OPERATIONS RESEARCH
Vol. 58, No. 4, Part 2 of 2, July–August 2010, pp. 1220–1234
ISSN 0030-364X | EISSN 1526-5463 | 10 | 5804 | 1220

DOI 10.1287/opre.1100.0821

A Soft Robust Model for Optimization Under Ambiguity

Aharon Ben-Tal
William Davidson Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology, Technion City, Haifa 32000, Israel, abental@ie.technion.ac.il

Dimitris Bertsimas
Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, dbertsim@mit.edu

David B. Brown
The Fuqua School of Business, Duke University, Durham, North Carolina 27708, dbbrown@duke.edu

In this paper, we propose a framework for robust optimization that relaxes the standard notion of robustness by allowing the decision maker to vary the protection level in a smooth way across the uncertainty set. We apply our approach to the problem of maximizing the expected value of a payoff function when the underlying distribution is ambiguous and therefore robustness is relevant. Our primary objective is to develop this framework and relate it to the standard notion of robustness, which deals with only a single guarantee across one uncertainty set. First, we show that our approach connects closely to the theory of convex risk measures. We show that the complexity of this approach is equivalent to that of solving a small number of standard robust problems. We then investigate the conservatism benefits and downside probability guarantees implied by this approach and compare to the standard robust approach. Finally, we illustrate the methodology on an asset allocation example consisting of historical market data over a 25-year investment horizon and find in every case we explore that relaxing standard robustness with soft robustness yields a seemingly favorable risk-return trade-off: each case results in a higher out-of-sample expected return for a relatively minor degradation of out-of-sample downside performance.

Subject classifications: robust optimization; ambiguity; convex risk measures; divergence measures; optimized certainty equivalent; portfolio optimization.
Area of review: Special Issue on Computational Economics.
History: Received March 2008; revision received September 2009; accepted December 2009. Published online in *Articles in Advance* June 3, 2010.

1. Introduction

There are many different approaches to modeling uncertainty in the context of optimization problems. Different criteria, ranging from tractability issues to practical appropriateness to modeling concerns, motivate these disparate approaches.

A typical critique of stochastic approaches is that probability distributions are often unknown in practice and can be computationally unwieldy. The robust optimization paradigm (advanced by Ben-Tal and Nemirovski 1998, El Ghaoui et al. 1998, and Bertsimas and Sim 2004, among others) instead relies on deterministic sets and boasts favorable tractability properties for many classes of optimization problems. Specifically, when attempting to maximize a payoff function $f(\mathbf{x}) := f(\mathbf{x}, \mathbf{u})$, where $\mathbf{x}$ are decision variables and $\mathbf{u}$ are uncertain parameters, the robust optimization framework focuses on subproblems of the form

$$\inf_{u\in\mathcal{U}} f(\mathbf{x}, \mathbf{u}). \tag{1}$$

On the other hand, in many applications, particularly financial ones, it is natural to have at least some probabilistic description of the world. In particular, $\mathbf{u}$ can be modeled as random and having a distribution that belongs to some set of probability measures $\mathcal{P}$. This allows for a very natural merging of the robust and the stochastic approaches via functions of the form

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[f(\mathbf{x}, \mathbf{u})]. \tag{2}$$

With (2), we are allowing for a stochastic model of the world; but rather than pinning down the distribution exactly, we provide some measure of robustness against distributional variation. Specifically, we allow for *ambiguity* described by the set of measures $\mathcal{P}$.

While the approach above can be motivated from a modeling standpoint, there are two clear and related critiques. First, how should one obtain a set of distributions $\mathcal{P}$? Second, why does one want to treat distributions within $\mathcal{P}$ equally (while entirely ignoring those outside $\mathcal{P}$!)? For instance, in an asset management setting, one can imagine f is the uncertain return of a portfolio $\mathbf{x}$ driven by, say, a factor model. Problem (2) might be too limiting to capture preferences adequately in such a situation. For instance, in scenarios in which the market crashes, we would likely desire guarantees that are inherently different than those in situations in which the market performs well. More generally, we might to have assurances on performance that vary in a smooth way with the performance of the market as a whole.

To this end, we propose a more general version of (2) that allows for a notion of robustness providing different guarantees for different subsets within $\mathcal{P}$. We call this approach the *soft robust approach*. In addition to providing a richer modeling framework, this approach preserves favorable convexity properties of the nominal problem. Finally, the structure of the robust model can be guided via the use of *convex risk measures* (Föllmer and Schied 2002), which we show to be closely connected.

