MANAGEMENT SCIENCE
Vol. 61, No. 1, January 2015, pp. 111–128
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2014.2059

Decision Making Under Uncertainty When Preference Information Is Incomplete

Benjamin Armbruster
Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208, armbruster@northwestern.edu

Erick Delage
Department of Decision Sciences, HEC Montréal, Montréal, Québec H3T 2A7, Canada, erick.delage@hec.ca

We consider the problem of optimal decision making under uncertainty but assume that the decision maker's utility function is not completely known. Instead, we consider all the utilities that 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- and second-order stochastic dominance framework. We then give tractable formulations for such decision-making problems. We formulate them as robust utility maximization problems, as optimization problems with stochastic dominance constraints, and as robust certainty equivalent maximization problems. We use a portfolio allocation problem to illustrate our results.

Keywords: expected utility; robust optimization; stochastic dominance; certainty equivalent
History: Received June 29, 2012; accepted August 5, 2014, by Dimitris Bertsimas, optimization.

1. Introduction

This paper questions a key and rarely challenged assumption of decision making under uncertainty: that decision makers can always, after a tolerable amount of introspective questioning, clearly identify the utility function that characterizes their attitude toward risk. The use of expected utility to characterize attitudes toward risk is pervasive. In large part, this is due to Von Neumann and Morgenstern (1944), who prove that any set of preferences that a decision maker may have among risky outcomes can be characterized by an expected utility measure if the preferences respect certain reasonable axioms (i.e., completeness, transitivity, continuity, and independence). Specifically, there exists a utility function $u: \mathbb{R} \to \mathbb{R}$ so that among two random variables (or lotteries), W and Y, the decision maker prefers W to Y if and only if $\mathbb{E}[u(W)] \geq \mathbb{E}[u(Y)]$.

There has been much effort on determining how to choose a utility function for a decision maker, and this work plays an integral part in the design of surveys for assessing tolerance to financial risk (Grable and Lytton 1999). The method for choosing a utility function proposed in most textbooks on decision analysis (see, for instance, Clemen and Reilly 2001) is to make a set of pairwise comparisons between lotteries (often using the Becker-DeGroot-Marschak reference lotteries; Becker et al. 1964) in order to identify the value of the utility function at a discrete set of points. The utility function is then completed by naïve interpolation. A more sophisticated approach assumes that the utility function has a parametric structure such as constant absolute or constant relative risk aversion. For example, if a decision maker can confirm that he is risk averse and that his preference between *any* two lotteries is invariant to the addition of *any* constant amount to all outcomes, then that decision maker has constant absolute risk aversion; thus, his utility is of the form $u(y) = 1 - e^{-\gamma y}$. Parameters are then resolved using a small number of pairwise comparisons between lotteries.

These approaches have important shortcomings. If they do not assume a parametric form, then the large or even continuous space of outcomes may require a lot of interpolation or asking the decision maker many questions. Even interpolation may not be easy, because if the questions to the decision maker are binary choices between two lotteries, then the answers will not provide the value of the utility function at any point; instead, each answer will provide merely a single linear constraint on the values of the utility function on the support of these two lotteries. To justify a parametric form for the utility function, a decision maker must be able to confidently address a question about an infinite number of lottery comparisons (such as that described above for utilities with constant absolute risk aversion). A more fundamental limitation is that all these procedures conclude by selecting a *single* "most likely" utility function given the evidence. In other words, these procedures entirely disregard other plausible choices and the inherent ambiguity of those choices. In this paper, we will focus on these instances where knowledge can only be gathered using a small number of simple questions, and meaningful decisions must be made even though no single utility can be unambiguously identified.

Our approach follows in spirit the line of work in the artificial intelligence literature on utility elicitation using optimization for problems that only involve a finite, although possibly large, set of outcomes. This line of work emphasizes that utility elicitation and decision analysis should be combined into a single process in order to use all the information collected about the true utility function when making a decision. In this context, Chajewska et al. (2000) represents the knowledge of the decision maker's preferences using a probability distribution over utility functions and then judges a decision by its expected utility averaged over the distribution of utilities. To increase their knowledge of the utility function, they use a value of information criterion to select the next question to the decision maker. In contrast to this probabilistic approach, in Boutilier et al. (2006), the authors construct the set $\mathcal{U}$ of all utility functions that do not contradict the available information. They then identify the decision that achieves minimum worst-case regret (i.e., regret experienced a posteriori once the true utility function is revealed) using a mixed-integer programming approach and exploiting the assumed "generalized additive" structure of the true utility. In comparison, our paper considers uncertain real-valued outcomes and proposes formulations that are more natural for decision making and can be reduced to convex optimization problems.

We motivate our discussion with the following stochastic program:

$$\max_{x\in\mathcal{X}} \mathbb{E}[u(h(x,\xi))],$$

where x is a vector of decision variables, $\mathcal{X}$ is a set of implementable decisions, and $h(x,\xi)$ is a function mapping the decision x to a random return indexed by the scenario ξ; the expectation is over the random scenarios ξ. We assume that we have not gathered enough information to uniquely specify $u(\cdot)$. Thus we build on the theory developed by Aumann (1962) of expected utility without the completeness axiom. This theory suggests that our incomplete preferences can be characterized by a *set* of utility functions $\mathcal{U}$ (Dubra et al. 2004). This set describes our incomplete information about $u(\cdot)$ and is known to contain the true utility function. Another situation where preferences are incomplete is when groups make decisions by consensus: here, $\mathcal{U}$ contains the utility functions of the group members, and two lotteries are incomparable if the group members do not agree on which is preferred.

The set $\mathcal{U}$ suggests that we face a robust optimization problem. Our approach will differ, however, from the typical robust optimization framework, which is robust to the possible realizations or distributions of ξ (see, for example, Ben-Tal and Nemirovski 1998, Delage and Ye 2010, and references therein). Instead, we are robust to the possible utilities in $\mathcal{U}$ and choose the worst-case utility function.

When the range of $h(x,\xi)$ is not restricted to a discrete set, the only existing way of dealing with ambiguity in the utility function is a stochastic program with a stochastic dominance constraint (Dentcheva and Ruszczyński 2003):

$$\max_{x\in\mathcal{X}} \ \mathbb{E}[f(x,\xi)]$$
$$\text{s.t.} \ \ h(x,\xi)\succeq Z,$$

with some objective function f. In these problems the stochastic dominance constraint, $h(x,\xi)\succeq Z$, is defined as $\mathbb{E}[u(h(x,\xi))]\geq \mathbb{E}[u(Z)]$ for all utility functions $u\in\mathcal{U}$. This constraint ensures that the random consequences of the chosen action, $h(x,\xi)$, are preferred to those of a baseline random variable Z for all utility functions in $\mathcal{U}$. For first-order dominance, $\mathcal{U}$ is the set of all increasing functions, and for second-order dominance, it is the set of all increasing concave functions. The limitations of stochastic dominance constraints are threefold: first, stochastic dominance does not provide guidance with respect to choosing an objective function f; second, the choice of a baseline Z is not a trivial one to make; and third, the set $\mathcal{U}$ is very large in the case of first- and second-order dominance, and thus the stochastic dominance constraint may be very restrictive.

We briefly describe the four main contributions of this paper.

1. In a context where preferences information is incomplete, to the best of our knowledge, we provide for the first time tractable solution methods that can account for information that takes the shape of comparisons between specific lotteries. In particular, we will show how the worst-case difference between expected utilities,

$$\inf_{u\in\mathcal{U}} \left(\mathbb{E}[u(h(x,\xi))]-\mathbb{E}[u(Z)]\right),$$

or even

$$\inf_{u\in\mathcal{U}} \mathbb{E}[u(h(x,\xi))],$$

can be expressed as the maximum of a linear programming problem of reasonable size. This is done by exploiting the fact that these comparisons can be represented as linear constraints in the space of utility functions. The importance of this contribution comes from the realization that lottery comparisons are fundamental building blocks for representing one's preferences regarding risk. It can, for instance, be observed in many risk tolerance assessment surveys used by financial advisers, as these typically involve questions such as the following, from Grable and Lytton (1999, p. 170):

> You are on a TV game show and can choose one of the following. Which would you take?
> (a) $1,000 in cash
> (b) A 50% chance at winning $5,000
> (c) A 25% chance at winning $10,000
> (d) A 5% chance at winning $100,000

That we can handle such comparisons in a tractable way opens the door to a wide range of possibilities, one of them being the allowance of more flexibility in describing the set of utilities $\mathcal{U}$ involved in a stochastic dominance constraint.[1]

2. We present for the first time the robust (i.e., worst-case) certainty equivalent formulation,

$$\max_{x\in\mathcal{X}} \inf_{u\in\mathcal{U}} u^{-1}(\mathbb{E}[u(h(x,\xi))]),$$

(see §2 for a precise definition of "certainty equivalent") and show how under mild conditions it can be reduced to solving a small number of linear programs of reasonable size. In fact, this performance measure is a natural one to employ when there is ambiguity about the decision maker's risk preferences, as it provides solutions that we know are preferred to the highest amount of guaranteed return. In particular, this measure has a meaningful set of units (the same ones as $h(x,\xi)$), unlike utility measures that can be scaled arbitrarily. The set of utilities required for tractability is the same as in first contribution and are discussed in §3. Note that the concept of *optimized certainty equivalent* defined in Ben-Tal and Teboulle (2007), which falls in the class of convex risk measures, is a completely different concept; intuitively, it can be seen as a best-case instead of a worst-case approach, and it does not involve ambiguity about the utility function.

3. Given a number of lottery comparisons, we provide a natural way of detecting when a decision maker is inconsistent in his stated preferences (i.e., makes a set of comparisons that together violate the axioms of the expected utility framework) by verifying whether or not a certain linear program is feasible. In case of inconsistency, we are able to identify the "closest" set of feasible preferences (or closest feasible utility function) and quantify the "degree of infeasibility."

4. We measure for the first time the potential value that is added to the decision as more knowledge of the decision maker's preferences is gathered, starting from simple knowledge of risk aversion to exact knowledge of the utility function that characterizes his preferences. We do this by evaluating the difference in the optimal worst-case certainty equivalent with and without the additional information. We believe similar insights should be obtained in situations where one is worried about worst-case expected utility. This idea could potentially be used to help choose among a set of questions/comparisons or when deciding whether the necessary effort required to ask these questions is worth the gain.

In the next section we describe three formulations (including the stochastic dominance formulation) that can be used instead of maximizing expected utility when the decision maker's utility function is only known to lie inside a set $\mathcal{U}$. In §3 we describe the sets of utilities $\mathcal{U}$ and how to optimize each formulation with these sets. We then present numerical examples involving a portfolio allocation problem in §4. Section 5 describes extensions of the framework to allow the detection and correction of inconsistent behavior and to account for characteristics of the utility function that are associated with "almost stochastic dominance." We conclude in §6.

2. Formulations

Our work examines three formulations for decision making when one knows the utility function is in some set $\mathcal{U}$. These formulations involve (1) optimizing with a stochastic dominance constraint,

$$\begin{aligned}\max_{x\in\mathcal{X}} \quad & f(x)\\ \text{s.t.} \quad & \mathbb{E}[u(h(x,\xi))] \ge \mathbb{E}[u(Z)] \quad \forall\, u\in\mathcal{U},\end{aligned} \tag{1}$$

where Z is some reference random variable; (2) maximizing the worst-case utility,

$$\max_{x\in\mathcal{X}} \inf_{u\in\mathcal{U}} \mathbb{E}[u(h(x,\xi))]; \tag{2}$$

and (3) maximizing the worst-case (or robust) certainty equivalent,

$$\max_{x\in\mathcal{X}} \inf_{u\in\mathcal{U}} \mathbb{C}_u[h(x,\xi)], \tag{3}$$

where the certainty equivalent of a lottery (i.e., random variable) X given a utility function u is typically defined as the amount for sure such that one would be indifferent between it and the lottery; that is, $u^{-1}(\mathbb{E}[u(X)])$. To ensure uniqueness, we slightly modify this definition to $\mathbb{C}_u[X] := \sup\{s\colon\ u(s) \le \mathbb{E}[u(X)]\}$. The robust certainty equivalent formulation (3) maximizes $\inf_{u\in\mathcal{U}} \mathbb{C}_u[h(x,\xi)]$, the largest amount of money we know for sure we would be willing to exchange

[1] Note that in this paper, we focus on the definition of stochastic dominance that involves the comparison of expected utility under a set of utility functions. We leave the question open as to how the conclusions that we will draw might be interpreted in terms of comparing the results of different integration operations on the cumulative density functions.

for the lottery $h(x, \xi)$. In the context of group decision making, using the worst-case utility function means accommodating the group's least favored member.

