Similar to Luedtke et al. (2010), we observe that the addition of inequalities (4.14), where $\gamma = \xi_k(x)$ and $r_j = r_{k,j}$, to (4.10) has not effect on the optimal objective value of its LP relaxation.

5 Asymptotic Convergence and Exponential Rate of Convergence for the Approximated Problem

The SAA approach (4.9) is developed to approximate problem (4.7). In this section we discuss the asymptotic convergence of the optimal value and solutions of the SAA problem (4.9). The converge of the optimal solutions is ensured for the deviation of the sets of the optimal solutions of the SAA problem from the sets of the optimal solutions of the true counterpart decreases to 0 as the sample size increases. Here, the deviation of two sets $\mathcal{A}$ and $\mathcal{B}$ is denoted as

$$\mathbb{D}(\mathcal{A}, \mathcal{B}) := \sup_{a \in \mathcal{A}} \inf_{b \in \mathcal{B}} \|a - b\|.$$

We also show that the optimal value of the SAA problem (4.9) converges to the true counterpart of problem (4.7) with an exponential rate. The novel aspect of the analysis in this section is that SAA is taken for a set in the functional (infinite) dimensional space.

We first present Lemma 5.1 which shows the boundedness of the sets of optimal solutions of the true problem (4.7) and its SAA (4.9). Then, Theorem 5.2 ensures the asymptotic convergence of the approximation problem (4.9).

Lemma 5.1 *Let $\mathcal{Z}^*$ be the set of optimal solutions of problem (4.7), and $\mathcal{Z}_N$ be that of problem (4.9). If Assumption (A1) - (A3) holds, then $\mathcal{Z}^*$ is bounded and $\mathcal{Z}_N$ is bounded a.s. for large enough N. In particular, $\mathcal{Z}^* \subseteq X \times \widehat{\Lambda}$, where*

$$\widehat{\Lambda} := \{\lambda \in \mathbb{R}^{M+1} \mid \psi(x, \lambda) \geq 0, \ x \in \mathcal{X}, \lambda_i \geq 0, \ i = 1, \ldots, M\}$$

is a nonempty compact set.

Theorem 5.2 *Let y^* and $\mathcal{Z}^*$ be the optimal value and the set of optimal solutions of problem (4.7), and y_N and $\mathcal{Z}_N$ be those of problem (4.9). If Assumptions (A1) - (A4) hold, then $y_N \to y^*$ and $\mathbb{D}(\mathcal{Z}_N, \mathcal{Z}^*) \to 0$ a.s. as $N \to \infty$.*

Proof: By Lemma 5.1, $\mathcal{Z}^*$ is bounded and $\mathcal{Z}_N$ is bounded a.s.. for large enough N under Assumptions (A1) - (A3). There exists a compact set $\mathcal{C} \subseteq \mathcal{X} \times \Lambda(:= \{\lambda \in \mathbb{R}^{M+1} \mid \lambda_i \geq 0, i = 1, \ldots, M\})$ such that $\mathcal{Z}^* \subseteq \mathcal{C}$ and $\mathcal{Z}_N \subset \mathcal{C}$ a.s. for large enough N. Problem (4.7) is equivalent to

$$\max_{(x,\lambda) \in \mathcal{C}} \psi(x, \lambda), \tag{5.1}$$

and for large enough N, the SAA problem (4.9) is equivalent to

$$\max_{(x,\lambda) \in \mathcal{C}} \psi_N(x, \lambda), \quad \text{a.s..} \tag{5.2}$$

We now claim that $\psi(x, \lambda)$ is continuous on $\mathcal{C}$ and $\psi_N(x, \lambda)$ converges to $\psi(x, \lambda)$ a.s. uniformly on $\mathcal{C}$. Let ζ be the random variable with uniformly distribution $\bar{U}$ on Θ, and

$$\Psi(x, \lambda, \zeta) := g_1(\zeta)[\mathcal{L}(x, \lambda)(\zeta)]_+ - g_2(\zeta)[\mathcal{L}(x, \lambda)(\zeta)]_- + h(\lambda),$$

be the integrand of $\psi(x, \lambda)$, i.e., $\psi(x, \lambda) = \mathbb{E}_\zeta[\Psi(x, \lambda, \zeta)]$. In order to apply Theorem 7.48 in Shapiro et al. (2009) (restated in Theorem B.4) to verify our claim, we need to show that $\Psi(\cdot, \cdot, \zeta)$ is continuous on $\mathcal{C}$ a.s. and is dominated by a function integrable under the uniform distribution $\bar{U}$.

It is easy to check that $\Psi(x, \lambda, \zeta)$ is continuous for λ a.s.. Since the continuity of $\bar{U}$ implies that, for a fixed $x \in \mathcal{X}$, $\Pr\{\zeta = \xi_k(x)\} = 0$ for $k = 1, \ldots, K$, and all $\xi_k(x)$ are continuous on $\mathcal{X}$ under Assumption (A4), $\varphi_x^0(\zeta) = \sum_{k=1}^K p_k \mathbf{1}\{\xi_k(x) \geq \zeta\}$ and $\Psi(x, \lambda, \zeta)$ are continuous for $x \in \mathcal{X}$ a.s..

We next build an $\bar{U}$-integrable function dominating $\Psi(\cdot, \cdot, \zeta)$. Since $\mathcal{C}$ is compact, there exists $B > 0$ such that $-B \leq \lambda_0 \leq B$ and $\lambda_i \leq B$, $i = 1, \ldots, M$, for $(x, \lambda) \in \mathcal{C}$. Then,

$$
\begin{aligned}
|\Psi(x, \lambda, \zeta)| &\leq (g_1(\zeta) + g_2(\zeta))|\mathcal{L}(x, \lambda)(\zeta)| + |h(\lambda)| \\
&\leq 2g_2(\zeta)\theta_2 - \theta_1 \left(|\varphi_x^0(\zeta)| + |\lambda_0| + \sum_{i=1}^M |\lambda_i||\varphi^i(\zeta)| \right) + |\lambda_0| + \sum_{i=1}^M |\lambda_i||c_i| \\
&\leq 2(1 + B)g_2(\zeta) + 2B \sum_{i=1}^M |\varphi^i(\zeta)|g_2(\zeta) + 1 + B \sum_{i=1}^M |c_i| \\
&=: \Phi(\zeta).
\end{aligned}
$$

The absolute continuity of G_2 under Assumption (A2) implies the boundedness of G_2 on θ, i.e., $b := G_2(\theta_2) < \infty$. Since all φ^i are Lebesgue-integral, we thus have $\int_\Theta |\varphi^i| dG_2/b < \infty$ for $i = 1, \ldots, M$. It follows that

$$
\int_\Theta \Phi(t) d\bar{U}(t) = 2(1 + B)(G_2(\theta_2) - G_2(\theta_1)) + 2bB \sum_{i=1}^M \int_\Theta |\varphi^i(t)| dG_2(t)/b + 1 + B \sum_{i=1}^M |c_i| < \infty,
$$

which verifies that $\Phi(\zeta)$ is integrable under the uniform distribution $\bar{U}$. Hence, $\Psi(x, \lambda, \zeta)$ is dominated by the $\bar{U}$-integrable function $\Phi(\zeta)$ for all $(x, \lambda) \in \mathcal{C}$.

Recall that $\mathcal{C}$ is a nonempty compact set, $\mathcal{Z} \in \mathcal{C}$, $\mathcal{Z}_N \in \mathcal{C}$ a.s. for large enough N. Theorem 5.3 in Shapiro et al. (2009) (restated in Theorem B.2). implies that $y_N \to y^*$ and $\mathbb{D}(\mathcal{Z}_N, \mathcal{Z}^*) \to 0$ a.s. as $N \to \infty$. $\qquad\square$

The following theorem shows that the optimal value of the SAA problem (4.9) converges to the true counterpart of problem (4.7) with an exponential rate.