The concept of ambiguity has been looked at from many angles in the economics literature; we provide only a few of the many references on the subject. The etymology of research on ambiguity is generally traced back to Ellsberg (1961), although there are certainly earlier discussions (e.g., Knight 1921). Schmeidler (1989) formally defines a notion of ambiguity aversion via nonadditive probability measures. Gilboa and Schmeidler (1989) extend classical utility theory to allow for multiple priors in a min-max way; Epstein and Wang (1994) provide a related model for asset pricing. More recently, Klibanoff et al. (2005) axiomatize a model of "smooth ambiguity-averse" preferences that involve the use of a second-order probability distribution. Hansen and Sargent (2001) have linked ideas from robust control to the problem of model misspecification for economic agents: they work with a model based on *multiplier preferences*, which is represented by a relative entropy penalty that arises as an important, special case of our framework as well. Maccheroni et al. (2006) derive a so-called *variational preferences* framework for handling both risk and ambiguity. Their resulting framework has similarities to ours from a modeling standpoint, but their focus is on the preferences that induce these types of functionals in the decision-making of economic agents. Rigotti et al. (2008) study a general model for accommodating risk and ambiguity based on convex preferences that includes several of the aforementioned preferences as special cases. Our focus here is not on preferences; rather, we will take the soft robust approach as a starting point, link it to the theory of convex risk measures, and then focus on its relationship to standard robustness in terms of problem complexity, conservativeness of solutions, and downside performance guarantees in the context of optimization under ambiguity.

There is a growing body of research in the optimization and robust optimization literature geared toward handling ambiguity. Calafiore and El Ghaoui (2006) consider linear optimization problems with chance constraints in which the underlying distribution is known to belong only within a given set. Erdogan and Iyengar (2006) develop a robust sampled version of ambiguous chance-constrained problems that is feasible with high probability. Chen et al. (2007) propose a tractable means of approximating distributionally robust optimization problems using deviation measures. Delage and Ye (2010) provide a polynomial-time algorithm for sample-driven robust stochastic programs with uncertainty in the mean and covariance.

Although our framework is more general, we explore the methodology with a particular aim toward asset allocation problems. Again, there are many references on robust portfolio optimization. Goldfarb and Iyengar (2003) propose a robust approach based on uncertainties in the parameters of an underlying factor model. El Ghaoui et al. (2003) provide tractable optimization models for value-at-risk constraints under uncertain first- and second-order moment information. More recently, Lim et al. (2010) develop a robust approach based on a relative benchmark objective that results in less pessimistic allocations. The use of convex risk measures in this setting has been explored as well (e.g., Lüthi and Doege 2005, Ben-Tal and Teboulle 1991 under a special class of convex risk measures known as optimized certainty equivalents; these risk measures are recently used by Natarajan et al. 2010 in a distribution-robust model applied to portfolio optimization as well). For asset allocation, our focus is on the ambiguity implications of such risk measures and the trade-off between expected return and downside performance that one can obtain by relaxing the traditional model of set-based robustness over probability distributions.

After completing this paper, we became aware of the work of Fischetti and Monaci (2008), who develop an approach called "light robustness" to cope with the issue of overly conservative solutions in robust optimization. This approach is somewhat different than ours. First, their main focus is on uncertainty in the parameters of LPs, whereas we are looking at ambiguity. In addition, the light robust approach places a hard upper bound on the objective value (to reduce conservatism) and then minimizes the degree of infeasibility, measured by a weighted sum of slack variables, with a fixed uncertainty set. In contrast, we do not impose an explicit optimality guarantee or a single uncertainty set, but instead allow the feasibility guarantees to vary across a family of uncertainty sets. Thus, in the light robust approach, optimality is enforced in hard fashion, and the level of feasibility is an output of the formulation; in the soft robust approach, it is the opposite: varying levels of feasibility are enforced up front, and the level of optimality is ultimately an output of the formulation.

The structure of the paper is as follows. In §2, we first state the problem and provide some of the necessary background on convex risk measures. We then introduce the soft robust approach and show a dual relationship between soft robustness and a representation with convex risk measures. The latter parts of this section deal with the relationship of the soft robust and standard robust approaches: first, in 2.4, we discuss complexity of the soft robust approach and show that, from an algorithmic standpoint, it can be dealt with by solving a sequence of standard robust problems; then, in 2.5, we compare the conservatism of the two approaches. Section 3 provides bounds on downside performance of both approaches under ambiguous distributions. Finally, §4 illustrates the approach on two applications: first, optimization of a portfolio of risky bonds with ambiguous default probabilities; then, asset allocation with re-balancing using historical market data. The appendix contains proofs of some of the results and results related to the asset allocation application.