Because we do not know the true utility function in these formulations, any choice from $\mathcal{U}$ is as justifiable as any other. We use the worst-case utility function for convenience: that choice turns out to make these formulations very tractable. In addition to convenience, we can also motivate the choice of utility from $\mathcal{U}$ with an analogy to Rawls' (1971) *A Theory of Justice*. Rawls proposes that one imagines deciding the structure of society behind a "veil of ignorance," i.e., without knowing one's place in society. Although our decision maker's choices are less weighty, his ignorance of their true utility function is somewhat analogous. Rawls then argues that this leads one to focus on the least advantaged in society and suggests a max-min principle for allocating goods. Similarly, we focus on the least favorable utility function using max-min formulations.

Because we seek convex formulations, we will assume that the feasible set $\mathcal{X}$ is convex, the objective function f in (1) is concave, the function $h(x, \xi)$ relating the action to a random outcome is concave in x, and the utilities in $\mathcal{U}$ are risk averse to ensure that the objective in (2) is concave in x (the only exception is when we discuss S-shaped utilities). For computational tractability we also assume that all the random variables have finite support. We assume that there are M scenarios for ξ, $\Omega := \{\xi_1, \ldots, \xi_M\}$ with associated probabilities $p_i := \mathbb{P}[\xi = \xi_i]$.

The key to our success is determining tractable representations of

$$\psi(x; \mathcal{U}, Z) := \inf_{u \in \mathcal{U}} \big(\mathbb{E}[u(h(x, \xi))] - \mathbb{E}[u(Z)]\big), \qquad (4)$$

where we sometimes drop the dependence on $\mathcal{U}$ and Z from our notation. Using $\psi(x; \mathcal{U}, Z)$, we can write the stochastic dominance formulation (1) as

$$\begin{aligned} \max_{x \in \mathcal{X}} \quad & f(x) \\ \text{s.t.} \quad & \psi(x; \mathcal{U}, Z) \geq 0 \end{aligned}$$

and the worst-case utility formulation (2) as

$$\max_{x \in \mathcal{X}} \psi(x; \mathcal{U}, 0),$$

where we chose $Z := 0$ a.s. Unlike the other formulations, the robust certainty equivalent formulation is not concave but quasiconcave (see the proof in Appendix A). Thus we can solve it using a bisection algorithm.

Remark 1. Although it might be tempting to straightforwardly adopt the worst-case expected utility formulation (2) when considering ambiguity about the choice of utility function, one must consider with care that when maximizing worst-case expected utility, one implicitly compares random variables using a hidden (and potentially meaningless) set of ordered lotteries, which tends in particular to favor a risk-neutral attitude. We refer the reader to Appendix C for a thorough discussion.

Remark 2. We do not study the worst-case regret formulation

$$\min_{x \in \mathcal{X}} \sup_{u \in \mathcal{U}} \max_{x' \in \mathcal{X}} \big(E[u(h(x', \xi))] - E[u(h(x, \xi))]\big)$$

proposed in Boutilier et al. (2006), for two reasons. First, from a decision-theoretic point of view, minimax regret as a choice function violates the independence to irrelevant alternatives condition, which is essential for rationalizing preferences (see Arrow 1959). That condition states that our preference between decision x_1 and x_2 should not be influenced by the set of alternatives $\mathcal{X}$. Second, it is likely to be an intractable problem when $\mathcal{U}$ is a general convex set. Intuitively, the reason is that evaluating the worst-case regret associated with a fixed x reduces to solving

$$\sup_{u \in \mathcal{U},\, x' \in \mathcal{X}} \int E[u(y)\delta_y(h(x', \xi)) - u(y)\delta_y(h(x, \xi))]\, dy,$$

where $\delta_y(\cdot)$ is the Dirac measure. Unfortunately, the cross term $u(y)\delta_y(h(x', \xi))$ prevents this from being a convex optimization problem.

3. Worst-Case Utilities

The following are three common hypotheses about a decision maker's utility function.

1. *Risk aversion:* A decision maker is risk averse if for any lottery X, he prefers $\mathbb{E}[X]$ for sure over the lottery X itself. This is characterized by the concavity of the utility function.

2. *S-shape:* Prospect theory was proposed by Kahneman and Tversky (1979) to bridge the gap between normative theories of rational behavior and behavior observed by experimentalists. This theory conjectures that preferences are affected by four factors. First, outcomes are evaluated with respect to a reference point. Second, decision makers are more affected by losses than by winnings. Third, the perception of winnings or losses is diminished as they get larger. Finally, the perception of probabilities is biased (i.e., overweighting smaller probabilities and underweighting larger ones). These observations suggest that the decision maker is risk averse with respect to *gains* and risk seeking with respect to *losses*. Specifically, it suggests an S-shaped utility function that is concave for gains and convex for losses. As is typically done in the context of a normative study, in what follows, we will

disregard the possibility of any probability assessment bias and focus on how to account for information that indicates that the utility function has this particular shape.

3 *Prudence:* In Eeckhoudt and Schlesinger (2006), prudence captures the fact that a decision maker is more risk tolerant in situations where he can achieve higher returns. In particular, given any lottery involving two outcomes with equal probability of occurring, a prudent decision maker will prefer adding a zero-mean risk Z to the outcome with the largest value.[2] Prudence is a stronger condition than risk aversion and, as shown in Appendix E, is characterized by the existence and convexity of the derivative of the utility function. It is also commonly referred to as decreasing absolute risk aversion.

In what follows, we present tractable reformulation for evaluating $\psi(\cdot;\mathscr{U})$ for three different types of sets $\mathscr{U}$ that are formed from intersections of the following sets of utility functions:

$$U_2 := \{u\colon u \text{ is nondecreasing and concave}\},$$

$$U_s := \{u\colon u \text{ is nondecreasing, convex on } (-\infty,0], \text{ and concave on } [0,\infty)\},$$

$$U_3 := \{u\colon u' \text{ exists and is convex}\},$$

$$U_n := \{u\colon \mathbb{E}[u(W_0)] - \mathbb{E}[u(Y_0)] = 1\},$$

$$U_a := \{u\colon \mathbb{E}[u(W_k)] \geq \mathbb{E}[u(Y_k)] \quad \forall\, k = 1,\ldots,K\}.$$

Here, $W_0,\ldots,W_K, Y_0,\ldots,Y_K$ are given random variables representing lotteries. The set of risk-averse utilities is denoted by U_2; the set of S-shaped convex–concave utilities (and the only exception to the assumption throughout the paper that utilities are concave) is denoted by U_s; the set of prudent utilities, those with convex u', is denoted by U_3; and the set of utilities that prefer lottery W_k to lottery Y_k for all k is denoted by U_a. Since adding a constant to a utility or multiplying it by a positive constant results in an equivalent utility, it is often necessary to normalize utilities. There are multiple ways of normalizing utilities. Here, we use U_n to specify the scaling, specifying that the utility difference between W_0 and Y_0 is 1. For example, assuming that $W_0 := 1$ and $Y_0 := 0$ a.s. enforces that $u(1) - u(0) = 1$.

As the choices of $\mathscr{U}$, we focus on $\mathscr{U}^2 := U_a \cap U_n \cap U_2$, $\mathscr{U}^s := U_a \cap U_n \cap U_s$, and $\mathscr{U}^3 := U_a \cap U_n \cap U_2 \cap U_3$. These choices all incorporate U_a, allowing one to tailor the problem to the specific preferences of a particular decision maker, whether he be entirely risk averse, risk seeking over losses, or prudent. For example, $\mathscr{U}^2$ with no specific preferences, i.e., $K = 0$, reduces to the set defining second-order dominance. We now present finite dimensional linear programming reformulations of $\psi(x;\mathscr{U},Z)$ for these choices of $\mathscr{U}$. Although the reformulations will be exact for $\mathscr{U}^2$ and $\mathscr{U}^s$, the reformulations will lead to a conservative approximation of high precision for $\mathscr{U}^3$. In the cases of $\mathscr{U}^2$ and $\mathscr{U}^3$, the reformulations can easily be reintegrated in the optimization model for x and give rise to a convex optimization problem of reasonable size.

The notation used in the following results will refer to $\mathscr{S}$ as the joint support of all static random variables, $\mathscr{S} := \operatorname{supp}(Z) \cup \bigcup_{k=0}^{K} (\operatorname{supp}(Y_k) \cup \operatorname{supp}(W_k))$, and we will use $\bar{y}_j$ to denote the jth smallest entry of $\mathscr{S}$. For clarity of exposure, scenarios in Ω will always be indexed by i, outcomes in $\mathscr{S}$ by j, and queries by k. Thus the size of our optimization problems is specified by the number of queries K, the number of scenarios M, and the size of the support $N := |\mathscr{S}|$.

3.1. Incorporating Lottery Comparisons

We first address how to account for the results of K lottery comparisons for a decision maker known to be risk averse. Specifically, in this case evaluating $\psi(x;\mathscr{U},Z)$ requires characterizing the optimal value of the infinite dimensional problem

$$\inf_{u \in \mathscr{U}^2} \big(\mathbb{E}[u(h(x,\xi))] - \mathbb{E}[u(Z)]\big).$$

Our main result states that this value can be computed by solving a finite dimensional linear program of reasonable size as it involves $2(N+M)$ variables and $MN + K + M + 2N - 1$ constraints (not counting the nonnegativity constraints).

Theorem 1. *The optimal value of the linear program*

$$\min_{\alpha,\beta,v,w} \sum_i p_i(v_i h(x,\xi_i) + w_i) - \sum_j \mathbb{P}[Z = \bar{y}_j]\alpha_j \tag{5a}$$

$$\text{s.t.} \quad \bar{y}_j v_i + w_i \geq \alpha_j \quad \forall\, i \in \{1,\ldots,M\},\ j \in \{1,\ldots,N\}, \tag{5b}$$

$$\sum_j \Big(\mathbb{P}[W_0 = \bar{y}_j]\alpha_j - \mathbb{P}[Y_0 = \bar{y}_j]\alpha_j\Big) = 1, \tag{5c}$$

$$\sum_j \mathbb{P}[W_k = \bar{y}_j]\alpha_j \geq \sum_j \mathbb{P}[Y_k = \bar{y}_j]\alpha_j \quad \forall\, k = 1,\ldots,K, \tag{5d}$$

$$(\alpha_{j+1} - \alpha_j) \geq \beta_{j+1}(\bar{y}_{j+1} - \bar{y}_j) \quad \forall\, j \in \{1,\ldots,N-1\}, \tag{5e}$$

$$(\alpha_{j+1} - \alpha_j) \leq \beta_j(\bar{y}_{j+1} - \bar{y}_j) \quad \forall\, j \in \{1,\ldots,N-1\}, \tag{5f}$$

$$v \geq 0, \quad \beta \geq 0, \tag{5g}$$

[2] In the economics literature (see, for instance, Leland 1968), a prudent attitude is said to be defined by the need for larger precautionary savings when facing a riskier situation. Here, we adopt a definition that does not rely on comparing amounts of money received now versus later and is therefore closer in spirit to the definition of risk aversion. Both of these definitions translate as imposing that $u'(y)$ exists and is convex.

equals $\psi(x;\mathcal{U}^2)$. Furthermore, a worst-case utility function (i.e., one achieving the infimum in (4)) is

$$u^*(y)=\begin{cases}\alpha_N & y\ge\bar{y}_N,\\ \dfrac{\alpha_{j+1}-\alpha_j}{\bar{y}_{j+1}-\bar{y}_j}y+\dfrac{\bar{y}_{j+1}\alpha_j-\bar{y}_j\alpha_{j+1}}{\bar{y}_{j+1}-\bar{y}_j} & \bar{y}_j\le y<\bar{y}_{j+1}\\ & \forall j\in\{1,\dots,N-1\},\\ -\infty & y<\bar{y}_1.\end{cases}\tag{6}$$

This is a piecewise linear function connecting the points $u(\bar{y}_j)=\alpha_j$, which equals $-\infty$ for $y<\bar{y}_1$ and α_N for $y\ge\bar{y}_N$ and which has supergradient $\beta_j\in\partial u(\bar{y}_j)$. Here, y is a dummy outcome variable and not related to the $\{\bar{y}_j\}$. Note that according to this utility function, outcomes below $\bar{y}_1$ are infinitely bad.

We present a detailed proof of this result because the ideas that are used will be reused in the proofs of Theorems 2 and 3. Intuitively, (5a) represents the difference in utilities, (5c) normalizes the utilities, (5e) and (5f) ensure concavity, and $\beta\ge 0$ ensures the utility is nondecreasing.