Theorem 5.3 *Suppose Assumptions (A1) and (A3) - (A5) hold. Let $(x, \lambda)^*$ and $(x, \lambda)_N^*$ be optimal solutions of problems (4.7) and (4.9). If $(x, \lambda)^*$ is unique, then, for any $\epsilon > 0$, there are $\alpha > 0$, $\beta > 0$ such that*

$$
Pr\{\psi((x, \lambda)_N^*) - \psi((x, \lambda)^*) \leq \epsilon\} \leq \alpha e^{-\beta N}, \quad N > 0.
$$

Proof: The absolute continuity of G_1 and G_2 is implied by their Lipschitz continuity under Assumption (A5). Therefore, under Assumptions (A1) and (A3) - (A5), Theorem 5.2 indicates $(x, \lambda)_N^* \to (x, \lambda)^*$ as $N \to \infty$ a.s.. Let K_1 denote the Lipschitz coefficient, i.e.,

$$
|G_2(t_2) - G_2(t_1)| \leq K_1|t_2 - t_1|, \quad t_1, t_2 \in \Theta.
$$

Also according to the boundedness of $\varphi^i(\cdot)$ on Θ, we can assume that there exists $K_2 > 0$ such that $|\varphi^i(t)| \leq K_2$ for all $t \in \Theta$ and $i = 1, \ldots, M$. Furthermore, it follows by Lemma 5.1 that $(x, \lambda)^* \in \mathcal{X} \times \widehat{\Lambda}$, and also $\widehat{\Lambda}$ is a compact set. Then, there exists $K_3 > 0$ such that $\widehat{\Lambda} \subseteq \mathfrak{K} := \{\lambda \in \Re^{m+1} \mid \|\lambda\|_\infty \leq K_3\}$.

Let

$$
\Psi(x, \lambda, \zeta) := \rho_1[\mathcal{L}(x, \lambda)(\zeta)]_+ - \rho_2[\mathcal{L}(x, \lambda)(\zeta)]_- + h(\lambda).
$$

It follows that $\Psi(\cdot)$ is bounded over $\mathcal{X} \times \mathfrak{K} \times \Theta$, since we have, for any $(x, \lambda, \zeta) \in \mathcal{X} \times \mathfrak{K} \times \Theta$,

$$
\begin{aligned}
|\Psi(x, \lambda, \zeta)| &\leq (g_1(\zeta) + g_2(\zeta))|\mathcal{L}(x, \lambda)(\zeta)| + |h(\lambda)| \\
&\leq 2g_2(\zeta)\left(|\varphi_x^0(\zeta)| + |\lambda_0| + \sum_{i=1}^{M} \lambda_i |\varphi^i(\zeta)|\right) + |\lambda_0| + \sum_{i=1}^{M} \lambda_i |c_i| \\
&\leq K_1(1 + K_3 + K_2 K_3 M) + K_3 + K_3 \sum_{i=1}^{M} |c_i|.
\end{aligned}
$$

Because of the uniqueness of $(x, \lambda)^*$, asymptotic convergence of $(x, \lambda)_N^*$, and boundedness of $\Psi(x, \lambda, \zeta)$, we can complete the proof by applying Theorem 3.3 in Dai et al. (2000) (restated in Theorem B.5).

5.1 Heuristic Algorithm

In this section, we harness the asymptotic convergence property of the SAA problem (Theorem 5.2) to provide a feasible solution for problem (4.10) with large sample size. This is done by utilizing the optimal solution of a problem with smaller sample size. We then use this feasible solution for warm start while solving the large sample size problem (LSP) to optimality. Given a LSP instance with sample size N_L and the optimal solution of a small sample size problem (SSP) with $N_S(< N_L)$ samples, our heuristic solves the LSP as follows. Let the optimal solution of SSP be denoted by $(\hat{x}, \hat{\lambda}, \hat{s}, \hat{r})$. First, we solve (4.10) for $N = N_L$ by setting $x = \hat{x}$ and obtain the optimal solution $(\hat{x}, \tilde{\lambda}, \tilde{s}, \tilde{r})$. Then, we resolve (4.10) for $N = N_L$ by setting $r = \tilde{r}$ and obtain a feasible solution for LSP which provides a warm start for solving the LSP to optimality. Our computational results in Section 6.1 show that for portfolio optimization problem (discussed in the next section), this heuristic provides a high quality feasible solution, reducing the integrality gap to on average 0.18% of the optimal solution value.

6 Portfolio Optimization Problem Model and Data

The uncertainty set for John's utility function is summarized as follows:

$$
\mathcal{U}(\rho_1, \rho_2) := \left\{ u \in \mathfrak{U} \,\middle|\, \begin{array}{l} \rho_1 \hat{u}_{\#}(2, 3) \preceq u \preceq \rho_2 \hat{u}_{\#}(2, 3), \\ 0.9 \leq \int_0^2 t\,du \leq 1, \\ 0.8 \leq \int_0^2 t^2\,du \leq 1 \end{array} \right\}.
$$

To check the cost of ambiguous utility discussed in Section 2.2, we consider adjusting the perturbation region around the reference utility as

$$
\widetilde{\mathcal{U}}(\kappa) := (1 - \kappa)\mathcal{U}(1, 1) + \kappa \mathcal{U}(0.5, 2),
$$

where the parameter $\kappa \in [0, 1]$ quantifies the size of a utility set. The cost of ambiguity is $\mathcal{C}(\widetilde{\mathcal{U}}(\kappa))$ by the definition (2.9). In this case, $\mathcal{C}(\widetilde{\mathcal{U}}(\kappa))$ can also be regarded as the sensitivity of the boundary conditions on the reference utility function. Note that the set $\mathcal{U}(1, 1)$ is either empty or it contains the reference utility function $u_{\#}(2, 3)$ only. If the set $\mathcal{U}(1, 1)$ is empty, it means that the boundary and auxiliary conditions give contradictory descriptions of John's risk preference to provide a feasible utility. Adjusting κ allows us to incorporate information from these two assessments.

We use data from the portfolio investment problem studied by Dentcheva and Ruszczyński (2003). This example has $J(= 8)$ assets which are widely used market indexes: U.S. three-month

treasury bills, U.S. long-term government bonds, S&P 500, Willshire 5000, NASDAQ, Lehman Brothers corporate bond index, EAFE foreign stock index, and gold. Dentcheva and Ruszczyński (2003) use $K(= 22)$ yearly returns of these assets as equally probable realizations (see Table 4 in Appendix A).

Let $\xi = (\xi_1, \ldots, \xi_J)$ be a random vector of yearly returns, with equally likely realizations $\xi^k = (\xi_1^k, \ldots, \xi_J^k)$, $k = 1, \ldots, K$, and $p_k = 1/K$. The random outcome of \$1 investment is given by $(1 + x^T \xi)$. Here $x \in \mathcal{X} := \{x \in \mathbb{R}_+^J \mid \sum_{j=1}^J x_j = 1\}$ represents the portfolio investment. John's portfolio investment model is now given as

$$\max_{x \in \mathcal{X}} \min_{u \in \widetilde{\mathcal{U}}(\kappa)} \sum_{k=1}^K p_k u(1 + x^T \xi^k). \tag{6.1}$$

6.1   Computational Experience

All computations are performed running Gurobi 6.0.0 on a PC which has Intel®Core™2 Duo E8400 3.0 GHz Processor and 8 GB of RAM. In view of Theorem 4.2 we reformulate model (6.1) by (4.7), and solve its approximation problem (4.9) with the sample size 300. The optimal portfolios for different $\kappa's$ are given in Table 1. This problem is infeasible for $\kappa \leq 0.3$, which shows the inconsistency in the utility assessments. The solution at $\kappa = 0.4$ is very unstable. Increasing κ to 0.5 results in a large change in the investment in Assets 4 and 5. In comparison, the solutions are stable for $\kappa \geq 0.5$. The optimal solution is unchanged for $0.5 \leq \kappa \leq 0.8$, and also the changes are small when further increasing κ to 0.9 and 1. Theorem 2.6 shows that the cost of ambiguous utility $\mathcal{C}(\widetilde{\mathcal{U}}(\kappa))$ is an increasing concave function of κ. This is shown in Figure 2 for κ in $[0.4, 1]$.