2. The Soft Robust Approach

In this section, we motivate the problem in question and provide some background on convex risk measures. We then show our central result, which is the connection between the soft robust approach and convex risk measures. Finally, we compare the soft robust and standard robust approaches in terms of complexity and conservatism.

2.1. Preliminaries

Consider the problem of maximizing an uncertain payoff function subject to constraints. Let $(\Omega, \mathcal{F})$ be an underlying measure space and let $X \subseteq \mathbb{R}^n$ be a set representing feasible choices for a decision vector $\mathbf{x}$. Let $\mathbf{u}$: $\Omega \to \mathbb{R}^m$ be a random variable that represents the uncertain parameters in the problem and let f: $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a payoff function, depending on both our decision $\mathbf{x} \in \mathbb{R}^n$ and the realization $\mathbf{u}(\omega)$ of the uncertain parameter. Assume further that $f(\mathbf{x}, \mathbf{u}(\omega))$ is concave on $\mathbf{x} \in X$ for all $\omega \in \Omega$ and that X is a nonempty, closed, convex set. These assumptions ensure that the nominal problem (i.e., the problem without uncertainty for fixed $\mathbf{u}$) is a convex optimization problem and therefore not hopeless to solve by itself.

We will restrict ourselves to the case when the random variable $f(\mathbf{x}, \mathbf{u})$ is bounded for every $\mathbf{x} \in X$. In most practical applications, we can always find reasonable bounds on our payoffs. For instance, in asset allocation, stock prices cannot fall below zero and are not likely to increase by more than several standard deviations above their mean. We also note that for the most part, the explicit dependency of f on the random parameter $\mathbf{u}$ is unnecessary, and we will simply write the uncertain payoff as $f(\mathbf{x})$, with the understanding that $f(\mathbf{x})$ is a random variable induced by $\mathbf{u}$.

Our objective is to maximize $f(\mathbf{x})$ in expectation. As one motivating example, consider the problem when $\mathbf{x}$ is a wealth allocation vector, $\mathbf{u}$ is a random return vector on the underlying assets, and $f(\mathbf{x}, \mathbf{u}) = U(\mathbf{u}'\mathbf{x})$ is the payoff function, where U: $\mathbb{R} \to \mathbb{R}$ is a utility function representing the investor's risk preferences.

The nominal problem of interest, then, is $\max\{\mathbb{E}_{\mathbb{P}}[f(\mathbf{x})]: \mathbf{x} \in X\}$. Notice, however, that this problem requires detailed (if not exact) knowledge of an underlying distribution $\mathbb{P}$. In practice, such a distribution is typically inferred from prior beliefs, an estimation procedure, or some mixture of the two.[1] Regardless, $\mathbb{P}$ is generally subject to *ambiguity* and, ideally, we would like to ensure some level of robustness against this ambiguity in the distribution.

A standard way to accomplish this is to instead solve the problem

$$\begin{aligned} &\underset{x \in X}{\text{maximize}} && \mathbb{E}_{\mathbb{P}}[f(\mathbf{x})], \\ &\text{subject to} && \mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})] \geqslant 0 \quad \forall \mathbb{Q} \in \mathcal{Q}, \end{aligned} \tag{3}$$

where $\mathcal{Q}$ is a set of probability measures on $(\Omega, \mathcal{F})$. This problem maximizes the expected payoff under the reference distribution, subject to the requirement that the expected payoff be nonnegative[2] provided the "true" distribution lies within the set $\mathcal{Q}$. Intuitively, the larger the set $\mathcal{Q}$ is, the more robust this approach will be (but, of course, the worse its performance level will be under $\mathbb{P}$).

A drawback of this approach is that it does not afford much flexibility in trading off robustness for performance under $\mathbb{P}$. In particular, if a solution above is deemed too conservative, the natural way to correct this is to reduce $\mathcal{Q}$ to a smaller set $\tilde{\mathcal{Q}}$. Doing this, however, discards entirely any guarantees on the performance under measures in $\mathcal{Q} \backslash \tilde{\mathcal{Q}}$. Related to this is the fact that (3) does not distinguish between measures within $\mathcal{Q}$; it enforces the same requirement for any measures in this set, even ones that are somehow very extreme.