Proof. We first partition the set of utility functions by their values at the points in $\mathcal{S}$, letting $U(\alpha):=\{u\colon u(\bar{y}_j)=\alpha_j\ \forall j\}$. Hence,

$$\psi(x;\mathcal{U}^2)=\min_{\alpha}\ \psi(x;U(\alpha)\cap\mathcal{U}^2),\qquad U(\alpha)\cap\mathcal{U}^2\neq\varnothing.$$

Note that $U(\alpha)$ is either a subset of U_a or is disjoint from it. The same is true with respect to U_n. Since $\mathcal{U}^2:=U_a\cap U_n\cap U_2$, it then follows that

$$\psi(x;\mathcal{U}^2)=\min_{\alpha}\ \psi(x;U(\alpha)\cap U_2),\qquad U(\alpha)\cap U_2\neq\varnothing,\ U(\alpha)\subseteq U_a,\ U(\alpha)\subseteq U_n.$$

The constraint $U(\alpha)\cap U_2\neq\varnothing$ is represented by (5e) and (5f), and $\beta\ge 0$, $U(\alpha)\subseteq U_a$ by (5d), and $U(\alpha)\subseteq U_n$ by (5c). Note that $\mathbb{E}[u(Z)]$ is a constant, $\sum_j\mathbb{P}[Z=\bar{y}_j]\alpha_j$, for $u\in U(\alpha)$. Thus, evaluating $\psi(x;U(\alpha)\cap U_2)$ is equivalent to minimizing $\mathbb{E}[u(h(x,\xi))]$ over $u\in U(\alpha)\cap U_2$. Among the nondecreasing concave functions in $U(\alpha)$, this is minimized by the piecewise linear function u^* in (6), which essentially forms a convex hull of the points $(\bar{y}_j,\alpha_j)$ with the additional requirement that the function be nondecreasing. Hence when $U(\alpha)\cap U_2\neq\varnothing$, then the function u^* in (6) is a worst-case utility function for $\psi(x;U(\alpha)\cap U_2)$ (i.e., achieves the infimum in (4)). Then, $\psi(x;U(\alpha)\cap U_2)=\mathbb{E}[u^*(h(x,\xi))]-\sum_j\mathbb{P}[Z=\bar{y}_j]\alpha_j$. Since u^* is concave and nondecreasing,

$$u^*(y)=\min_{v\ge 0,w}\ vy+w\tag{7a}$$

$$\text{s.t.}\quad v\bar{y}_j+w\ge\alpha_j\quad\forall j\in\{1,\dots,N\}.\tag{7b}$$

Substituting $y=h(x,\xi)$ for every i gives us the objective (5a) and the constraints (5b) and $v\ge 0$. □

Remark 3. An alternative way to ensure concavity of the utility functions would be to replace constraints (5e), (5f), and $\beta\ge 0$ by $\alpha_{j+1}=\alpha_j+\beta_j(\bar{y}_{j+1}-\bar{y}_j)$ and $\beta_{j+1}\le\beta_j$ for all $j\in\{1,\dots,N-1\}$, where we consider $\beta_N=0$. We used the form that is presented, as it relates more naturally to the definition of a concave function $u(\bar{y}_{j+1})\le u(\bar{y}_j)+\nabla u(\bar{y}_j)^T(\bar{y}_{j+1}-\bar{y}_j)$, where $\nabla u(y)$ refers to a supergradient of $u(y)$. This form could therefore easily be generalized to the context of multiattribute utility functions, which we leave as a future direction of research to explore.

This formulation allows us to efficiently solve problems (1), (2), and (3). To solve (1) and (3), we look at the dual of problem (5). This allows us to write $\psi(x;\mathcal{U}^2)\ge 0$ using the dual variables $\mu\in\mathbb{R}^{N\times M}$, $\nu_0\in\mathbb{R}$, $\nu\in\mathbb{R}^K$, $\lambda^{(1)}\in\mathbb{R}^{N-1}$, and $\lambda^{(2)}\in\mathbb{R}^{N-1}$ as well as the following constraints:

$$\nu_0\ge 0,\tag{8a}$$

$$\sum_i\mu_{i,j}-(\mathbb{P}(W_0=\bar{y}_j)-\mathbb{P}(Y_0=\bar{y}_j))\nu_0-\sum_k(\mathbb{P}(W_k=\bar{y}_j)-\mathbb{P}(Y_k=\bar{y}_j))\nu_k+(\lambda^{(1)}_j-\lambda^{(1)}_{j-1})-(\lambda^{(2)}_j-\lambda^{(2)}_{j-1})=\mathbb{P}(Z=\bar{y}_j)\quad\forall j,\tag{8b}$$

$$\lambda^{(2)}_j(\bar{y}_{j+1}-\bar{y}_j)-\lambda^{(1)}_{j-1}(\bar{y}_j-\bar{y}_{j-1})\le 0\quad\forall j,\tag{8c}$$

$$\sum_j\bar{y}_j\mu_{i,j}\le p_ih(x,\xi_i)\quad\forall i,\tag{8d}$$

$$\sum_j\mu_{i,j}=p_i\quad\forall i,\tag{8e}$$

$$\mu\ge 0,\ \nu\ge 0,\ \lambda^{(1)}\ge 0,\ \lambda^{(2)}\ge 0,\tag{8f}$$

where we consider $\lambda^{(1)}_0=\lambda^{(2)}_0=0$. All constraints are linear in the decision variables except for (8d), which is a convex constraint in x if $h(\cdot,\xi)$ is concave. For the stochastic dominance constrained problem (1), we simply add these constraints and variables to the problem; for the robust certainty equivalent problem (3), we check their feasibility a small number of times. In the case of the robust utility maximization problem (2), we let $Z=0$, then take the dual formulation, and then combine the two stages of minimization to get

$$\max_{x\in\mathcal{X}}\inf_{u\in\mathcal{U}}\ \mathbb{E}[u(h(x,\xi))]=\max_{\substack{x\in\mathcal{X}\\ \mu,\nu_0,\nu,\lambda^{(1)},\lambda^{(2)}}}\ \nu_0\quad\text{s.t. (8b)–(8f).}$$

3.2. Incorporating S-Shape Information

We assume that $y=0$ is the reference point (i.e., inflection point) for the S-shaped utility function. For simplicity, we will include 0 in $\mathcal{S}$ and define the sets $\mathcal{J}^+=\{j\colon\bar{y}_j\ge 0\}$ and $\mathcal{J}^-=\{j\colon\bar{y}_j\le 0\}$. The following theorem is similar to Theorem 1. Since the proof is also similar, we defer it to the Appendix B.

Theorem 2. *The optimal value of the linear program*

$$\min_{\alpha,\beta,v,w,s} \sum_i p_i\big(1\{h(x,\xi_i)<0\}s_i+1\{h(x,\xi_i)\geq 0\} \cdot(v_i h(x,\xi_i)+w_i)\big)-\sum_j \mathbb{P}[Z=\bar{y}_j]\alpha_j \tag{9a}$$

$$\text{s.t.}\quad \bar{y}_j v_i+w_i\geq \alpha_j \quad \forall i\in\{1,\ldots,M\}, j\in\mathcal{J}^+, \tag{9b}$$

$$s_i\geq \beta_j(h(x,\xi_i)-\bar{y}_j)+\alpha_j \quad \forall i\in\{1,\ldots,M\}, j\in\mathcal{J}^-, \tag{9c}$$

$$\sum_j \mathbb{P}[W_0=\bar{y}_j]\alpha_j-\sum_j \mathbb{P}[Y_0=\bar{y}_j]\alpha_j=1, \tag{9d}$$

$$\sum_j \mathbb{P}[W_k=\bar{y}_j]\alpha_j\geq \sum_j \mathbb{P}[Y_k=\bar{y}_j]\alpha_j \quad \forall k=1,\ldots,K, \tag{9e}$$

$$\alpha_{j+1}-\alpha_j\geq \beta_{j+1}(\bar{y}_{j+1}-\bar{y}_j) \quad \forall j\in\mathcal{J}^+\backslash\{N\}, \tag{9f}$$

$$\alpha_{j+1}-\alpha_j\leq \beta_j(\bar{y}_{j+1}-\bar{y}_j) \quad \forall j\in\mathcal{J}^+\backslash\{N\}, \tag{9g}$$

$$\alpha_j-\alpha_{j-1}\leq \beta_j(\bar{y}_j-\bar{y}_{j-1}) \quad \forall j\in\mathcal{J}^-, \tag{9h}$$

$$\alpha_j-\alpha_{j-1}\geq \beta_{j-1}(\bar{y}_j-\bar{y}_{j-1}) \quad \forall j\in\mathcal{J}^-, \tag{9i}$$

$$v\geq 0,\ \beta\geq 0, \tag{9j}$$

equals $\psi(x;\mathcal{U}^s)$. *Furthermore, a worst-case utility function (i.e., one achieving the infimum in* (4)) *is*

$$u^*(y)=\begin{cases}\alpha_N & y\geq \bar{y}_N,\\ \dfrac{\alpha_{j+1}-\alpha_j}{\bar{y}_{j+1}-\bar{y}_j}y+\dfrac{\bar{y}_{j+1}\alpha_j-\bar{y}_j\alpha_{j+1}}{\bar{y}_{j+1}-\bar{y}_j} & \bar{y}_j\leq y<\bar{y}_{j+1}\ \forall j\in\mathcal{J}^+,\\ \max_{j'\in\{j,j+1\}}(\beta_{j'}(y-\bar{y}_{j'})+\alpha_{j'}) & \bar{y}_j\leq y<\bar{y}_{j+1}\ \forall j\in\mathcal{J}^-,\\ \beta_1(y-\bar{y}_1)+\alpha_1 & y<\bar{y}_1.\end{cases} \tag{10}$$

This is a piecewise linear function connecting the points $u(\bar{y}_j)=\alpha_j$.

Unfortunately, the general problems (1), (2), and (3) are probably hard to solve under $\mathcal{U}^s$ because even maximizing expected utility with an S-shaped utility function may lead to multiple local maxima. Nevertheless, Theorem 2 allows us to evaluate $\psi(x;\mathcal{U}^s)$ (and, potentially, its derivatives using linear programming sensitivity analysis), despite its infinite dimensional nature. This suggests that nonlinear optimization methods that accept black-box representations of the objective function should be applicable. Such methods rely on an oracle that can evaluate efficiently the objective function $g(x)$ for a fixed x. In particular, considering the robust certainty equivalent formulation presented in problem (3), one could easily consider applying derivative-free optimization methods (see Conn et al. 2009 for a complete survey) to the problem $\max_{x\in\mathcal{X}} g(x)$, where $g(x):=\inf_{u\in\mathcal{U}}\mathbb{C}_u[h(x,\xi)]$. Here, $g(x)$ can be evaluated by applying a bisection algorithm to find the largest value z such that the optimal value of the linear program (9) with $Z:=z$ almost surely is greater than or equal to 0.

3.3. Incorporating Prudence Information

Our results are weaker for $\mathcal{U}^3$. We will assume that $[a,b]$ contains the support of all the random variables involved in this problem. We then discretize this interval, adding values to $\mathcal{S}$ to minimize the largest gap $\bar{y}_{j+1}-\bar{y}_j$.

Theorem 3. *The optimal value of the linear program*

$$\hat{\psi}(x):=\min_{\alpha,\beta,\gamma,v,w}\left\{\sum_i p_i(v_i h(x,\xi_i)+w_i)-\sum_j \mathbb{P}[Z=\bar{y}_j]\alpha_j\right\} \tag{11a}$$

$$\text{s.t.}\quad \bar{y}_j v_i+w_i\geq \alpha_j \quad \forall i\in\{1,\ldots,M\},\quad j\in\{1,\ldots,N\}, \tag{11b}$$

$$\sum_j \mathbb{P}[W_0=\bar{y}_j]\alpha_j-\sum_j \mathbb{P}[Y_0=\bar{y}_j]\alpha_j=1, \tag{11c}$$

$$\sum_j \mathbb{P}[W_k=\bar{y}_j]\alpha_j\geq \sum_j \mathbb{P}[Y_k=\bar{y}_j]\alpha_j \quad \forall k=1,\ldots,K, \tag{11d}$$

$$\alpha_{j+1}-\alpha_j\geq \beta_{j+1}(\bar{y}_{j+1}-\bar{y}_j) \quad \forall j\in\{1,\ldots,N-1\}, \tag{11e}$$

$$\alpha_{j+1}-\alpha_j\leq \beta_j(\bar{y}_{j+1}-\bar{y}_j) \quad \forall j\in\{1,2,\ldots,N-1\}, \tag{11f}$$

$$\beta_{j+1}-\beta_j\leq \gamma_{j+1}(\bar{y}_{j+1}-\bar{y}_j) \quad \forall j\in\{1,\ldots,N-1\}, \tag{11g}$$

$$\beta_{j+1}-\beta_j\geq \gamma_j(\bar{y}_{j+1}-\bar{y}_j) \quad \forall j\in\{1,\ldots,N-1\}, \tag{11h}$$

$$\beta\geq 0,\ \boldsymbol{\gamma}\leq 0,\ \mathrm{v}\geq 0, \tag{11i}$$

is a lower bound for $\psi(x;\mathcal{U}^3)$. *Furthermore, an approximate worst-case utility function is the piecewise linear function*

$$\hat{u}^*(y)=\begin{cases}\alpha_N & y\geq \bar{y}_N,\\ \dfrac{\alpha_{j+1}-\alpha_j}{\bar{y}_{j+1}-\bar{y}_j}y+\dfrac{\bar{y}_{j+1}\alpha_j-\bar{y}_j\alpha_{j+1}}{\bar{y}_{j+1}-\bar{y}_j} & \bar{y}_j\leq y<\bar{y}_{j+1}\ \forall j\in\{1,\ldots,N-1\},\\ -\infty & y<\bar{y}_1.\end{cases} \tag{12}$$

The proof can be found in the Appendix D. As the discretization becomes finer, we expect that the approximate value $\hat{\psi}(x)$ and approximate worst-case utility function $\hat{u}^*(\cdot)$ converge, respectively, to

$\psi(x; \mathcal{U}^3)$ and its worst-case utility function. Note that replacing $\psi(x)$ with its lower bound $\hat{\psi}(x)$ in any of the three formulations (1), (2), or (3) will always return a solution that is conservative in the sense that it is ensured to be feasible and to achieve at least the level of performance dictated by the approximate optimal value.