Table 1: Optimal Investment Portfolios (in %)

κ	Asset 1	Asset 2	Asset 3	Asset 4	Asset 5	Asset 6	Asset 7	Asset 8
0 - 0.3	-	-	-	-	-	-	-	-
0.4	0	1.49	0.8	38.53	0	4.14	53.57	1.46
0.5	0	0	0.53	3.38	38.93	0	57.07	0.08
0.6	0	0	0.63	3.27	38.95	0	57.08	0.08
0.7	0	0	0.53	3.38	38.93	0	57.07	0.08
0.8	0	0	0.63	3.27	38.95	0	57.08	0.08
0.9	0.1	0	0	0	41.40	0	58.38	0.11
1	0	0	0	0.21	41.28	0	58.37	0.14

Each row of Table 2 gives the computational performance for solving the SAA of problem (6.1), where $\kappa = 1$, using N samples. We obtain a feasible solution (or lower bound z_S) for a problem instance by utilizing the optimal solution of the problem with N_S samples (refer to our heuristic in Section 5.1 for details), and report the time taken to run the heuristic, denoted by T_S. Note that T_S does not include the time taken to solve the problem with N_S samples and in Table 2, $N_S = 0$ implies that no SSP is solved to get a feasible solution for the LSP instance. We solve formulation (4.10) corresponding to the instance using Gurobi 6.0.0 (default settings) with the feasible solution obtained by our heuristic as a starting solution, and report the optimal solution z_{opt} and initial integrality gap $G\% := 100 \times (z_{opt} - z_S)/z_{opt}$. We set a time limit of 10 hours (or 36000 seconds), i.e. if the optimal solution is not found within 10 hours, the algorithm is terminated and the best solution found within 10 hours is reported. The investment portfolios obtained from these runs are
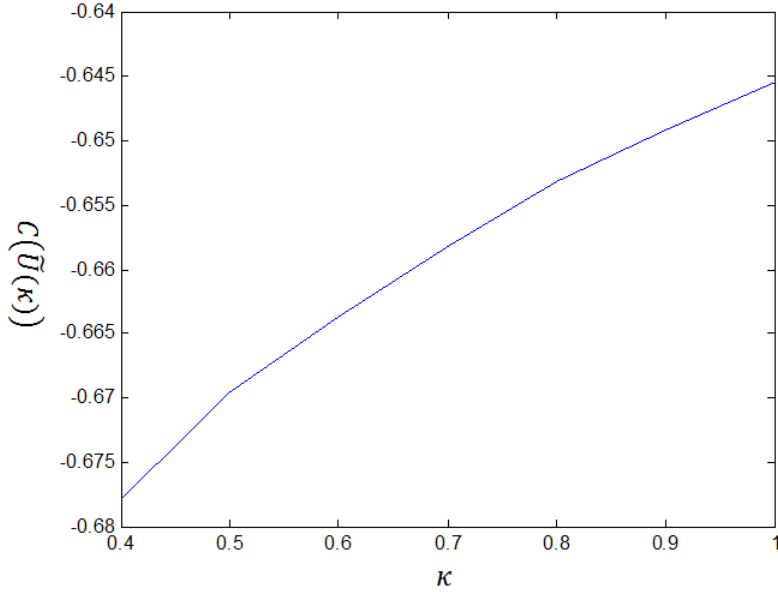
Figure 2: Cost of Ambiguous Utility

Table 2: Results of computational experiments

Sample Size	Problem Size		Heuristic			MIP					
(N)	# BinVar	# Cons	N_S	T_S	z_S	T_1	$Node_1$	z_{opt}	T_{opt}	$Node$	$G\%$
150	3300	3451	0	-	-	11	151	0.6522	57	19714	-
300	6600	6901	0	-	-	39	143	0.6434	212	22099	-
			150	4	0.6398	43	85	0.6434	224	21568	0.56
600	13200	13801	0	-	-	413	3169	0.6466	802	26838	-
			150	12	0.6348	306	667	0.6466	1156	31744	0.44
			300	10	0.6461	233	567	0.6466	1070	31363	0.07
900	19800	20701	0	-	-	7127	51876	0.6438	10279	97565	-
			300	22	0.6434	66	0	0.6438	5423	83353	0.06
			600	25	0.6435	1010	807	0.6438	6326	83281	0.04
1200	26400	27601	0	-	-	3843	20948	0.6438	36000+	61274+	-
			300	64	0.6423	12896	18451	0.6438	36000+	57011+	0.23
			600	62	0.6423	2073	1504	0.6438	36000+	58046+	0.23
1800	39600	41401	0	-	-	15153	20948	0.6438	36000+	58800+	-
			300	115	0.6423	20548	20996	0.6438	36000+	34543+	0.22
			600	103	0.6423	5962	1207	0.6438	36000+	57538+	0.23

given in Table 3. In Table 2, T_{opt} and $Node$ reveal the time (in seconds) taken and total number of branching nodes explored in solving the instance to optimality. However, the solver finds a feasible solution, that is later proved to be the optimal, in T_1 time (seconds) after exploring $Node_1$ nodes.

We make the following observations. The robust utility-based model (P-RVM) is rather challenging to solve. It is because this model is a probability maximization problem. We find that the objective function value and the investment portfolio is accurate to about two decimal places with a SAA of nearly 1000 samples. While the allocation to Assets 5 and 7 are stable, the small fractional investment in Assets 1-3 and 8 change with the sample size. The running time T_{opt} with and

Table 3: Optimal Values and Solutions for Different Sample Sizes

Sample Size	Optimal Value	Optimal Solution (in %)							
		Asset 1	Asset 2	Asset 3	Asset 4	Asset 5	Asset 6	Asset 7	Asset 8
150	0.6522	0.86	0	0	1.31	58.52	0	39.30	0
300	0.6434	0	0	0	0.21	41.28	0	58.37	0.14
600	0.6466	0	0.09	0	0	40.96	0	58.87	0.07
900	0.6438	0	0	0	0.34	41.27	0	58.39	0
1200	0.6438	0	0	0.38	0.88	57.57	0	41.17	0
1800	0.6438	0	0	0	1.04	57.50	0	41.45	0

without using our heuristic also shows a significant growth with the sample size N. However, it is important to observe that for all instances, our heuristic gives a high quality feasible solution in less than 1.8% of the total solution time T_{opt}, resulting in a significant reduction in the integrality gap $G\%$ (which is 0.18% on average). We also notice that for instances which are solved to optimality within time limit, T_1 is on average 27% of T_{opt} and $Node_1$ is on average 8% of $Node$. This implies that a feasible solution, which is the optimal solution, is found very quickly and the remaining time $T_{opt} - T_1$ is spent in proving the optimality. In addition, we compare the results for solving an instance with $(N, N_S) \in \{(600, 150), (600, 300), (900, 300), (900, 600), (1200, 600), (1800, 600)\}$, i.e. N sample sized instance which utilizes the optimal solution of SSP with N_S samples for warm start, to the results for solving the N sample sized instance with no-warm start; and observe that T_1 and $Node_1$ for the former instance are smaller than the ones for latter instance. In fact, for instance with $(N, N_S) = (900, 300)$, the optimal solution is found at the root node with the help of the feasible solution found by our heuristic. This demonstrates the importance of our heuristic.