Instead, we propose an approach that allows us to provide guarantees that are ever weakening for probability measures that are more extreme. In particular, we can generalize (3) to the problem

$$\begin{aligned} &\underset{x \in X}{\text{maximize}} && \mathbb{E}_{\mathbb{P}}[f(\mathbf{x})], \\ &\text{subject to} && \mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})] \geqslant -\epsilon \quad \forall \mathbb{Q} \in \mathcal{Q}(\epsilon),\ \epsilon \geqslant 0, \end{aligned} \tag{4}$$

where $\{\mathcal{Q}(\epsilon) \subseteq \mathcal{Q}\}_{\epsilon \geqslant 0}$ is a family of sets of measures that is nondecreasing on $\epsilon \geqslant 0$ (i.e., $\mathcal{Q}(\epsilon_1) \subseteq \mathcal{Q}(\epsilon_2)$ for all $\epsilon_2 \geqslant \epsilon_1 \geqslant 0$). Instead of forcing $f(\mathbf{x})$ to be nonnegative in expectation for all measures in the set $\mathcal{Q}$, (4) imposes a weaker requirement. Namely, this problem indexes probability measures within $\mathcal{Q}$ according to some nonnegative parameter ϵ and then requires, for any measure in $\mathcal{Q}(\epsilon)$, that the expectation be bounded below by the relaxed value of $-\epsilon$. Roughly speaking, one can visualize the parameter ϵ as being proportional to some notion of distance between probability measures: as ϵ grows, the set $\mathcal{Q}(\epsilon)$ increases, but the demand placed by the right-hand side of the constraint in (4) decreases as well.

Notice that (3) is just a special case of (4) with the assignment $\mathcal{Q}(\epsilon) = \mathcal{Q}$ for all $\epsilon \geqslant 0$. Therefore, (4), which we will call the *soft robust approach*, is more general than (3), which we will call the *standard robust approach*. The formulation of soft robustness and its relationship to standard robustness will be a key theme of this paper.

2.2. Convex Risk Measures

It turns out that problem (4), which we have motivated purely from ambiguity and robustness considerations, is related to the theory of *convex risk measures*, which we now briefly describe. For more detailed background, see Föllmer and Schied (2004).

Consider a set of scenarios Ω, and let $\mathcal{Y}$ be a linear space of bounded functions representing the payoffs of various "positions" as a function of the set of outcomes Ω. For any $Y_1, Y_2 \in \mathcal{Y}$, the shorthand notation $Y_1 \geqslant Y_2$ denotes $Y_1(\omega) \geqslant Y_2(\omega)\ \forall\, \omega \in \Omega$. We define the following.

Definition 2.1. A function $\rho\colon \mathcal{Y} \to \mathbb{R}$ is a *risk measure* if it satisfies the following, for all $Y_1,\ Y_2 \in \mathcal{Y}$:
1. If $Y_1 \geqslant Y_2$, then $\rho(Y_1) \leqslant \rho(Y_2)$.
2. If $c \in \mathbb{R}$, then $\rho(Y_1 + c) = \rho(Y_1) - c$.

The interpretation of a risk measure is as a capital requirement: $\rho(Y)$ represents the sure amount by which a position needs to be augmented in order to make it acceptable. Put another way, a position is acceptable if and only if $\rho(Y) \leqslant 0$, which means we require no additional payment to be willing to accept the position.

In particular, it is easy to see that $\rho(Y + \rho(Y)) = 0$, so by adding $\rho(Y)$ to position Y, the position is now deemed acceptable. This is the intuition behind the translation invariance property. The monotonicity property simply imposes the reasonable requirement that positions that dominate other positions should also have less risk associated with them.

For our purposes, we will focus on the following special class of risk measures.

Definition 2.2. A risk measure $\rho\colon \mathcal{Y} \to \mathbb{R}$ is *convex* if it satisfies, for all $Y_1, Y_2 \in \mathcal{Y}$:

$$\rho(\lambda Y_1 + (1-\lambda) Y_2) \leqslant \lambda \rho(Y_1) + (1-\lambda)\rho(Y_2) \quad \forall\, \lambda \in [0, 1].$$

Convexity simply makes explicit the idea that diversification of positions cannot increase risk. Put another way, convex risk measures encourage allocation into "bundled" positions as opposed to extreme ones.