Remark 4. This theorem can be used to solve problems with third-order stochastic dominance constraints because third-order dominance of $h(x, \xi)$ over Z is equivalent to $\psi(x; \mathcal{U}^3) \geq 0$ with $K = 0$. The existing approach for such problems uses the fact that $\psi(x; \mathcal{U}^3) \geq 0$ is equivalent to $\mathbb{E}[\max(0, y - h(x, \xi))^2] \leq \mathbb{E}[\max(0, y - Z)^2]$ for all $y \in \mathbb{R}$ (Ogryczak and Ruszczyński 2001). Verifying this inequality at a discrete set of points is a tractable approximation. However, Theorem 3 leads to an approximation that has certain advantages: (1) it only imposes linear constraints instead of quadratic ones, (2) it provides a conservative (i.e., inner instead of an outer) approximation for the set of feasible x ensuring that dominance holds for all feasible points in the approximation, and (3) it allows us to account for additional information about the utility function.

4. Numerical Study

In this section, we use a portfolio optimization problem to illustrate the gains that can be achieved by adopting formulations that account for the preference information that is available. In this portfolio optimization problem, we assume that there are n assets, and we let x_i be the proportion of the total budget allocated to asset i. Since we do not consider short positions, the feasible set for the vector of allocations x is the convex set $\mathcal{X} := \{x \in \mathbb{R}^n : x \geq 0,\ x \cdot 1 = 1\}$. Let ξ_i be the random weekly return of asset i. Then we let the random outcome $h(x, \xi) := x \cdot \xi$ be the return of the portfolio.

We consider two formulations. First, we consider a formulation that attempts to maximize the certainty equivalent of the constructed portfolio:

$$\max_{x \in \mathcal{X}} \mathbb{C}_{\bar{u}}(x \cdot \xi), \tag{13}$$

where $\bar{u}$ is the utility function that would capture exactly the complete preference of our decision maker, the investor. When preference information is incomplete, i.e., only K pairwise comparisons have been made by the decision maker, the utility function is only known to lie in a set of type $\mathcal{U}^2$. Hence, one can either use this information to estimate the true utility function by some function $\hat{u}$ and solve problem (13) with $\hat{u}$ instead of $\bar{u}$ or solve the robust certainty equivalent formulation (3) with $h(x, \xi) = x \cdot \xi$. The latter will effectively return a portfolio that is preferred to the bank account with the largest fixed interest rate.

Alternatively, our second formulation attempts to maximize expected return of the portfolio under the constraint that this portfolio is preferred by the investor to the return of a given benchmark portfolio Z. Specifically, we are interested in solving

$$\max_{x \in \mathcal{X}} \ \mathbb{E}[x \cdot \xi] \tag{14a}$$

$$\text{s.t.}\ \ \mathbb{E}[\bar{u}(x \cdot \xi)] \geq \mathbb{E}[\bar{u}(Z)]. \tag{14b}$$

This time, in the case of incomplete preference information, although one could replace $\bar{u}$ by some estimated $\hat{u}$, we will follow the spirit of stochastic dominance, as presented in Dentcheva and Ruszczyński (2006), which suggests replacing constraint (14b) with

$$\mathbb{E}[u(x \cdot \xi)] \geq \mathbb{E}[u(Z)] \quad \forall\, u \in U_2\,.$$

Note that this approach disregards all the preference information except for the fact that the investor is risk averse. By allowing one to replace U_2 by $\mathcal{U}^2$ in problem (14), our approach corrects for this weakness.

After presenting the data used to parameterize these problems, in what follows we present empirical results that demonstrate how, in a context with incomplete preference information, decisions can improve (1) by using a worst-case analysis that accounts appropriately for this information instead of simply using an estimate $\hat{u}$ or being overly conservative through replacing $\mathcal{U}^2$ with U_2, and (2) by gathering preference information that is pertinent with respect to the nature of the decision that needs to be made. Indeed, as we ask more questions, and K increases, we expect the set of potential utilities $\mathcal{U}^2$ to shrink as our knowledge becomes better, and our portfolio performance should improve.

4.1. Data

We gathered the weekly returns of the companies in the S&P 500 index from March 30, 1993 to July 6, 2011. We focused on the 351 companies that were continuously part of the index during this period. Although not including companies that were removed from the index creates some survivorship bias, our results should remain meaningful because the absolute returns are not our focus. For each run, we randomly chose 10 companies from the pool of 351 to be our $n = 10$ assets. We considered $M = 50$ equally likely scenarios for the weekly asset returns, which we choose by randomly selecting a contiguous 50-week period of historical returns for the selected companies from the data. For the stochastic dominance formulation, the distribution of the benchmark return Z is given by the weekly return of the S&P 500 index during the same period.

4.2. Effectiveness of Robust Approach

Our first numerical study attempts to determine whether there is something to gain by accounting explicitly for available preference information in our portfolio optimization model instead of assuming more naïvely that the utility function takes on one of the popular shapes. In our simulation, the decision maker is risk averse and agrees with the axioms of expected utility, yet he is unaware which utility function captures his risk attitude. Information about this attitude can be obtained through comparison of randomly generated pairs of lotteries (using the "random utility split method" described in §4.3.1) and thus can be represented by $\mathcal{U}^2$.[3] Although he is unaware of this, the simulated decision maker, when making a comparison, acts according to the utility function $\bar{u}(y) = -20E_i(20/y) + y\exp(20/y)$, where E_i stands for the exponential integral $E_i(y) := -\int_{-x}^{\infty} \exp(-t)/t\,dt$. Our experiments consist of comparing four utility function selection strategies with respect to their average performance at maximizing the portfolio's certainty equivalent over a random sets of 10 companies and 50 scenarios, which are drawn as described in §4.1.

Remark 5. The function $\bar{u}(y) = -20E_i(20/y) + y \cdot \exp(20/y)$ was chosen because it has the property that $-u''(y)y^2/u'(y) = 20$. Hence, if the decision maker is only asked to compare lotteries that involve weekly returns close to 0%, then one might conclude that the absolute risk aversion of this decision maker is constant (i.e., his utility function takes the exponential form) when in fact his absolute risk aversion is decreasing and scales proportionally to $1/x^2$.

4.2.1. Utility Function Selection Strategies. We consider four different approaches to dealing with incomplete preference information that takes the form of a set of pairwise comparisons under the risk aversion hypothesis, i.e., $\mathcal{U}^2$.

1. *Exponential fit:* This approach simply suggests approximating $u(\cdot)$ with $\hat{u}(\cdot)$ obtained by fitting an exponential utility function of the form $u_c(y) = (1 - \exp(-cy))$ to the available information. For implementation details, we refer the reader to Appendix 6. It is interesting to note that, when a decision maker has constant absolute risk aversion, it is sufficient to identify the certainty equivalent of a single lottery to learn exactly the values that c should take. Unfortunately, here, the decision maker has decreasing risk aversion; hence, as $\mathcal{U}^2 \to \{\bar{u}(\cdot)\}$, the best-fitted function will become unable to fit $\bar{u}(\cdot)$ exactly.

2. *Piecewise linear fit:* This approach simply suggests approximating $u(\cdot)$ with $\hat{u}(\cdot)$ obtained by fitting a piecewise linear concave utility function of the form $u_{\alpha,\beta}(y) = \min_i(\alpha_i y + \beta_i)$ to the available information about the true utility function. Our implementation follows similar lines as used for the exponential utility function with the single exception that we enforce that $u_{\alpha,\beta}$ be in $\mathcal{U}^2$. The best-fitted piecewise linear utility function does have a more complex representation: for instance, in our implementation, the number of linear pieces was comparable to the size $|S|$. For implementation details, we refer the reader to Appendix G.

3. *Worst-case utility function:* This approach suggests decisions that achieve the best worst-case performance over the set of potential risk-averse utility functions. See §3.1 for implementation details.

4. *Worst-case prudent utility function:* This approach suggests decisions that achieve the best worst-case performance over the set of potential prudent utility functions. We used a discretization of 250 points to approximate the true problem as described in §3.3.

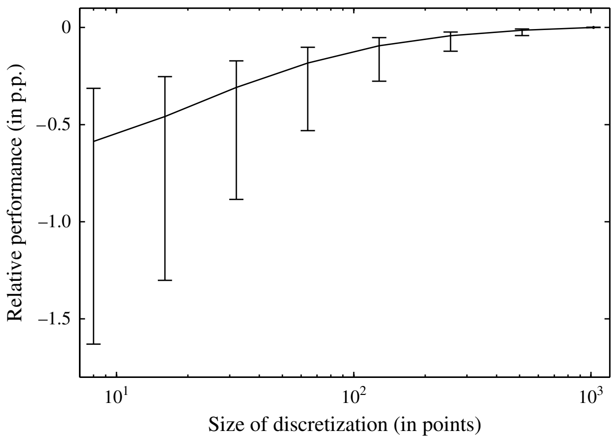
In addition, the true utility function approach plays the role of a reference for the best performance that can be achieved in each decision context. This is done by assuming that the decision maker actually knows that his preference can be represented by the form $u(y) = -20E_i(20/y) + y\exp(20/y)$. Although we argue that this situation is unlikely to occur in practice, we hope to verify that the approaches based on a piecewise linear fit or the worst-case utility functions are consistent in the sense that the decisions they suggest will actually converge, as more information is obtained about the decision maker's preferences, to the decisions that should be taken if the true utility function was known.

Remark 6. We performed a short experiment to verify that the approximation method based on discretization was accurate enough when 250 points are used. To do so, we fixed the number of lottery comparison to 40 and evaluated the effect of using a more refined discretization grid on the value of the approximated optimal worst-case certainty equivalent on 6,000 random problem instances. We observed in these experiments that when it was possible to improve the performance by more than 0.2 percentage points through a 1,000-point discretization this was nearly always already achieved using a discretization of 250 points. Figure 1 presents statistics of this convergence to the value achieved with a discretization of 1,000 points.

4.2.2. Results. Table 1 presents a comparison of the first percentiles and averages of certainty equivalents achieved in 10,000 experiments when

[3] To implement, in each simulation, we used as reference lotteries for W_0 and Y_0 the minimum and maximum return that could be achieved according to the 50 selected scenarios.

Figure 1 Statistics of Convergence of the Approximate Optimal Worst-Case Certainty Equivalent When Prudence Is Accounted for Using a Discretization Grid of Growing Size



Notes. The average, 10th, and 90th centiles of the performance relative to the performance achieved with a grid of 1,000 points are presented for a set of 6,000 experiments. p.p., percentage points.

maximizing the certainty equivalent under incomplete preference information using the four utility function selection strategies that described above. Note that the certainty equivalents, whose statistics are reported in this table, were evaluated using the true utility function. Because our simulations did not include a risk-free option, optimal portfolios had negative certainty equivalents on occasion in contexts where the 50 scenarios were taken from a period with a declining economy. An approximate method might also suggest a portfolio with negative certainty equivalent if the utility function that is used to measure performance actually overestimates its certainty equivalent.

First, we can confirm that the piecewise linear and worst-case utility function approaches suggest decisions whose respective performance converges, in terms of first percentile and average value, to the performance achieved knowing the true utility function; this is because they always employ utility functions that are members of $\mathcal{U}^2$ and because $\mathcal{U}^2 \to \{\bar{u}\}$. It is also as expected that making the false assumption that absolute risk aversion is constant, i.e., using an exponential utility function, can potentially lead to a significant loss in performance, especially when a large quantity of information about the decision maker's risk attitude has been gathered. Indeed, the results indicate that in these experiments, after 80 queries were performed, the method that used the best-fitted exponential utility function proposed portfolios that on average were equivalent to a negative guaranteed return, whereas other methods were able to suggest portfolios that on average were equivalent to a 0.06% guaranteed weekly return on investment (i.e., 3.1% annually) in terms of the decision maker's preferences. Finally, we can confirm that choosing a portfolio based on the worst-case utility function is statistically more robust, in terms of average and first percentile of the performance, when little preference information is available. It is also clear from Table 1 that accounting for information about prudence can increase the worst-case certainty equivalent by an average of 0.02 percentage points.