Remark 6.1 *In a separate run for each problem instance, we add Gomory Mixed Integer (GMI) cuts, derived from rows of the simplex tableau, at the root node. This is done as follows. Given an instance and the optimal basis of its LP relaxation, we randomly select a row u in the tableau, defined by*

$$r_{B_u} + \sum_{t \in NB} \bar{a}_{ut} v_t = \bar{a}_{u0}, \tag{6.2}$$

such that the basic variable $r_{B_u} \in \{r_{k,j} : k = 1, \ldots, K, j = 1, \ldots, N\}$ and $\bar{a}_{u0} \notin \{0, 1\}$, and generate a GMI cut of Gomory (1963) for the row u that is violated by the LP relaxation optimal solution (refer Page 129 of Wolsey (1998) for details), which is updated after adding each cut. After considering N rows for cut generation, we remove the inactive cuts and solve the instance using Gurobi 6.0.0 with its default settings. It is important to note that the GMI cuts are at least as strong as the disjunctive cuts and intersection cuts defined by splits of Balas (1971). We found that the addition of these cuts improve the LP relaxation gap on average from 7.32% to only 5.38% but increased the size of the problem, resulting in an increase in the total solution time. As a topic of future research, we may exploit more efficient methods for solving model 6.1.

6.2 Concluding Remarks

We studied a robust utility-based decision making model to address the issue of ambiguity and inconsistency in utility assessments. In this model, we assumed that the utility function belongs to a set which is described by boundary constraints, using a preference relationship, and auxiliary

requirements, given by 'linear' inequalities. Interestingly, this model is equivalent to the model obtained by using the distributionally robust approach to solve a random target based decision making problem. Note that a target based decision making approach chooses the best policy under which the random outcome meets a given target with the highest probability (e.g. Manski (1988)). Borch (1968) further studied the case with a random target. However, we studied the random target based decision making problem where the random outcome is a function of (continuous or/and discrete) decision variables, the target is random, and only partial information is available on the cdf of this random target.

In order to restrict the risk arising from the uncertainty of the utility function, we maximized the worst-case expected utility of random outcome over the utility set. First, we studied the effect of the uncertain utility on the objective function value of the aforementioned robust utility-based decision making model and then studied the Lagrangian dual of this model by reformulating the Sample Average Approximation (SAA) of the Lagrangian dual problem as a mixed integer linear program. We showed that the optimal value of the SAA problem converges to the optimum value of its true counterpart at an exponential rate. Moreover, we presented a heuristic to provide a feasible solution of the SAA problem by harnessing its convergence property. Finally, we computationally evaluated the effectiveness of our heuristic and the mixed integer linear program in solving the SAA of a portfolio optimization problem.

References

Abbas, A. E. 2007. Moments of utility functions and their applications. *European Journal of Operational Research* **180** 378–395.

Armbruster, B., E. Delage. 2015. Decision making under uncertainty when preference information is incomplete. *Management Science* **61**(1) 111128.

Arrow, Kenneth. J. 1965. *Aspects of the Theory of Risk Bearing*. Yrjö Jahnssonin Säätiö.

Balas, E. 1971. Intersection cuts - a new type of cutting planes for integer programming. *Operations Research* **19** 23–85.

Bazaraa, M. S., H. D. Sherali, C. M. Shetty. 2006. *Nonlinear Programming: Theory and Algorithm*. 3rd ed. John Wiley & Sons, New Jersey.

Berhold, Marvin. H. 1973. The use of distribution functions to represent utility functions. *Management Science* **19**(7) 825–829.

Bertsimas, D., X. V. Doan, K. Natarajan, C. Teo. 2010. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research* **35**(3) 580–602.

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

Bertsimas, D., M. Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

Borch, K. 1968. Decision rules depending on the probability of ruin. *Oxford Economic Papers* **20** 1–10.

Bordley, R., M. LiCalzi. 2000. Decision analysis using targets instead of utility functions. *Decisions in Economics and Finance* **23** 53–74.

Calafiore, Giuseppe. C. 2007. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization* **18**(3) 853–877.

Castagnoli, E., M. LiCalzi. 1996. Expected utility without utility. *Theory and Decision* **41**(3) 281–301.

Chajewska, U., D. Koller, R. Parr. 2000. Making rational decisions using adaptive utility elicitation. *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-00)*. 363–369.

Dai, L., C. H. Chen, J. R. Birge. 2000. Convergence properties of two-stage stochastic programming. *Journal of Optimization Theory and Applications* **106**(3) 489–509.

Delage, Erick., Yinyu. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* **58**(3) 595–612.

Dentcheva, D., A. Ruszczyński. 2003. Optimization with stochastic dominance constraints. *SIAM J. Optim.* **14**(2) 548–566.

Dentcheva, D., A. Ruszczyński. 2004. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Math. Programming* **99** 329–350.

Dentcheva, D., A. Ruszczyński. 2008. stochastic dynamic optimization with discounted stochastic dominance constraints. *SIAM Journal on Control and Optimization* **47**(5) 2540–2556.

Dentcheva, D., A. Ruszczyński. 2009. Optimization with multivariate stochastic dominance constraints. *Math. Programming* **117** 111–127.

Dupacová, J. 1987. The minimax approach to stochastic programming and an illustrative application. *Stochastics* **20** 73–88.

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

Fromberg, D. G., R. L. Kane. 1989. Methodology for measuring health-state preferences - ii: Scaling methods. *Journal of Clinical Epidemiology* **42**(5) 459–471.

Gomory, Ralph E. 1963. An algorithm for integer solutions to linear programs. Robert L. Graves, Philip Wolfe, eds., *Recent Advances in Mathematical Programming*. McGraw-Hill, New York, 269–302.

Hershey, John. C., Paul J. H. Schoemaker. 1985. Probability versus certainty equivalence methods in utility measurement: Are they equivalent? *Management Science* **31**(10) 1213–1231.

Homem-de-Mello, T., S Mehrotra. 2009. A cutting surface method for linear programs with linear stochastic dominance constraints. *SIAM Journal on Optimization* **20**(3) 1250–1273.

Hu, J., T. Homem-de-Mello, S Mehrotra. 2012. Sample average approximation of stochastic dominance constrained programs. *Mathematical Programming* **133**(1-2) 171–201.

Hu, J., T. Homem-de-Mello, S Mehrotra. 2013. Stochastically weighted stochastic dominance concepts with an application in capital budgeting. *European Journal of Operational Research* Publish online at `http://dx.doi.org/10.1016/j.ejor.2013.08.007`.

Hu, J., S. Mehrotra. 2012. Robust and stochastically weighted multi-objective optimization models and reformulations. *Operations Research* **60**(4) 936–953.

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

Jacquet-Lagrèze, E., Y. Siskos. 1982. Assessing a set of additive utility functions for multicriteria decision making: The uta method. *European Journal of Operational Research* **10**(2) 151–164.

Karmarkar, Uday. S. 1978. Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance* **21**(1) 61–72.

Keeney, R. L., H. Raiffa. 1976. *Decisions with multiple objectives: preferences and value tradeoffs.* John Wiley & Sons, New York.