Now consider the case when we have a measure space $(\Omega, \mathcal{F})$, as is the case in our development above. Let $\mathcal{P}$ be the set of all probability measures on $(\Omega, \mathcal{F})$. Föllmer and Schied (2002) show, under some technical conditions, that convex risk measures on bounded random variables Y are representable as

$$\rho(Y) = \sup_{\mathbb{Q} \in \mathcal{P}} \{-\mathbb{E}_{\mathbb{Q}}[Y] - \alpha(\mathbb{Q})\},$$

where $\alpha\colon \mathcal{P} \to \mathbb{R}$ is a convex function. We say that ρ is *generated* by the function α. Notice that such a representation does not rely on specification of any reference distribution $\mathbb{P}$. Typically, we will assume the presence of a reference distribution, but it is not explicitly necessary for our general approach. This type of representation will allow us to connect soft robustness to convex risk measures.

In what follows, we will focus on convex risk measures with the following property.

Definition 2.3. We say a convex risk measure $\rho\colon \mathcal{Y} \to \mathbb{R}$ is *normalized* if the corresponding α is nonnegative for all $\mathbb{Q} \in \mathcal{P}$ and there exists a $\mathbb{Q} \in \mathcal{P}$ such that $\alpha(\mathbb{Q}) = 0$.

Notice that a normalized convex risk measure has the natural property that $\rho(0) = 0$.

2.3. Soft Robustness and Convex Risk Measures

Now let us return to the problem at hand, which is protecting ourselves, in a relaxed way, against ambiguity in the distribution of the payoff function f. At one extreme, we have the approach that ignores ambiguity entirely and only focuses on expectation over a single measure $\mathbb{P}$. This is potentially dangerous, because solutions could be very sensitive to the choice of $\mathbb{P}$. At the other extreme, we have the robust approach above, which protects equally against *all* measures in some set of measures $\mathcal{Q} \subseteq \mathcal{P}$, which can be too conservative. The soft robust approach (4) lies between these two extremes.

Using the representation above, we now formalize soft robustness and link it explicitly to convex risk measures. In everything that follows, we will denote the family of sets of probability measures by $\mathbf{Q}$, i.e.,

$$\mathbf{Q} = \{\mathcal{Q}(\epsilon) \subseteq \mathcal{P}\colon \epsilon \geqslant 0\},$$

where $\mathcal{Q}(\epsilon)$ is nondecreasing and convex[3] and $\mathcal{Q}(0)$ is nonempty. With this shorthand, we define the soft robust feasible set as

$$X^{Q}(\delta) \triangleq \Big\{\mathbf{x} \in X\colon \inf_{\mathbb{Q} \in \mathcal{Q}(\epsilon)} \mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})] \geqslant -\epsilon\ \forall\, \epsilon \in [0, \delta]\Big\} \qquad (5)$$

at level $\delta \geqslant 0$. We can now connect this to convex risk measures.

Theorem 2.1. *For any $\delta \geqslant 0$,*

$$X^{Q}(\delta) = \left\{\mathbf{x} \in X\colon \min_{\lambda \geqslant 0}\left\{\lambda\delta + (\lambda+1)\rho^{Q}\left(\frac{f(\mathbf{x})}{\lambda+1}\right)\right\} \leqslant 0\right\}, \qquad (6)$$

where ρ^{Q} is the normalized, convex risk measure generated by the penalty function α^{Q}:

$$\alpha^{Q}(\mathbb{Q}) = \inf\{\epsilon \geqslant 0\colon \mathbb{Q} \in \mathcal{Q}(\epsilon)\}.$$

Conversely, if ρ is any normalized, convex risk measure generated by a penalty function α, then for any $\delta \geqslant 0$, the equivalence (6) *holds with ρ on the right and the convex, nondecreasing family of measures $\mathbf{Q}^{\alpha} = \{\mathcal{Q}^{\alpha}(\epsilon) \subseteq \mathcal{P}\colon \epsilon \geqslant 0\}$ on the left, where*

$$\mathcal{Q}^{\alpha}(\epsilon) = \big\{\mathbb{Q} \in \mathcal{P}\colon \alpha(\mathbb{Q}) \leqslant \epsilon\big\},$$

and $\mathcal{Q}^{\alpha}(0)$ is nonempty.