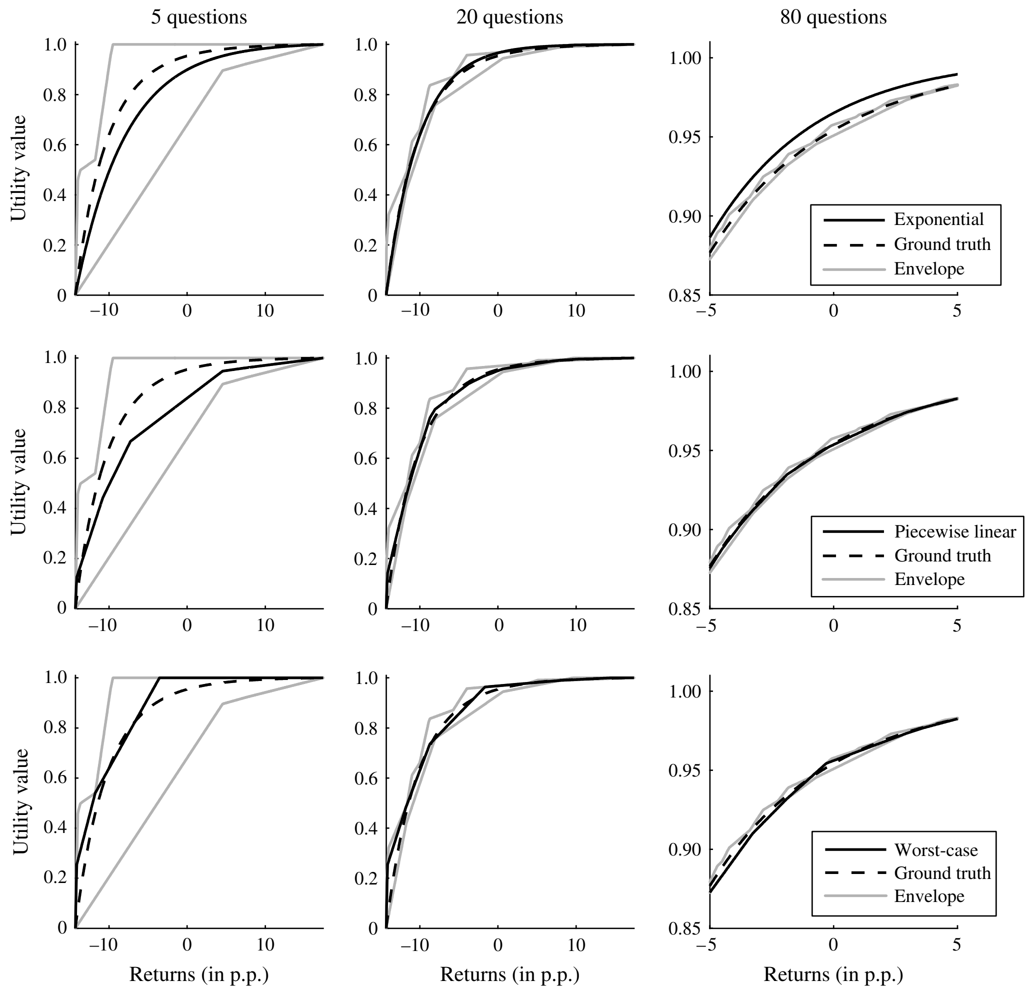
We wish to provide slightly more intuition about how the uncertainty about the utility function is reduced as more questions are answered and how the respective approaches succeed at fitting the unknown utility function. For this purpose, Figure 2 presents a set of illustrations that describe the shape of the uncertainty region together with the fitted functions as more information was obtained in one of the above experiments. In the five-questions scenario, it is clear that there is too little information to make a good choice of utility function—hence the need for a method that accounts for this ambiguity. In the 20-questions scenario, all three methods seem to provide a good estimate of the utility function. Note that although in this scenario the exponential function seems to fit the function best, we notice at a finer resolution (in the plot for 80 questions) that it will never exactly replicate the true attitude toward risk. It is harder to distinguish in these illustrations

Table 1 Comparison of the 99% Confidence Intervals of the First Percentile and Average of Certainty Equivalents Achieved in 10,000 Experiments by Maximizing the Certainty Equivalent Under Incomplete Preference Information Using Four Utility Function Selection Strategies

	Certainty equivalent (in %)					
	5 queries		20 queries		80 queries	
Approach	1st %ile	Average	1st %ile	Average	1st %ile	Average
Exponential fit	-3.8 ± 1.3	-0.05 ± 0.03	-6.0 ± 1.6	-0.12 ± 0.05	-6.1 ± 2.6	-0.13 ± 0.06
Piecewise linear fit	-8.0 ± 1.0	-0.60 ± 0.04	-3.8 ± 0.4	-0.11 ± 0.02	-3.0 ± 0.3	0.05 ± 0.02
Worst-case utility function	-2.6 ± 0.2	-0.14 ± 0.01	-2.6 ± 0.2	-0.08 ± 0.01	-2.3 ± 0.2	0.06 ± 0.01
Worst-case prudent utility function	-2.6 ± 0.2	-0.13 ± 0.01	-2.6 ± 0.2	-0.05 ± 0.01	-2.2 ± 0.2	0.08 ± 0.01
True utility function	-2.0 ± 0.2	0.12 ± 0.01	-2.0 ± 0.2	0.12 ± 0.01	-2.0 ± 0.2	0.12 ± 0.01

Notes. An experiment consists of randomly sampling a set of 10 companies as candidates for investment; a set of 50 return scenarios; and a set of 5, 20, or 80 answered queries. Also, %ile stands for percentile.

Figure 2 **Evolution of the Bounding Envelope of Utility Functions in $\mathcal{U}^2$ and of the Utility Functions Used by the Different Approaches as Observed in One Experiment for a Growing Number of Answered Questions**



Notes. The ground truth function refers to $\bar{u}(y)$. The worst-case utility function is obtained by solving the robust certainty equivalent optimization problem. Note also that for the 80-questions scenario, magnified versions of the curves are provided to highlight the irreducible fitting error of the exponential utility. Finally, p.p. stands for percentage points.

between the quality of the fitted piecewise linear utility function and the worst-case utility. Note, however, that although neither of them will ever be an exact fit (given that we can see parts of the dotted line), with the worst-case analysis approach, we can be reassured by the fact that the "misadjusted" utility function that is used is guaranteed to provide a conservative estimate of the certainty equivalent.

4.3. Effectiveness of Elicitation Strategies

The following results shed some light on how decisions might be improved by gaining more information about the preferences of the decision maker. In particular, we compare how performance is improved as we increase the number of questions the decision maker is asked using the four different elicitation strategies presented in §4.3.1. For simplicity, in our simulation, the decision maker's true utility function over the weekly return now has a constant absolute risk aversion level of 10: $\bar{u}(y) := 1 - e^{-10y}$. Note that although the decision maker is unaware that his preferences can be represented by this function, we assume that he never contradicts the conclusions suggested by such a utility function when comparing lotteries. Our experiments consist of evaluating, as the number of queries is increased, the average performance achieved by the robust approach over random sets of 10 companies and 50 scenarios, which are drawn as described in §4.1.

4.3.1. Elicitation Strategies. We elicit information about the investors' preferences by asking them to choose between the preferred two random outcomes. For simplicity, we only consider questions that compare a certain outcome to a risky gamble with two outcomes (a.k.a. the Becker-DeGroot-Marschak reference lottery; Becker et al. 1964). In other words, each query can be described by four values $r_1 \le r_2 \le r_3$ and a probability p. These four values specify the question, "Do you prefer a certain return of r_2 or a lottery where the return will be r_3 with probability p and r_1 with probability $1-p$?" If we normalize the utilities such that $u(r_1)=0$ and $u(r_3)=1$, then this query will identify whether $u(r_2)>p$ or not. We now describe three different schemes for sequentially choosing questions to ask the investor.

1. *Random utility split:* This scheme lets r_1 and r_3 be the worst and best possible returns, respectively, and chooses r_2 uniformly from $[r_1, r_3]$. The scheme then seeks to reduce by half the interval $I := \{u(r_2)\colon u \in \mathcal{U}^2\}$ of potential utility values at r_2. Thus we choose p so that $pu(r_3)+(1-p)u(r_1)$ is the midpoint of I.

2. *Random relative utility split:* This scheme differs from the previous by choosing r_1 and r_3 uniformly at random from the range of potential returns and then setting $r_2 := (r_1+r_3)/2$. Like the previous scheme, we seek to reduce by half the interval $I := \{u(r_2)\colon u \in \mathcal{U}^2\}$, and thus, we choose p so that $pu(r_3)+(1-p)u(r_1)$ is the midpoint of I.

3. *Objective-driven relative utility split:* Unlike the previous schemes, this scheme takes the optimization objective into account and seeks to improve the optimal objective value as much as possible regardless of the answer (i.e., positive or negative) to the query. To do so, it generates 10 queries using the random relative utility split scheme and for each calculates the smaller of the optimal objective value that would be reached either with a positive answer or a negative answer. It then selects among the 10 queries the query that will give the greatest improvement in the optimal objective value in the most pessimistic scenario with respect to whether the answer will be positive or negative. Mathematically speaking, in the case of the robust certainty equivalent model, this elicitation scheme will suggest the i^*th query in the list according to

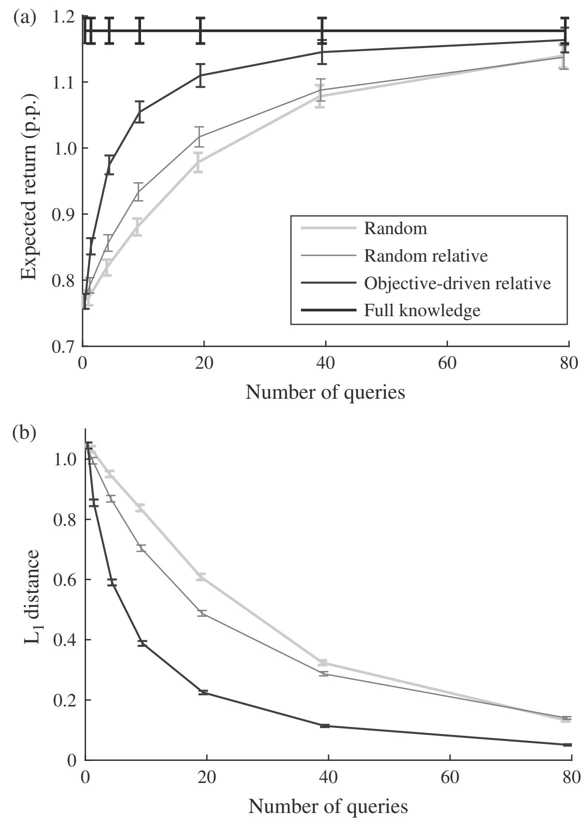
$$i^* = \underset{i\in\{1,2,\ldots,10\}}{\arg\max} \min_{\alpha\in\{-1,1\}} \max_{x\in\mathcal{X}} \min_{u\in\{u\in\mathcal{U}\mid\alpha(\mathbb{E}[u(W_i)]-\mathbb{E}[u(Y_i)])\geq 0\}} \mathbb{C}_u(x\cdot\xi),$$

where $\alpha\in\{-1,1\}$ captures the fact that the answer we might get from the investor might be that $\mathbb{E}[u(W_i)]\geq \mathbb{E}[u(Y_i)]$ or that $\mathbb{E}[u(W_i)]\leq\mathbb{E}[u(Y_i)]$.

4.3.2. Results. Whereas Figure 3 relates to the stochastic dominance formulation, Figure 4 relates to the robust certainty equivalent formulation. Panel (a) in Figures 3 and 4 shows how our objective value improves as we gain more knowledge about the investor's preferences. Panel (b) in Figures 3 and 4 focuses on the convergence of the optimal allocation.