Li, Jonathan. Y., Roy. H. Kwon. 2013. Portfolio selection under model uncertainty: a penalized moment-based optimization approach. *Journal of Global Optimization* **56**(1) 131–164.

Luedtke, James, Shabbir Ahmed, George L. Nemhauser. 2010. An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming* **122**(2) 247–272.

Manski, C.F. 1988. Ordinal utility models of decision making under uncertainty. *Theory and Decision* **25** 79–104.

Mehrotra, Sanjay., David. Papp. 2014. A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. `http://arxiv.org/abs/1306.3437` to appear SIAM Journal on Optimization.

Müller, A., D. Stoyan. 2002. *Comparison Methods for Stochastic Models and Risks.* John Wiley & Sons, Chichester.

Nord, Erik. 1992. Methods for quality adjustment of life years. *Social Science and Medicine* **34**(5) 559–569.

Pflug, G., D. Wozabal. 2007. Ambiguity in portfolio selection. *Quantitative Finance* **7**(4) 435–442,.

Pflug, Georg., Alois. Pichlera, David. Wozabalb. 2012. The 1/n investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance* **36**(2) 410–417.

Pratt, John. W. 1964. Risk aversion in the small and in the large. *Econometrica* **32** 122–136.

Prékopa, A. 1995. *Stochastic Programming.* Kluwer Academic.

Rockafellar, R. Tyrrell. 1970. *Convex Analysis.* Princeton University Press.

Scarf, H. 1958. A min-max solution of an inventory problem. *Studies in the Mathematical Theory of Inventory and Production*, chap. 12. 201–209.

Shapiro, A., D. Dentcheva, A. Ruszczyński. 2009. *Lectures on stochastic programming : Modeling and theory.* SIAM.

Shapiro, Alexander., Shabbir. Ahmed. 2004. On a class of minimax stochastic programs. *SIAM J. Optim.* **14**(4) 1237–1249.

Thurstone, L. L. 1927. A law of comparative judgement. *Psychological Review* **34** 278–286.

Train, K. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York.

Tversky, A., D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5** 297–323.

von Neumann, J., O. Morgenstern. 1947. *Theory of Games and Economic Behavior*. 2nd ed. Princeton University Press, Princeton, NJ.

Wakker, Peter., Daniel. Deneffe. 1996. Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Science* **42**(8) 1131–1150.

Weber, M. 1987. Decision making with incomplete information. *European Journal of Operational Research* **28**(1) 44–57.

Wolsey, L. A. 1998. *Integer Programming*. Wiley, New York, USA.

Wozabal, David. 2012. A framework for optimization under ambiguity. *Annals of Operations Research* **193**(1) 21–47.

A   Asset Return Data

The following table provides asset returns given in Dentcheva and Ruszczyński (2003).

B   Known Results used in the paper

Lemma B.1 *[Theorem 3.2 in Rockafellar (1970)] Let $\mathfrak{U}$ be a convex set. For $\kappa_1, \kappa_2 \geq 0$, we have $(\kappa_1 + \kappa_2)\mathfrak{U} = \kappa_1\mathfrak{U} \oplus \kappa_2\mathfrak{U}$.*

Let $\mathcal{X}$ be a convex set and ξ^i, $i = 1, \ldots, N$, be iid samples of a random vector ξ with support Ξ. Denote $g(x) := \mathbb{E}[G(x, \xi)]$ and $g_N(x) := \frac{1}{N}\sum_{i=1}^{N} G(x, \xi^i)$ where $G(\cdot)$ is a well-defined function on $\mathcal{X} \times \Xi$. By these notions, two optimization problems are described as

$$y* := \min_{x \in \mathcal{X}} g(x), \quad y_N := \min_{x \in \mathcal{X}} g_N(x).$$

Let $\mathcal{Z}*$ and $\mathcal{Z}_N$ be the sets of optimal solutions x^* and x_N of these optimization problems respectively.

Theorem B.2 (Theorem 5.3 in Shapiro et al. (2009)) *Suppose that there exists a compact set $\mathcal{C} \in \mathcal{X}$ such that: (i) $\mathcal{Z}*$ is nonempty and is contained in $\mathcal{C}$, (ii) the function $g(x)$ is finite valued and continuous on $\mathcal{C}$, (iii) $g_N(x)$ converges to $g(x)$ a.s. as $N \to \infty$ uniformly in $x \in \mathcal{C}$, and (iv) Z_N is nonempty and is contained in $\mathcal{C}$ for N large enough a.s.. Then $y_N \to y^*$ and $\mathbb{D}(Z_N, Z^*) \to 0$ a.s. as $N \to \infty$.*

Theorem B.3 (Theorem 5.4 in Shapiro et al. (2009)) *Suppose that: (i) the integrand function G is random lower semicontinuous, (ii) for almost every $\xi \in \Xi$ the function $G(\cdot, \xi)$ is convex, (iii) the set $\mathcal{X}$ is closed and convex, (iv) the expected value function g is lower semicontinuous and there exists a point $\bar{x} \in \mathcal{X}$ such that $g(x) < \infty$ for all x in a neighborhood of $\bar{x}$, (v) the set Z^* of optimal solutions of the true problem is nonempty and bounded, and (vi) the LLN holds pointwise. Then $y_N \to y^*$ and $\mathbb{D}(Z_N, Z^*) \to 0$ a.s. as $N \to \infty$.*

Table 4: Asset Returns (in %) in Dentcheva and Ruszczyński (2003)

Year	Asset 1	Asset 2	Asset 3	Asset 4	Asset 5	Asset 6	Asset 7	Asset 8
1	7.5	-5.8	-14.8	-18.5	-30.2	2.3	-14.9	67.7
2	8.4	2.0	-26.5	-28.4	-33.8	0.2	-23.2	72.2
3	6.1	5.6	37.1	38.5	31.8	12.3	35.4	-24.0
4	5.2	17.5	23.6	26.6	28.0	15.6	2.5	-4.0
5	5.5	0.2	-7.4	-2.6	9.3	3.0	18.1	20.0
6	7.7	-1.8	6.4	9.3	14.6	1.2	32.6	29.5
7	10.9	-2.2	18.4	25.6	30.7	2.3	4.8	21.2
8	12.7	-5.3	32.3	33.7	36.7	3.1	22.6	29.6
9	15.6	0.3	-5.1	-3.7	-1.0	7.3	-2.3	-31.2
10	11.7	46.5	21.5	18.7	21.3	31.1	-1.9	8.4
11	9.2	-1.5	22.4	23.5	21.7	8.0	23.7	-12.8
12	10.3	15.9	6.1	3.0	-9.7	15.0	7.4	-17.5
13	8.0	36.6	31.6	32.6	33.3	21.3	56.2	0.6
14	6.3	30.9	18.6	16.1	8.6	15.6	69.4	21.6
15	6.1	-7.5	5.2	2.3	-4.1	2.3	24.6	24.4
16	7.1	8.6	16.5	17.9	16.5	7.6	28.3	-13.9
17	8.7	21.2	31.6	29.2	20.4	14.2	10.5	-2.3
18	8.0	5.4	-3.2	-6.2	-17.0	8.3	-23.4	-7.8
19	5.7	19.3	30.4	34.2	59.4	16.1	12.1	-4.2
20	3.6	7.9	7.6	9.0	17.4	7.6	-12.2	-7.4
21	3.1	21.7	10.0	11.3	16.2	11.0	32.6	14.6
22	4.5	-11.1	1.2	-0.1	-3.2	-3.5	7.8	-1.0