Proof. First consider the case when $\delta > 0$. We have

$$\begin{aligned}
&\inf\{\mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})]+\epsilon\colon \epsilon\in[0,\delta],\ \mathbb{Q}\in\mathscr{Q}(\epsilon)\}\\
&\quad=\inf\{\mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})]+\alpha^{Q}(\mathbb{Q})\colon \mathbb{Q}\in\mathscr{P},\ \alpha^{Q}(\mathbb{Q})\leqslant\delta\}\\
&\quad=\max_{\lambda\geqslant 0}\Big\{-\lambda\delta+\inf_{\mathbb{Q}\in\mathscr{P}}\{\mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})]+(\lambda+1)\alpha^{Q}(\mathbb{Q})\}\Big\}\\
&\quad=\max_{\lambda\geqslant 0}\Big\{-\lambda\delta-(\lambda+1)\sup_{\mathbb{Q}\in\mathscr{P}}\Big\{-\mathbb{E}_{\mathbb{Q}}\Big[\frac{f(\mathbf{x})}{\lambda+1}\Big]-\alpha^{Q}(\mathbb{Q})\Big\}\Big\}\\
&\quad=\max_{\lambda\geqslant 0}\Big\{-\lambda\delta-(\lambda+1)\rho^{Q}\Big(\frac{f(\mathbf{x})}{\lambda+1}\Big)\Big\}\\
&\quad=-\min_{\lambda\geqslant 0}\Big\{\lambda\delta+(\lambda+1)\rho^{Q}\Big(\frac{f(\mathbf{x})}{\lambda+1}\Big)\Big\},
\end{aligned}$$

where in the first line we are using the definition of α^Q. The second equality follows by the fact that $\delta > 0$ and $\mathscr{Q}(0)$ is nonempty, so the Slater condition holds; moreover, the inf is finite because by assumption $f(\mathbf{x})$ is bounded for all $\mathbf{x} \in X$. Therefore, by invoking standard convex duality results (e.g., Luenberger 1969, Chapter 8), strong duality holds, which justifies the second equality. The other equalities follow by simple algebra and the definition of ρ^Q. This shows the equivalence when $\delta > 0$.

For the case when $\delta = 0$, we note that for any normalized convex risk measure ρ and any bounded random variable Y,

$$\min_{\lambda\geqslant 0}\Big\{(\lambda+1)\rho\Big(\frac{Y}{\lambda+1}\Big)\Big\}=\lim_{\lambda\to\infty}(\lambda+1)\rho\Big(\frac{Y}{\lambda+1}\Big).$$

To see this, note that for any $w \geqslant 1$, we have $Y/w = (1/w)Y + (1 - 1/w)0$; convexity and normalization of ρ then imply $\rho(Y/w) \leqslant (1/w)\rho(Y) + (1 - 1/w)\rho(0) = (1/w)\rho(Y)$, i.e., $w\rho(Y/w) \leqslant \rho(Y)$. Noting again that $f(\mathbf{x})$ is bounded for all $\mathbf{x} \in X$, we have

$$\begin{aligned}
&\lim_{\lambda\to\infty}(\lambda+1)\rho^{Q}\Big(\frac{f(\mathbf{x})}{\lambda+1}\Big)\\
&\quad=\lim_{\lambda\to\infty}\sup_{\mathbb{Q}\in\mathscr{P}}\{-\mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})]-(\lambda+1)\alpha^{Q}(\mathbb{Q})\}\\
&\quad=\sup_{\mathbb{Q}\in\mathscr{Q}(0)}\{-\mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})]\}.
\end{aligned}$$

This is clearly equivalent to the soft robust form for $\delta = 0$.

To see that ρ^Q is convex, it suffices to note that convexity of the family Q implies convexity of α^Q. Finally, because α^Q is by construction nonnegative and $\mathscr{Q}(0)$ is nonempty, there exists a $\mathbb{Q} \in \mathscr{P}$ such that $\alpha^Q(\mathbb{Q}) = 0$, i.e., ρ^Q is normalized.