For both formulations, we observe that the total gain between no knowledge of preferences except risk aversion and full knowledge is worth, on average, 0.4 percentage points of weekly return. We can also

Figure 3 **Effect of Increasing Numbers of Questions in a Stochastic Domination Formulation**

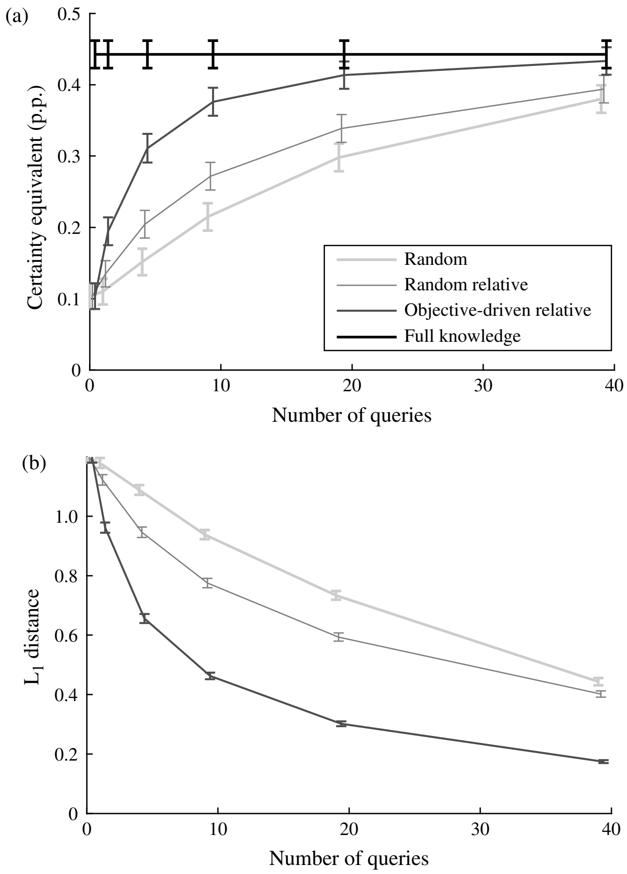


Notes. Panel (a) presents the expected return (in percentage points (p.p.)), and panel (b) presents the L_1 distance between optimal allocation with K queries and the optimal allocation with full knowledge. Shown are averages and standard errors from 1,500 simulations.

see that the improvement in performance is quick for the initial 10–20 queries. In fact, for the robust certainty equivalent formulation, four questions chosen with the objective-driven questioning scheme increase the average certainty equivalent of the weekly return by 0.2 percentage points. After these first queries, the gains from additional information decrease. This seems to indicate that there is considerable value in using all the preference information that is available, even if minimal, thus encouraging the use of our stochastic dominance formulation instead of the one presented in Dentcheva and Ruszczyński (2006), which here would achieve the performance associated to zero queries.

Finally, for both formulations, it is quite noticeable that the choice of questions to ask the decision maker also has an important impact on performance: the improvement is faster for the more sophisticated objective-driven elicitation scheme than for the simpler schemes. We believe this should justify further research on what constitutes an optimal learning strategy in this context.

Figure 4 **Effect of Increasing Numbers of Questions in a Robust Certainty Equivalent Formulation**



Notes. Panel (a) presents the certainty equivalent (in percentage points (p.p.)) of the optimal portfolio as measured with respect to the true utility function, and panel (b) presents the L_1 distance between optimal allocation with K queries and the optimal allocation with full knowledge. Shown are averages and standard errors from 500 simulations.

5. Extensions

In this section we discuss two extensions of the framework. In the first subsection we consider the case where the decision maker's preferences among the surveyed lotteries (i.e., that he prefers W_k to Y_k for all $k = 1, \ldots, K$) is inconsistent with respect to the axioms of the expected utility framework. Our proposed solution will either correct for the inconsistency by finding a consistent utility function that is closest to being able to justify the stated preferences or correct for the inconsistencies by permitting a bounded perturbation of the comparisons. The second subsection extends the framework to account for the notion of almost stochastic dominance. We have identified three different flavors of this concept and propose methods of integrating each of them. In both subsections, we argue that many of these extensions lead to only minor modifications of our framework with little loss in tractability.

5.1. Accounting for Elicitation Errors

There are many reasons why comparisons that are made by a decision maker might be inconsistent with the theory of expected utility theory. This could be because the decision maker's actual preferences do not satisfy the axioms of expected utility (such as in Allais or Ellsberg paradoxes). Alternatively, many recent studies have identified cognitive biases that can lead a decision maker to misperceive either the size of a probability or the gravity of an outcome Tversky and Kahneman (1974). In particular, the work of Kahneman and Tversky has led to an entirely new field studying behavioral decision making (see Kahneman and Tversky 1979). In view of these important issues concerning the hypotheses made by expected utility theory and of the possibility of inaccurate comparisons, our proposed approach is prescriptive in nature. Specifically, our main objective is to help decision makers that believe in the axioms of expected utility theory to identify which decision most truthfully reflects their attitude toward risk. Similar to what is done in Bertsimas and O'Hair (2013), when there is a set of preferences that does not satisfy one of the axioms, we believe the framework should identify and work with utility functions that are closest to being able to explain the incoherent preferences. Practically speaking, this means that inconsistencies can be treated as small "measurement" errors that need to be corrected for to identify how the decision maker truly wishes to act although he might be unable to express it. Note that one might want to report to the decision maker the amount of correction that needs to be applied to have coherent preferences in order to give a signal regarding whether the expected utility framework is well suited to describe his preferences.

Technically speaking, in this framework a set of comparisons (i.e., that W_k is preferred to Y_k for all $k = 1, \ldots, K$) can be identified as inconsistent when linear programs (5), (9), or (11), depending on assumptions made about the prudent or S-shaped attitude, are diagnosed as infeasible. When inconsistencies are detected or assumed (indeed, "to err is human"), we suggest accounting for "error" margins in the formulations. Below we describe three different types of errors that can easily be accounted for. Note that although we focus on the formulation presented in §3.1, similar conclusions can be drawn for the formulations of §§3.2 and 3.3.

1. If we wish to consider that noise is corrupting the expected utility evaluation at the moment when a comparison is made, then we can easily replace the condition $\mathbb{E}[u(W_k)] \geq \mathbb{E}[u(Y_k)]$ with

$$\mathbb{E}[u(Y_k)] - \mathbb{E}[u(W_k)] \leq \gamma_k, \qquad (15)$$

where $\gamma_k \geq 0$ is some positive error term (or margin) for the kth comparison. This would lead to a minor change in constraint (5d). The smallest total $\sum_k \gamma_k$ needed for the feasibility of problem (5) to hold can then be considered a measure of the size of inconsistency. This quantity, together with the description of the closest consistent comparisons, can easily be obtained by solving linear program (5) after replacing constraint (5d) with (15) and replacing the objective of (5) with $\sum_k \gamma_k$. Alternatively, one could assume a budget Γ for the total amount of inconsistency, $\sum_k \gamma_k \leq \Gamma$, and propose a solution that maximizes worst-case certainty equivalent in this context. This can easily be done by replacing constraint (5d) in problem (15) and adding $\gamma_k \geq 0, \forall k$ and

$$\sum_k \gamma_k \leq \Gamma$$

before applying duality when formulating the equivalent augmented linear constraints for $\psi(x; \mathcal{U}, Z) \geq 0$.

2. A second option assumes that the error is in the perception of the outcome values: that random variable W is perceived as $W + \delta$. If $\delta^{\min} \leq \delta \leq \delta^{\max}$ a.s., then we could replace the condition $\mathbb{E}[u(W_k)] \geq \mathbb{E}[u(Y_k)]$ with $\mathbb{E}[u(W_k + \delta^{\max})] \geq \mathbb{E}[u(Y_k + \delta^{\min})]$. In our formulations we would then need to replace the parameters W_k with $W_k + \delta^{\max}$ and Y_k by $Y_k + \delta^{\min}$, which retains the linear structure of the problem.

3. In the spirit of Bertsimas and O'Hair (2013), we could require that $1 - \epsilon$ of the K lottery comparisons hold: that the decision maker is mistaken about at most ϵK of his lottery comparisons. In that case we would introduce binary variables δ_i into (5), which would be 1 if the decision maker is mistaken about lottery i, and add the constraint $\sum_{i=1}^{K} \delta_i \leq K\epsilon$. We would then replace the condition $\mathbb{E}[u(W_k)] \geq \mathbb{E}[u(Y_k)]$ with the two constraints $\delta_i \mathcal{M} + \mathbb{E}[u(W_k)] \geq \mathbb{E}[u(Y_k)]$ and $(1 - \delta_i)\mathcal{M} + \mathbb{E}[u(Y_k)] \geq \mathbb{E}[u(W_k)]$, where $\mathcal{M}$ is a large constant ("big M"). Since this turns the calculation of $\psi(x)$ (5) into a mixed-integer linear program, solving the master problem becomes harder but potentially solvable using cutting-plane methods.

5.2. Almost Stochastic Dominance

The idea of reducing the severity of stochastic dominance constraints by assuming additional structure is not a recent one. Since the introduction of the notion of stochastic dominance, there have been a few attempts at reducing the severity of the constraint. Of course, the earliest appearance would be the idea that a higher-order stochastic dominance constraint is less restrictive. This translates as imposing the concavity/convexity of a higher-order derivative of the utility function. Three other instances are presented below. The first two are close in spirit to our framework, as they make assumptions about the utility function—that is, properties that the first and second derivative must satisfy. Unlike our framework, however, it is unclear how one might validate with the decision maker such hypotheses about derivatives and how one might perform optimization in the resulting space. The third instance is more similar in flavor to the methods that are proposed in §5.1, as it suggests inflating the set of feasible random variables by adding random variable that are "close enough" to a nondominated one.[4] Although this approach appears more tractable, nothing is known as to what type of preference axioms would suggest using this approach. For all three instances, we propose ways of extending our results to implement the proposed relaxation.

5.2.1. Meyer's Relaxation. Meyer (1977) appears to be the first mention of the idea of relaxing the stochastic dominance constraint by imposing structural properties on the utility functions in $\mathcal{U}$ that go beyond the sign of derivatives. Specifically, Meyer suggests imposing bounds on the Arrow–Pratt measure of absolute risk aversion:

$$U_{\mathrm{M}}(r_1, r_2) := \{u\colon\, r_1(x) \leq -u''(x)/u'(x) \leq r_2(x)\}.$$

He explains how to identify for a specific pair (X, Z) the worst-case utility function using dynamic programming. It is unclear, however, how one would go about optimizing when the stochastic dominance constraint involves this utility set. The following corollary sheds some light on the question.

Corollary 1. *The optimal value of the linear program* (11) *with the additional constraints*

$$r_1(\bar{y}_j)\beta_j \leq -\gamma_j \quad \forall\, j \in \{1, \ldots, N\},$$
$$r_2(\bar{y}_j)\beta_j \geq -\gamma_j \quad \forall\, j \in \{1, \ldots, N\},$$

is a lower bound for $\psi(x; \mathcal{U}^3 \cap U_{\mathrm{M}})$.

Indeed, when we account for prudence, our approximate linear program optimizes β_j and γ_j variables that play the respective role of first and second derivatives of the utility function; thus it is possible to further impose that the Arrow–Pratt measure fall in the appropriate range at the $\bar{y}_j$ locations. This gives rise, through duality theory, to conservative approximations for the robust certainty equivalent problem and the stochastic dominance problem that account for information about absolute risk aversion. Once again, as the interval of realizations is further discretized, it is expected that the approximation will converge to the true optimal value.

5.2.2. Leshno and Levy's Relaxation. To reduce the severity of stochastic dominance constraints, Leshno and Levy (2002) suggest intersecting the utility sets associated with first- or second-order stochastic dominance constraint either with

$$U_{\mathrm{LL1}}(\varepsilon) = \{u : \inf_{y'} u'(y') \leq u'(y) \leq \inf_{y'} u'(y')(1/\varepsilon - 1) \forall y \in \Re\}$$

[4] Note that this is different to what is done in §5.1 because we suggest enlarging the set of utility functions (as opposed to the set of feasible random variables).

or with

$$U_{\text{LL2}}(\varepsilon)=\{u:\inf_{y'}u''(y')\leq u''(y)\leq\inf_{y'}u''(y')(1/\varepsilon-1)\forall y\in\Re\}$$

for some $0<\varepsilon<0.5$. They motivate the use of these utility sets, arguing that most decision makers have "bounded" risk aversion. In a more recent paper, Lizyayev and Ruszczyński (2012) argue that optimizing with this form of relaxed dominance is actually intractable. We disagree in part with this statement as both $U_{\text{LL1}}(\varepsilon)$ and $U_{\text{LL2}}(\varepsilon)$ represent convex sets of utility functions. In particular, it is possible to impose both types of structural information in our framework. For instance, $\psi(x;\mathcal{U}_2\cap U_{\text{LL1}}(\varepsilon))$ can be shown to be equivalent to the optimal value of linear program (5) after simply adding the constraint

$$\beta_1\leq\beta_N(1/\varepsilon-1).$$

Similarly, $\psi(x;\mathcal{U}_3\cap U_{\text{LL2}}(\varepsilon))$ can be bounded below by the optimal value of the linear program (11) in which we add

$$\gamma_1\geq\gamma_N(1/\varepsilon-1).$$

The question of how to formulate a linear program that might evaluate $\psi(x;\mathcal{U}_2\cap U_{\text{LL2}}(\varepsilon))$ efficiently is left open.