23

Theorem B.4 (Theorem 7.48 in Shapiro et al. (2009)) *Let $\mathcal{X}$ be a nonempty compact set and suppose that: (i) $G(\cdot, \xi)$ is continuous on $\mathcal{X}$ a.s. and (ii) $G(x, \xi)$, $x \in \mathcal{X}$, is dominated by an integrable function. Then the expected value function $g(x)$ is finite valued and continuous on $\mathcal{X}$, and $g_N(x)$ converges to $g(x)$ a.s. uniformly on $\mathcal{X}$.*

Theorem B.5 (Theorem 3.3 in Dai et al. (2000)) *Suppose that x^* is unique and $x_N \to x^*$ a.s., and that there exist a neighborhood $\widehat{X} \subset \mathcal{X}$ of x^* and positive constants γ_0, M such that*

$$\mathbb{E}\left[e^{\gamma |G(x, \xi)|}\right] \leq C$$

for all $x \in \widehat{X}$ and for all $\gamma \in [0, \gamma_0]$. Then, for any $\epsilon > 0$, there are $\alpha > 0$, $\beta > 0$ such that

$$P(y_N - y* \geq \epsilon) \leq \alpha e^{-\beta N}, \quad \text{for all } N > 0.$$

C   Proof of Proposition 2.4

Proof: Given $u_1, u_2 \in \mathfrak{U}$, let $u := (1 - \kappa)u_1 + \kappa u_2$ for an arbitrarily chosen $\kappa \in [0, 1]$. It is easy to see $u \in \mathcal{M}$ and u satisfies the probability measure condition (2.5b). Since $(1 - \kappa)G_1 \preceq (1 - \kappa)u_1$ and $\kappa G_1 \preceq \kappa u_2$, we have

$$G_1 = (1 - \kappa)G_1 + \kappa G_1 \preceq (1 - \kappa)u_1 + \kappa u_2 = u,$$

and similarly, $u \preceq G_2$. It shows that u satisfies the boundary conditions (2.5c). We now check the satisfaction of auxiliary condition (2.5d). It follows that

$$\int_\Theta \varphi^i(t)du(t) = (1 - \kappa)\int_\Theta \varphi^i(t)du_1(t) + \kappa \int_\Theta \varphi^i(t)du_2(t) \leq (1 - \kappa)c_i + \kappa c_i = c_i.$$

$\square$

D   Proof of Lemma 2.5

Proof: By definition, for any $u \in \mathfrak{U}(\kappa_1)$, there exist $u_1 \in \mathfrak{U}_1$ and $u_2 \in \mathfrak{U}_2$ such that

$$u = (1 - \kappa_1)u_1 + \kappa_1 u_2$$

$$= \left(1 - \frac{\kappa_1}{\kappa_2}\right)u_1 + \frac{\kappa_1}{\kappa_2}((1 - \kappa_2)u_1 + \kappa_2 u_2).$$

Because $u_1 \in \mathfrak{U}_1 \subseteq \mathfrak{U}(\kappa_2)$, $(1 - \kappa_2)u_1 + \kappa_2 u_2 \in \mathfrak{U}(\kappa_2)$, and $\mathfrak{U}(\kappa_2)$ is a convex set by Theorem 3.1 in Rockafellar (1970), we have $u \in \mathfrak{U}(\kappa_2)$. $\square$

E   Proof of Theorem 2.6

The proof here follows the steps in Hu and Mehrotra (2012) where a similar result is obtained in the context of ambiguous trade-off weights in multi-objective optimization. *Proof:* Let us choose

κ_1, κ_2 such that $0 \le \kappa_1 \le \kappa_2 \le 1$. It follows by Lemma 2.5 that $\mathfrak{U}(\kappa_1) \subseteq \mathfrak{U}(\kappa_2)$ so that $C(\mathfrak{U}(\kappa_1)) \le C(\mathfrak{U}(\kappa_2))$. We now prove the concavity of $C(\mathfrak{U}(\cdot))$. For $\delta \in [0,1]$, we have

$$
\begin{aligned}
C(\mathfrak{U}((1-\delta)\kappa_1 + \delta\kappa_2)) &= -\max_{x \in \mathcal{X}} \min_{u \in \mathfrak{U}((1-\delta)\kappa_1 + \delta\kappa_2)} \int_\Theta \varphi_x^0(t) du(t) \\
&= -\max_{x \in \mathcal{X}} \min_{u \in (1-\delta)\mathfrak{U}(\kappa_1) \oplus \delta\mathfrak{U}(\kappa_2)} \int_\Theta \varphi_x^0(t) du(t) \qquad \text{(by Lemma B.1)} \\
&= -\max_{x \in \mathcal{X}} \min_{\substack{u_1 \in \mathfrak{U}(\kappa_1) \\ u_2 \in \mathfrak{U}(\kappa_2)}} \int_\Theta \varphi_x^0(t) \, d[(1-\delta)u_1(t) + \delta u_2(t)] \\
&\ge -\left((1-\delta) \max_{x \in \mathcal{X}} \min_{u_1 \in \mathfrak{U}(\kappa_1)} \int_\Theta \varphi_x^0(t) du_1(t) + \delta \max_{x \in \mathcal{X}} \min_{u_2 \in \mathfrak{U}(\kappa_2)} \int_\Theta \varphi_x^0(t) du_2(t) \right) \\
&= (1-\delta)C(\mathfrak{U}(\kappa_1)) + \delta C(\mathfrak{U}(\kappa_2)).
\end{aligned}
$$

$\square$

F Proof of Theorem 4.1

We fix $x \in \mathcal{X}$ and first prove the weak duality. We denote

$$
\alpha(u) := \int_\Theta \varphi_x^0(t) du(t)
$$

$$
\beta_0(u) := \int_\Theta du(t) - 1
$$

$$
\beta_i(u) = \int_\Theta \varphi^i(t) du(t) - c_i, \ i = 1, \ldots, M.
$$

For any feasible solution $u \in \mathfrak{U}$ of the primal problem (2.5) and feasible solution λ with $\lambda_i \ge 0$ for $i = 1, \ldots, M$ of the dual problem (4.4) we have

$$
\psi(x, \lambda) \le L(x, u, \lambda) = \alpha(u) + \sum_{i=0}^M \lambda_i \beta_i(u) \le \alpha(u),
$$

which implies weak duality. Next, we show the strong duality under Assumption (A1). Let γ be the optimal value of the primal problem (2.5). By Assumption (A1), problem (2.5) is feasible, and thus $0 \le \gamma \le 1$ because of (2.2) and (2.3). Let us now consider the system

$$
\alpha(u) - \gamma < 0, \ \beta_0(u) = 0, \ \beta_i(u) \le 0, \ i = 1, \ldots, M, \ \text{for some } u \in \mathcal{D}, \tag{F.1}
$$

and the set

$$
\Lambda = \left\{ (p, q) \in \mathbb{R}^{M+2} \mid p > \alpha(u) - \gamma, \ q_0 = \beta_0(u), \ q_i \ge \beta_i(u), \ i = 1, \ldots, M, \ \text{for some } u \in \mathcal{D} \right\}.
$$

Since $\alpha(\cdot)$ and $\beta_i(\cdot)$, $i = 0, \ldots, M$, are affine, and $\mathcal{D}$ is convex, it is easy to see that Λ is a convex set. Since γ is the optimal objective value of (2.5), the system (F.1) has no solution, and hence, Λ excludes the zero vector. Now using the separating hyperplane theorem (Bazaraa et al., 2006, Corollary 1 to Theorem 2.4.7), there exists a nonzero $(\hat{\lambda}, \lambda)$ such that

$$
\hat{\lambda} p + \sum_{i=0}^M \lambda_i q_i \ge 0, \quad \text{for all } (p, q) \in \text{cl } \Lambda.
$$