5.2.3. Lizyayev and Ruszczyński's Relaxation. Following up on the work of Leshno and Levy (2002), Lizyayev and Ruszczyński (2012) propose a version of ε-almost stochastic dominance that is believed to be more tractable. They propose identifying X as ε-almost stochastically dominating Z if there exists a nonnegative random variable Y with $\mathbb{E}[Y]\leq\varepsilon$ such that $X+Y$ stochastically dominates Z. This is a relaxation of stochastic dominance that can easily be integrated to our framework. This is done by considering that the stochastic program with stochastic dominance constant

$$\begin{aligned}\max_{x\in\mathcal{X}}\ &\mathbb{E}[f(x,\xi)]\\ \text{s.t. }&\mathbb{E}[u(h(x,\xi))]\geq\mathbb{E}[u(Z)]\quad\forall u\in\mathcal{U}\end{aligned}$$

is replaced with

$$\max_{x\in\mathcal{X},\,y:\Omega\to\mathbb{R}^+}\ \mathbb{E}[f(x,\xi)] \tag{16a}$$

$$\text{s.t. }\mathbb{E}[u(h(x,\xi)+y(\xi))]\geq\mathbb{E}[u(Z)]\quad\forall u\in\mathcal{U}, \tag{16b}$$

$$\mathbb{E}[y(\xi)]\leq\varepsilon, \tag{16c}$$

where $y\colon\Omega\to\mathbb{R}^+$ is a random variable over which we have complete control. When ξ has a finite number of scenarios, all the results of this paper can be extended to this version of ε-almost stochastic dominance by considering that $\psi(x,\mathcal{U})$ is actually

$$\psi([x,y],\mathcal{U}):=\inf_{u\in\mathcal{U}}\big(\mathbb{E}[u(g([x,y],\xi))]-\mathbb{E}[u(Z)]\big),$$

with $g([x,y],\xi):=h(x,\xi)+y(\xi)$ being a concave function of the concatenated vector of the decision variables $[x,y]$. In particular, the ε-almost stochastic dominance version of Theorem 1 would have Equation (5a) replaced with

$$\min_{\alpha,\beta,v,w}\left\{\sum_i p_i(v_i(h(x,\xi_i)+y_i)+w_i)-\sum_j\mathbb{P}[Z=\bar{y}_j]\alpha_j\right\},$$

where $y\in\mathbb{R}_+^M$ with each y_i capturing the decision for $y(\xi_i)$. Consequently, the almost stochastic dominance constraint (16b) can be replaced by the set of constraints (8) with a simple modification to constraint (8d) so that it takes the form

$$\sum_j\bar{y}_j\mu_{i,j}\leq p_ih(x,\xi_i)+p_iy_i\quad\forall i,$$

with the μ, ν, $\lambda^{(1)}$, and $\lambda^{(2)}$ variables added as extra decision variables to the problem.

6. Conclusion

In this paper we presented tractable approaches to dealing with incomplete information about the utility function. We looked at three different formulations of our aims and three different types of utility function sets. Particularly useful is how our formulations can incorporate information such as that the decision maker would always choose lottery A over lottery B. Particularly novel are the models involving S-shaped utilities and prudent utilities. There is little work on optimizing with such utilities despite them being common in behavioral economics and finance. Also quite innovative is the robust certainty equivalent formulation. This formulation appears to be useful in situations where one currently would use a stochastic dominance constraint. Not only does the robust certainty equivalent formulation have a natural interpretation, but unlike the formulation with a stochastic dominance constraint, there is no need to define a separate objective as well. We have presented an exact tractable reformulation for risk-averse utility functions and a tractable conservative approximation for prudent functions, which in practice should be arbitrarily precise. The only exception is for S-shaped utility functions, which are already hard to optimize when the function is known. We finally believe that our proposed framework leaves space for further study of many interesting questions regarding (1) how to obtain accurate information about risk preference from the decision maker (or correct for errors appropriately), (2) how to choose the questions that will most efficiently guide the optimization to a good decision, and (3) what other characteristics of reasonable preference systems can be accounted for in a tractable way using this framework.

Acknowledgments
The authors are thankful to the referees for their constructive criticism. E. Delange gratefully acknowledges the support of the Canadian Natural Sciences and Engineering Research Council [Grant 386416-2010].

Appendix A. Quasiconcavity of the Objective in Problem (3)

For any fixed u, the certainty equivalent $\mathbb{C}_u[h(x,\xi)]$ is quasiconcave in x because $\mathbb{C}_u[h(x,\xi)] \geq t$ is equivalent to $\mathbb{E}[u(h(x,\xi))] \geq u(t)$ and the left-hand side is concave in x. Thus, the pointwise infimum $\inf_{u\in\mathcal{U}} \mathbb{C}_u[h(x,\xi)]$ is quasiconcave in x with $\inf_{u\in\mathcal{U}} \mathbb{C}_u[h(x,\xi)] \geq t$ equivalent to $\psi(x;\mathcal{U},t) \geq 0$. Hence the optimum of (3) is greater than t if and only if the constraint $\psi(\cdot;\mathcal{U},t)\geq 0$ has a feasible solution $x\in\mathcal{X}$. Thus if we can bound the optimum to an interval $[t_0,t_1]$, then we can use the bisection algorithm on t to solve the problem to precision ϵ by solving $O(\log(1/\epsilon))$ feasibility problems of the type $\psi(\cdot;\mathcal{U},t)\geq 0$. In fact, any bound on $h(x,\xi)$ gives a bound on the optimum: if $t_0 \leq h(x,\xi)\leq t_1$ for all ξ and all $x\in\mathcal{X}$, then the optimum is in $[t_0,t_1]$.

Appendix B. Proof of Theorem 2

The proof is similar to that of Theorem 1, and we start by defining

$$U(\alpha,\beta) := \{u\colon u(\bar{y}_j)=\alpha_j,\, \beta_j\in u'(\bar{y}_j)\,\forall\, j\}.$$

Then by a similar argument,

$$\psi(x;\mathcal{U}^s) = \min_{\alpha}\ \psi(x;U(\alpha,\beta)\cap U_s)$$
$$U(\alpha,\beta)\cap U_s \neq \varnothing,\ \ U(\alpha,\beta)\subseteq U_a,$$
$$U(\alpha,\beta)\subseteq U_n.$$

The constraint $U(\alpha,\beta)\cap U_s\neq\varnothing$ is represented by (9f)–(9i) and $\beta\geq 0$, $U(\alpha,\beta)\subseteq U_n$ by (9d), and $U(\alpha,\beta)\subseteq U_n$ by (9e).

Again, since $\mathbb{E}[u(Z)]$ is a constant for $u\in U(\alpha,\beta)$, evaluating $\psi(x;U(\alpha,\beta)\cap U_s)$ is equivalent to minimizing $\mathbb{E}[u(h(x,\xi))]$ over $u\in U(\alpha,\beta)\cap U_s$. Among the S-shaped functions in $U(\alpha,\beta)$, this is minimized by the piecewise linear function u^* in (10). The reason u^* is minimal is for the concave portion is the same as in the proof of Theorem 1. For the convex portion, u^* is minimal because the lines $\{(y,s)\colon s=\beta_j(y-\bar{y}_j)+\alpha_j\}$ with $j\in\mathcal{J}^-$ are supporting hyperplanes for the convex portion of any function in $U(\alpha,\beta)$. Thus, when $U(\alpha,\beta)\cap U_s\neq\varnothing$,

$$\begin{aligned}\psi(x;U(\alpha,\beta)\cap U_s) &= \mathbb{E}[u^*(h(x,\xi))1\{h(x,\xi)<0\}]\\ &\quad+\mathbb{E}[u^*(h(x,\xi))1\{h(x,\xi)\geq 0\}]\\ &\quad-\sum_j \mathbb{P}[Z=\bar{y}_j]\alpha_j.\end{aligned}$$

Since u^* is concave for $y\geq 0$, the formulation (7) can be used for that part. When $y<0$, u^* is convex, and thus,

$$\begin{aligned}u^*(y) = \min_s\ & s\\ \text{s.t.}\ & s\geq \beta_j(h(x,\xi_i)-\bar{y}_j)+\alpha_j \quad \forall\, j\in\mathcal{J}^-.\end{aligned}$$

Putting these pieces together gives us the objective (9a) and the constraints (9b), $v\geq 0$, and (9c). □

Appendix C. Worst-Case Expected Utility Formulation Biased Toward Risk Neutrality

When using the worst-case expected utility approach, if there is a large enough amount of ambiguity in the risk attitude to implement, a risk-neutral attitude is used to select the decision. Take, for instance, the case of $\mathcal{U}^2$, where no information is known about the risk attitude (i.e., $K=0$ so that no lotteries have been compared). If the normalization scheme is such that $u(a)=0$ and $u(b)=1$, then one can show that the linear utility $u^*(y)=(y-a)/(b-a)$ is a worst-case utility for evaluating any lottery Z supported on the interval $[a,b]$. When comparing the application of the robust certainty equivalent formulation (3) to this context, one can easily demonstrate that the infimum is achieved as ϵ goes to 0 in $u_Z^\epsilon(y)=\min(1;(1-\epsilon)(y-a)/(\inf_\xi Z(\xi)-a))$ such that $\lim_{\epsilon\searrow 0}\mathbb{C}_{u_Z^\epsilon}(Z)=\inf_\xi Z(\xi)$ under this alternative formulation, thus implementing extreme risk aversion. Based on this example, one might argue that the use of "worst-case expected utility" might be misleading because it actually does not return cautious decisions in situations where it could.

In fact, when maximizing the worst-case expected utility, one should be aware of the following connection with the robust certainty equivalent formulation. It is actually the case that both methods are special cases of an approach that seeks the optimal solution to

$$\text{maximize}_{x\in\mathcal{X}}\ \sup\{v\in\mathcal{V}\colon\ \mathbb{E}[u(Z_v)]\leq\mathbb{E}[u(h(x,\xi))]\quad\forall\, u\in\mathcal{U}\},$$

where $\{Z_v\}_{v\in\mathcal{V}}$, for some $\mathcal{V}\subseteq\mathbb{R}$, is an indexed set of lotteries parameterized by v such that $v'>v$ implies that $Z_{v'}\succ Z_v$. In other words, this more general formulation suggests selecting the decision for which the random variable $h(x,\xi)$ is known to be preferred to the most attractive lottery in $\{Z_v\}_{v\in\mathcal{V}}$. When Z_v is simply the certain amount v, this reduces to maximizing the robust certainty equivalent formulation. Perhaps less trivial is the fact that when $\mathcal{U}$ is normalized so that all utility functions return $u(a)=0$ and $u(b)=1$, then choosing $\{Z_v\}_{v\in\mathcal{V}}$ to be the lotteries that return b with probability v and return a with probability $1-v$ makes this general approach reduce to the worst-case expected utility formulation. More formally, we have that

$$\begin{aligned}&\sup\{v\in[0,1]\colon\ \mathbb{E}[u(Z_v)]\leq\mathbb{E}[u(h(x,\xi))]\,\forall\, u\in\mathcal{U}\}\\ &=\sup\{v\in[0,1]\colon\ vu(b)+(1-v)u(a)\leq\mathbb{E}[u(h(x,\xi))]\,\forall\, u\in\mathcal{U}\}\\ &=\sup\{v\in[0,1]\colon\ v\leq\mathbb{E}[u(h(x,\xi))]\,\forall\, u\in\mathcal{U}\}\\ &=\inf_{u\in\mathcal{U}}\mathbb{E}[u(h(x,\xi))].\end{aligned}$$

It is because of this somewhat arbitrary choice of the set $\{Z_v\}_{v\in\mathcal{V}}$, which is implicitly used by the worst-case expected utility formulation to map and compare $h(x, \xi)$, that one ends up obtaining decisions that are biased in a potentially unappealing way.

Appendix D. Proof of Theorem 3

As in the proof of Theorem 1, we define $U(\alpha) := \{u\colon u(\bar{y}_j) = \alpha_j \ \forall j\}$. Then by an argument similar to one used in the proof of Theorem 1,

$$\psi(x; \mathcal{U}^3) = \min_{\alpha} \ \psi(x; U(\alpha) \cap U_2 \cap U_3)$$
$$U(\alpha) \cap U_2 \cap U_3 \neq \varnothing, U(\alpha) \subseteq U_a, U(\alpha) \subseteq U_n.$$

The constraint $U(\alpha) \subseteq U_n$ is represented by (11c) and $U(\alpha) \subseteq U_n$ by (11d). Since we only seek a lower bound, we can represent $U(\alpha) \cap U_2 \cap U_3 \neq \varnothing$ by the constraints (11e)–(11h), $\beta \geq 0$, and $\gamma \leq 0$.