Now fix $u \in \mathcal{D}$. Since p and q_i, $i = 1, \ldots, M$, can be made arbitrarily large, the above inequality holds only if $\hat{\lambda} \geq 0$ and $\lambda_i \geq 0$ for $i = 1, \ldots, M$. Furthermore, $(\alpha(u) - \gamma, \beta(u)) \in \text{cl } \Lambda$ for all $u \in \mathcal{D}$. Therefore, we have

$$\hat{\lambda}(\alpha(u) - \gamma) + \sum_{i=0}^{M} \lambda_i \beta_i(u) \geq 0, \text{ for all } u \in \mathcal{D}. \tag{F.2}$$

We now claim that $\hat{\lambda} > 0$. By contradiction, suppose that $\hat{\lambda} = 0$. By Assumption (A1) there exists an interior point $\hat{u} \in \mathfrak{U}$, i.e., $G_1 \prec \hat{u} \prec G_2$, $\beta_0(\hat{u}) = 0$, and $\beta_i(\hat{u}) < 0$ for $i = 1, \ldots, M$. Substituting in (F.2), it follows that $\sum_{i=1}^{M} \lambda_i \beta_i(\hat{u}) \geq 0$, which implies that $\lambda_i = 0$ since $\beta_i(\hat{u}) < 0$ and $\lambda_i \geq 0$. Therefore, we have $\lambda_0 \beta_0(u) \geq 0$ for all $u \in \mathcal{D}$. Since $G_1 \prec \hat{u}$, we have $\beta_0(G_1) < 0$, and thus $\lambda_0 = 0$. It contradicts with the fact that $(\hat{\lambda}, \lambda)$ is nonzero in the separating hyperplane theorem.

Dividing (F.2) by $\hat{\lambda}$ and denoting $\tilde{\lambda}_i = \lambda_i / \hat{\lambda}$ for $i = 0, \ldots, M$, we get

$$\alpha(u) + \sum_{i=0}^{M} \tilde{\lambda}_i \beta_i(u) \geq \gamma, \text{ for all } u \in \mathcal{D},$$

which shows that $\psi(x, \tilde{\lambda}) \geq \gamma$. It implies that the optimal value of the dual problem (4.4) is greater than or equal to the optimal value of the primal problem (2.5). The strong duality follows since we have already established the weak duality result.

G   Lemma G.1

Lemma G.1 *Let $\hat{u} \in \mathcal{M}$ satisfy $\int_\Theta \varphi^i(t) d\hat{u}(t) < c_i$. For $\kappa_1 \in [0, 1)$ and $\kappa_2 \in (1, \infty)$, denote*

$$\widehat{\mathcal{D}}(\hat{u}) := \{u \in \mathcal{M} \mid \kappa_1 \hat{u} \preceq u \preceq \kappa_2 \hat{u}\}, \tag{G.1}$$

and consider the value of function

$$\phi_{\hat{u}}(x, \lambda) := \inf_{u \in \widehat{\mathcal{D}}(\hat{u})} L(x, u, \lambda).$$

At an unbounded $\hat{\lambda}$, i.e., $\|\hat{\lambda}\| = \infty$, we have

$$\phi_{\hat{u}}(x, \hat{\lambda}) = -\infty, \quad x \in \mathcal{X}.$$

Proof: We state the proof for an arbitrarily given $x \in \mathcal{X}$, and express the function $\phi_{\hat{u}}(x, \lambda)$ as

$$\begin{aligned}
\phi_{\hat{u}}(x, \lambda) &= \inf_{u \in \widehat{\mathcal{D}}(\hat{u})} L(x, u, \lambda) \\
&= \kappa_1 \int_\Theta [\mathcal{L}(x, \lambda)(t)]_+ d\hat{u}(t) - \kappa_2 \int_\Theta [\mathcal{L}(x, \lambda)(t)]_- d\hat{u}(t) + h(\lambda) \\
&= \kappa_1 \int_\Theta \mathcal{L}(x, \lambda)(t) d\hat{u}(t) - (\kappa_2 - \kappa_1) \int_\Theta [\mathcal{L}(x, \lambda)(t)]_- d\hat{u}(t) + h(\lambda) \\
&= \kappa_1 \int_\Theta \varphi_x^0(t) d\hat{u}(t) + (\kappa_1 - 1)\lambda_0 + \sum_{i=1}^{M} \lambda_i \left(\kappa_1 \int_\Theta \varphi^i(t) d\hat{u}(t) - c_i\right) \\
&\quad - (\kappa_2 - \kappa_1) \int_\Theta \left[\varphi_x^0(t) + \lambda_0 + \sum_{i=1}^{M} \lambda_i \varphi^i(t)\right]_- d\hat{u}(t).
\end{aligned}$$

26

Let

$$\widehat{\phi}_{\hat{u}}(x, \lambda) := \kappa_1 \int_\Theta \varphi_x^0(t) d\hat{u}(t) + (\kappa_1 - 1)\lambda_0 + \sum_{i=1}^M \lambda_i \left(\kappa_1 \int_\Theta \varphi^i(t) d\hat{u}(t) - c_i \right)$$

$$- (\kappa_2 - \kappa_1) \left[\int_\Theta \varphi_x^0(t) + \lambda_0 + \sum_{i=1}^M \lambda_i \varphi^i(t) \ d\hat{u}(t) \right]_- .$$

Since $\int_\Theta -[g(t)]_- d\hat{u}(t) \leq -[\int_\Theta g(t) d\hat{u}(t)]_-$ for any measurable function g, we have

$$\widehat{\phi}_{\hat{u}}(x, \lambda) \geq \phi_{\hat{u}}(x, \lambda). \tag{G.2}$$

In the following two cases we first discuss the value of the function $\widehat{\phi}_{\hat{u}}(x, \hat{\lambda})$ at an unbounded $\hat{\lambda}$.
Case 1: If

$$\int_\Theta \varphi_x^0(t) + \lambda_0 + \sum_{i=1}^M \lambda_i \varphi^i(t) d\hat{u}(t) \leq 0, \tag{G.3}$$

then we have

$$\widehat{\phi}_{\hat{u}}(x, \lambda) = (\kappa_2 - 1) \int_\Theta \varphi_x^0(t) + \lambda_0 + \sum_{i=1}^M \lambda_i \varphi^i(t) \ d\hat{u}(t) + \int_\Theta \varphi_x^0(t) d\hat{u}(t) + \sum_{i=1}^M \lambda_i \left(\int_\Theta \varphi^i(t) d\hat{u}(t) - c_i \right)$$

$$\leq \int_\Theta \varphi_x^0(t) d\hat{u}(t) + \sum_{i=1}^M \lambda_i \left(\int_\Theta \varphi^i(t) d\hat{u}(t) - c_i \right).$$

Since $\int_\Theta \varphi^i(t) d\hat{u}(t) < c_i$, any unbounded $\hat{\lambda}_i$ for $i \in \{1, \ldots, M\}$ results in $\hat{\phi}_{\hat{u}}(x, \hat{\lambda}) = -\infty$. Clearly, $\hat{\phi}_{\hat{u}}(x, \hat{\lambda}) = -\infty$ when $\hat{\lambda}_0 = -\infty$ and $\hat{\lambda}_i < \infty$ for $i \in \{1, \ldots, M\}$. Now suppose $\hat{\lambda}_0 = \infty$. To satisfy (G.3), we have $\hat{\lambda}_i = \infty$ for some $i \in \{1, \ldots, M\}$ so that $\hat{\phi}_{\hat{u}}(x, \hat{\lambda}) = -\infty$.
Case 2: If

$$\int_\Theta \varphi_x^0(t) + \lambda_0 + \sum_{i=1}^M \lambda_i \varphi^i(t) d\hat{u}(t) \geq 0, \tag{G.4}$$

then

$$\widehat{\phi}_{\hat{u}}(x, \lambda) = (\kappa_1 - 1) \int_\Theta \varphi_x^0(t) + \lambda_0 + \sum_{i=1}^M \lambda_i \varphi^i(t) d\hat{u}(t) + \int_\Theta \varphi_x^0(t) d\hat{u}(t) + \sum_{i=1}^M \lambda_i \left(\int_\Theta \varphi^i(t) d\hat{u}(t) - c_i \right)$$

$$\leq \int_\Theta \varphi_x^0(t) d\hat{u}(t) + \sum_{i=1}^M \lambda_i \left(\int_\Theta \varphi^i(t) d\hat{u}(t) - c_i \right).$$