Again, since $\mathbb{E}[u(Z)]$ is a constant for $u \in U(\alpha)$, evaluating $\psi(x; U(\alpha) \cap U_2 \cap U_3)$ is equivalent to minimizing $\mathbb{E}[u(h(x, \xi))]$ over $u \in U(\alpha) \cap U_2 \cap U_3$. Among the utilities in $U(\alpha) \cap U_2$ (thus giving a lower bound), this is minimized by the function $\hat{u}^*$ in (12), as in the proof of Theorem 1. Thus, we seek to minimize $\mathbb{E}[\hat{u}^*(h(x, \xi))] - \sum_j \mathbb{P}[Z = \bar{y}_j]\alpha_j$. Since $\hat{u}^*$ is concave, we use the formulation (7), which then gives us the objective (11a) and the constraints (11b) and $v \geq 0$. □

Appendix E. Prudence Implies the Existence and Convexity of $u'(\cdot)$

Based on the definition of prudence as put forth by Eeckhoudt and Schlesinger (2006), we can conclude that for a prudent decision maker,

$$w \geq v \ \ \Rightarrow \ \ E[u(w+Z)] - u(w) \geq E[u(v+Z)] - u(v),$$

for any pair $(v, w) \in \mathbb{R}^2$ and for any random variable Z with zero mean. Here, we will first demonstrate that if a decision maker is prudent, then the derivative of $u(\cdot)$ must exist on its domain. We follow with a proof that $u'(\cdot)$ is convex.

Proposition 1. *If a decision maker is prudent and risk averse, then the utility function that captures his attitude with respect to risk must be differentiable everywhere in the interior of its domain.*

It is a well-known fact that risk aversion implies that the utility function is monotonic and concave. It must therefore be differentiable almost everywhere and semidifferentiable everywhere. Let us assume that at w_0 in the interior of the domain, the utility function is not differentiable. Since it is semidifferentiable at w_0, we must have that

$$\lim_{\varepsilon \searrow 0} \frac{u(w_0+\varepsilon) - u(w_0)}{\varepsilon} = u'_+(w_0)$$

exists and is strictly smaller than

$$\lim_{\varepsilon \searrow 0} \frac{u(w_0) - u(w_0-\varepsilon)}{\varepsilon} = u'_-(w_0)$$

by concavity. Furthermore, since the utility function is differentiable almost everywhere, there must also exist a value $v_0 < w_0$ where the utility function is differentiable. Hence we have that if Z_ε is a random variable that puts half of the weight on ε and half on $-\varepsilon$, then

$$\lim_{\varepsilon \searrow 0} (1/\varepsilon)(E[u(v_0+Z_\varepsilon)] - u(v_0))$$
$$= \lim_{\varepsilon \searrow 0} (1/\varepsilon)(0.5u(v_0-\varepsilon) + 0.5u(v_0+\varepsilon) - u(v_0)) = 0.$$

Yet, since $v_0 \leq w_0$, by the prudence hypothesis we must also have that for any $\varepsilon > 0$,

$$E[u(v_0+Z_\varepsilon)] - u(v_0) \leq E[u(w_0+Z_\varepsilon)] - u(w_0).$$

By dividing both sides of the inequality by ε and taking the limit as ε goes to zero, we get

$$\begin{aligned}
&\lim_{\varepsilon \searrow 0} (1/\varepsilon)(E[u(v_0+Z_\varepsilon)] - u(v_0)) \\
&\quad \leq \lim_{\varepsilon \searrow 0} (1/\varepsilon)(E[u(w_0+Z_\varepsilon)] - u(w_0)) \\
&\quad = \lim_{\varepsilon \searrow 0} (1/\varepsilon)(0.5u(w_0+\varepsilon) + 0.5u(w_0-\varepsilon) - u(w_0)) \\
&\quad = 0.5u'_+(w_0) - 0.5u'_-(w_0) \\
&\quad < 0\,.
\end{aligned}$$

This establishes that the expression $\lim_{\varepsilon \searrow 0}(1/\varepsilon)(E[u(v_0+Z_\varepsilon)] - u(v_0))$ is both strictly smaller than zero and equal to zero, which is a contradiction. □

Note that this proof excludes the boundaries of the domain. We are left with proving that the first derivative of the utility function is convex.

Let w be any value in the interior of the domain of $u(\cdot)$, let $\varepsilon > 0$, and let Z be any zero-mean random variable supported on two points in the domain of $u(\cdot)$. We have from Eeckhoudt and Schlesinger (2006) their definition of prudence that

$$E[u(w+\varepsilon+Z)] - u(w+\varepsilon) \geq E[u(w+Z)] - u(w)$$

and therefore that

$$(1/\varepsilon)(E[u(w+\varepsilon+Z)] - u(w+\varepsilon)) \geq (1/\varepsilon)(E[u(w+Z)] - u(w)).$$

Taking the limit of the difference between the left and right sides, we get

$$\lim_{\varepsilon \searrow 0} (1/\varepsilon)(E[u(w+\varepsilon+Z)] - E[u(w+Z)] - u(w+\varepsilon) + u(w)) \geq 0$$

so that

$$E[u'(w+Z)] - u'(w) \geq 0.$$

The last inequality can be shown to be equivalent to the definition of convexity. □

Appendix F. Fitting an Exponential Utility Function to $\mathcal{U}^2$

To fit a utility function, common practice typically suggests fixing the utility value at two reference points $u(\bar{y}_0) = 0$ and $u(w_0) = 1$ and using queries to locate the relative utility values achieved at a set of returns $u_j \approx u(\bar{y}_j) \ \forall j = 1, 2, \ldots, J$. The "best-fitted" function is then the one that maximizes the following mean square error problem:

$$\begin{aligned}
\min_{a,b,c} \ & \sum_{j=1}^{J} (a(1 - \exp(-c\bar{y}_j)) + b - u_j)^2 \\
\text{s.t.} \ & a(1 - \exp(-c\bar{y}_0)) = 0 \text{ and } a(1 - \exp(-cw_0)) = 1, \\
& a \geq 0, \ c \geq 0.
\end{aligned}$$

Although nonconvex, this problem is typically considered computationally feasible since it reduces to a search over the single parameter c. We adapt this procedure to the context where the preference information takes the shape of $\mathcal{U}^2$. Specifically, without loss of generality, we first let Y_0 and W_0 be certain lotteries and fixed $\mathbb{E}[u(Y_0)] = 0$ and $\mathbb{E}[u(W_0)] = 1$. Next, for a set of $\{\bar{y}_j\}_{j=1}^J$, we can use the information in $\mathcal{U}^2$ to evaluate a range of possible utility values at each $\bar{y}_j$. We let u_j take on the midvalue of this interval, $u_j := (\min_{u\in\mathcal{U}^2} u(\bar{y}_j) + \max_{u\in\mathcal{U}^2} u(\bar{y}_j))/2$, hence capturing the fact that we wish the exponential utility function pass as close as possible to the center of the intervals in which we know the function should pass. We solve the same mean square error problem to select our best-fitted exponential utility function $\hat{u}(y)$. Note that this approach reduces to the method described above when $\mathcal{U}^2 = \{u(\cdot) | u(\bar{y}_j) = u_j\}$. For computational reasons, our implementation used the set $\{\bar{y}_j\}_{j=0}^J := \bigcup_{k=0}^K (\text{supp}(Y_k) \cup \text{supp}(W_k))$, which uniformly spanned the range of possible returns.

Appendix G. Fitting a Piecewise Linear Utility Function to $\mathcal{U}^2$

To fit a piecewise linear utility function, we follow a similar procedure as for fitting an exponential function. Namely, for a set of $\{\bar{y}_j\}_{j=1}^J$ that includes $\bar{y}_0$ and w_0, after considering that $u(\bar{y}_0) = 0$ and $u(w_0) = 1$, we can use the information in $\mathcal{U}^2$ to evaluate a range of possible utility values at each $\bar{y}_j$. We let u_j take on the midvalue of this interval, $u_j := (\min_{u\in\mathcal{U}^2} u(\bar{y}_j) + \max_{u\in\mathcal{U}^2} u(\bar{y}_j))/2$, hence capturing the fact that we wish the utility function pass as close as possible to the center of the intervals in which we know the function should pass. Based on the discretization $\{\bar{y}_j\}_{j=1}^J$, we parameterize the piecewise linear function using the value and supergradient at each point in the set. We are left with solving the following optimization problem:

$$\begin{aligned}
\min_{\alpha,\beta} \quad & \sum_{j=1}^J (\alpha_j - u_j)^2 \\
\text{s.t.} \quad & \alpha_{j+1} - \alpha_j \geq \beta_{j+1}(\bar{y}_{j+1} - \bar{y}_j) \quad \forall j, \\
& \alpha_{j+1} - \alpha_j \leq \beta_j(\bar{y}_{j+1} - \bar{y}_j) \quad \forall j, \\
& \alpha_j - \alpha_{j-1} \leq \beta_j(\bar{y}_j - \bar{y}_{j-1}) \quad \forall j, \\
& \alpha_j - \alpha_{j-1} \geq \beta_{j-1}(\bar{y}_j - \bar{y}_{j-1}) \quad \forall j, \\
& \alpha_{j(\bar{y}_0)} = 0, \;\; \alpha_{j(w_0)} = 1, \;\; \beta \geq 0,
\end{aligned}$$

where $j(\bar{y}_0)$ and $j(w_0)$ are the respective indexes of the $\bar{y}_0$ and w_0 terms in the set $\{\bar{y}_j\}_{j=0}^J$. Again, for computational reasons, our implementation used the set $\{\bar{y}_j\}_{j=0}^J := \bigcup_{k=0}^K (\text{supp}(Y_k) \cup \text{supp}(W_k))$, which uniformly spanned the range of possible returns.

References

Arrow KJ (1959) Rational choice functions and orderings. *Economica* 26(102):121–127.

Aumann RJ (1962) Utility theory without the completeness axiom. *Econometrica* 30(3):445–462.

Becker GM, DeGroot MH, Marschak J (1964) Measuring utility by a single-response sequential method. *Behavioral Sci.* 9(3):226–232.

Ben-Tal A, Nemirovski A (1998) Robust convex optimization. *Math. Oper. Res.* 23(4):769–805.

Ben-Tal A, Teboulle M (2007) An old-new concept of convex risk measures: The optimized certainty equivalent. *Math. Finance* 17(3):449–476.

Bertsimas D, O'Hair A (2013) Learning preferences under noise and loss aversion: An optimization approach. *Oper. Res.* 61(5): 1190–1199.

Boutilier C, Patrascu R, Poupart P, Schuurmans D (2006) Constraint-based optimization and utility elicitation using the minimax decision criterion. *Artificial Intelligence* 170(8–9): 686–713.

Chajewska U, Koller D, Parr R (2000) Making rational decisions using adaptive utility elicitation. *Proc. 17th Natl. Conf. Artificial Intelligence* (American Association for Artificial Intelligence, Palo Alto, CA), 363–369.

Clemen RT, Reilly T (2001) *Making Hard Decisions with DecisionTools Suite*, 2nd ed. (Duxbury, Pacific Grove, CA).

Conn AR, Scheinberg K, Vicente LN (2009) *Introduction to Derivative-Free Optimization* (Society for Industrial and Applied Mathematics, Philadelphia).

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):595–612.

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

Dentcheva D, Ruszczyński A (2006) Portfolio optimization with stochastic dominance constraints. *J. Banking Finance* 30(2): 433–451.

Dubra J, Maccheroni F, Ok E (2004) Expected utility theory without the completeness axiom. *J. Econom. Theory* 115(1):118–133.

Eeckhoudt L, Schlesinger H (2006) Putting risk in its proper place. *Amer. Econom. Rev.* 96(1):280–289.

Grable J, Lytton RH (1999) Financial risk tolerance revisited: The development of a risk assessment instrument. *Financial Services Rev.* 8(3):163–181.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291.

Leland HE (1968) Saving and uncertainty: The precautionary demand for saving. *Quart. J. Econom.* 82(3):465–473.

Leshno M, Levy H (2002) Preferred by "all" and preferred by "most" decision makers: Almost stochastic dominance. *Management Sci.* 48(8):1074–1085.

Lizyayev A, Ruszczyński A (2012) Tractable almost stochastic dominance. *Eur. J. Oper. Res.* 218(2):448–455.

Meyer J (1977) Choice among distributions. *J. Econom. Theory* 14(2): 326–336.

Ogryczak W, Ruszczyński A (2001) On stochastic dominance and mean-semideviation models. *Math. Programming* 89(2): 217–232.

Rawls J (1971) *A Theory of Justice* (Harvard University Press, Cambridge, MA).

Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131.

Von Neumann J, Morgenstern O (1944) *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, NJ).