Similarly, $\hat{\phi}_{\hat{u}}(x, \hat{\lambda}) = -\infty$ when any $\hat{\lambda}_i$ is unbounded or $\hat{\lambda}_0 = \infty$. Also to satisfy (G.4), $\hat{\lambda}_0 = -\infty$ needs some unbounded $\hat{\lambda}_i$. Consequently, $\hat{\phi}_{\hat{u}}(x, \hat{\lambda}) = -\infty$.
The above shows that $\widehat{\phi}_{\hat{u}}(x, \hat{\lambda}) = -\infty$ for any unbounded $\hat{\lambda}$. The inequality (G.2) shows that

$$\phi_{\hat{u}}(x, \hat{\lambda}) \leq \widehat{\phi}_{\hat{u}}(x, \hat{\lambda}) = -\infty.$$

□

H   Proof of Lemma 5.1

At the first step of the proof, we show that $\widehat{\Lambda}$ is a nonempty compact set and $\mathcal{Z}^*$ is bounded. Next, we verify the almost sure boundedness of $\mathcal{Z}_N$ for large enough sample size N.

Step 1. We first state the proof for a fixed $x \in \mathcal{X}$. By Assumption (A1), let $\hat{u}$ be an interior point of $\mathfrak{U}$. There are $\kappa_1 \in [0, 1)$ and $\kappa_2 \in (1, \infty)$ such that

$$G_1 \preceq \kappa_1 \hat{u} \prec \kappa_2 \hat{u} \preceq G_2.$$

By Lemma G.1 we have that, at an unbounded $\hat{\lambda}$,

$$\phi_{\hat{u}}(x, \hat{\lambda}) = \inf_{u \in \widehat{\mathcal{D}}(\hat{u})} L(x, u, \hat{\lambda}) = -\infty.$$

Recall that $\mathcal{D}(\hat{u})$ is denoted as in (G.1). Since $\widehat{\mathcal{D}}(\hat{u}) \subseteq \mathcal{D}$, we have

$$\psi(x, \hat{\lambda}) \leq \phi_{\hat{u}}(x, \hat{\lambda}) = -\infty. \tag{H.1}$$

Since $\psi(x, 0) = \int_\Theta \varphi_x^0(t) dG_1(t) \geq 0$ and $\psi(x, \cdot)$ is a continuous concave function for all $x \in \mathcal{X}$, $\widehat{\Lambda}$ is a closed nonempty set. Also, its boundedness is ensured by (H.1). On the other hand, denoting $\Lambda^*(x)$ as the set of optimal solutions of problem (4.4), we have a uniform lower bound on the optimal value of problem (4.4) unrelated to this arbitrary $x \in \mathcal{X}$, $\psi(x, \lambda^*(x)) \geq \psi(x, 0)$ for any $\lambda^*(x) \in \Lambda^*(x)$. Therefore, $\Lambda^*(x) \subseteq \widehat{\Lambda}$ for any $x \in \mathcal{X}$. The compactness of $\mathcal{X}$ is given by Assumption (A3). Then $\mathcal{Z}^* \subseteq \mathcal{X} \times \widehat{\Lambda}$ is bounded.

Step 2. Under Assumption (A2), the SAA problem of (4.4) for any $x \in \mathcal{X}$ is given by

$$\max_\lambda \psi_N(x, \lambda)$$

$$\text{s.t. } \lambda_i \geq 0, \quad i = 1, \ldots, M.$$

Let us denote the set of its optimal solutions by $\Lambda_N(x)$. Recall that $\psi(x, \cdot)$ is continuous and concave in λ, and $\psi(x, 0) \geq 0$. It implies that there exists an interior point $\bar{\lambda} \in \{\lambda \in \mathbb{R}^{M+1} \mid \lambda_i \geq 0, i = 1, \ldots, M\}$, which is the feasible region of problem (4.4), such that $\psi(x, \lambda) > -\infty$ for all λ in a neighborhood of $\bar{\lambda}$. It is easy to see that the feasible region of problem (4.4) is closed and convex and that the corresponding integrand shown in (4.8) is concave in λ. The proof at step 1 also shows that the set of optimal solutions $\Lambda^*(x)$ is nonempty and is contained in the bounded set $\widehat{\Lambda}$. It follows from Theorem 5.4 in Shapiro et al. (2009) (restated in Theorem B.3 for completeness) that $\mathbb{D}(\Lambda_N(x), \Lambda^*(x)) \to 0$ a.s. as $N \to \infty$. For a given $\epsilon > 0$, let us choose a compact neighborhood $\mathcal{C}$ of $\widehat{\Lambda}$,

$$\mathcal{C} := \left\{ \lambda \in \mathbb{R}^{M+1} \,\middle|\, \inf_{\hat{\lambda} \in \widehat{\Lambda}} \|\lambda - \hat{\lambda}\|_2 \leq \epsilon \right\}.$$

Then, there exists $\hat{N}(x) > 0$ such that, for all $N \geq \hat{N}(x)$, $\Lambda_N(x) \subseteq \mathcal{C}$ a.s.. We now claim that

$$N^* := \max_{x \in \mathcal{X}} \hat{N}(x) < \infty, \text{a.s.}.$$

Otherwise, because $\mathcal{X}$ is compact under Assumption (A3), there is a sample region Ω with $\bar{U}(\Omega) > 0$ ($\bar{U}$ is the uniform distribution on Θ) such that, for any sample realization of $\omega \in \Omega$, $\hat{N}(\hat{x}) = \infty$ for some $\hat{x} \in \mathcal{X}$. Then,

$$\Pr\left\{ \limsup_{N \to \infty} \mathbb{D}(\Lambda_N(\hat{x}), \Lambda^*(\hat{x})) > \epsilon/2 \right\} \geq \bar{U}(\Omega) > 0,$$

which is a contradiction. Therefore, we have $\Lambda_N(x) \subseteq \mathcal{C}$ for all $x \in \mathcal{X}$ when $N \geq N^*$ a.s.. Then $\mathcal{Z}_N \subseteq \mathcal{X} \times \mathcal{C}$ for $N \geq N^*$ a.s..

I   Proof of Proposition 3.1

Using integration by parts, we have

$$\int_{\theta_1}^{\theta_2} k(t - \theta_1)^{k-1} u(t) dt = (\theta_2 - \theta_1)^k - \int_{\theta_1}^{\theta_2} t^k du(t) = (\theta_2 - \theta_1)^k - \mathbb{E}[\zeta^k].$$

for all $k \in \mathbb{Z}_+$. It thus follows that

$$\mathbb{E}[\zeta^k] = (\theta_2 - \theta_1)^k - (\theta_2 - \theta_1)^k \int_{\theta_1}^{\theta_2} u(t) dt^k (\theta_2 - \theta_1)^{-k} = \theta^k (1 - \mathbb{E}[u(\vartheta_k)]).$$