For the converse, assume ρ, generated by α, is given. Clearly, since α is convex and ρ is normalized, we have $\mathbf{Q}^\alpha$ as a nondecreasing, convex family and $\mathscr{Q}^\alpha(0)$ nonempty. Having already established (6) for any $\delta \geqslant 0$ and any nondecreasing, convex family $\mathbf{Q}$ with $\mathscr{Q}(0)$ nonempty, we need only show that $\alpha^{Q^\alpha} = \alpha$, and the result will follow. We have

$$\begin{aligned}
\alpha^{Q^\alpha}(\mathbb{Q})&=\inf\{\epsilon\geqslant 0\colon \mathbb{Q}\in\mathscr{Q}^\alpha(\epsilon)\}\\
&=\inf\{\epsilon\geqslant 0\colon \mathbb{Q}\in\{\widehat{\mathbb{Q}}\in\mathscr{P}\colon \alpha(\widehat{\mathbb{Q}})\leqslant\epsilon\}\}\\
&=\alpha(\mathbb{Q}),
\end{aligned}$$

which completes the proof. □

Theorem 2.1 explicitly connects soft robustness and convex risk measures. It states that we can always view the soft robust approach as imposing an appropriately defined form of a convex risk measure, and vice versa. This provides some kind of economic interpretation to the soft robust approach: indeed, it is easy to verify that for any convex risk measure ρ and any $\delta \geqslant 0$, the function

$$\mu(Y)=\inf_{\lambda\geqslant 0}\Big\{\lambda\delta+(\lambda+1)\rho\Big(\frac{Y}{\lambda+1}\Big)\Big\}$$

is itself a convex risk measure. The result above states we can generate the corresponding convex risk measure ρ^Q in the representation according to the penalty function α^Q.

Conversely, given any normalized, convex risk measure generated by α, the representation given by (6) corresponds to a soft robust formulation with the family of sets $\mathbf{Q}^\alpha$ given as above.

In optimization terms, for $\delta > 0$, we have $\mathbf{x} \in X^Q(\delta)$ if and only if there exists a $\lambda \geqslant 0$ such that

$$\lambda\delta+(\lambda+1)\rho\Big(\frac{f(\mathbf{x})}{\lambda+1}\Big)\leqslant 0.$$

The function $(\lambda+1)\rho(f(\mathbf{x})/(\lambda+1))$ is jointly convex on $(\lambda, \mathbf{x})$, with $\lambda \geqslant 0$, so this is always a convex constraint.

To make the discussion more concrete, we now provide a few examples.

Example 2.1 (Standard Robustness and Coherent Risk Measures). Fix some set $\mathscr{Q} = \mathscr{Q}(\delta) \subseteq \mathscr{P}$ and let $\mathscr{Q}(\epsilon) = \mathscr{Q}$ for all $\epsilon \in [0, \delta]$. In this case, $\mathbf{x} \in X^Q(\delta)$ is equivalent to

$$\inf_{\mathbb{Q}\in\mathscr{Q}}\mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})]\geqslant 0,$$

which is equivalent to a *coherent* risk measure (Artzner et al. 1999) on $f(\mathbf{x})$. Coherent risk measures are convex risk measures with the additional property that $\rho(\lambda Y) = \lambda\rho(Y)$ for all $Y \in \mathscr{Y}$, $\lambda \geqslant 0$.

This assures that we "break even" in expectation for all probability measures $\mathbb{Q} \in \mathscr{Q}$ and have no guarantees outside this set; this is the standard notion of robustness. As above, $\mathscr{Q}$ can be thought of as ambiguity representing our uncertainty in the underlying measure $\mathbb{P}$. Following similar steps to those above, one can show under mild conditions (see Proposition 2.2) that the feasible set can alternatively be expressed in dual fashion as

$$X(\delta)=\Big\{\mathbf{x}\in X\colon \inf_{\lambda>0}\Big\{\lambda\delta+\lambda\rho^{Q}\Big(\frac{f(\mathbf{x})}{\lambda}\Big)\Big\}\leqslant 0\Big\} \tag{7}$$

for the same convex risk measure ρ^Q as in Theorem 2.1. From here on, we will refer to the formulation (7) as the *standard robust* approach at level $\delta \geqslant 0$. The connection between coherent risk measures and robust optimization have been explored recently in more detail by Bertsimas and Brown (2009) and Natarajan et al. (2009).

As another example[4] for a particular structure of f, consider the special case when we can decompose uncertainty as follows: $f(\mathbf{x}) = g(\mathbf{x}) + h$, for some random variable h, independent of $\mathbf{x}$, supported on $[0, \delta]$. Letting $\mathcal{Q}(\epsilon) = \{\mathbb{Q} \in \mathcal{Q}\colon \mathbb{E}_{\mathbb{Q}}[h] \leqslant \epsilon\}$, and noting then that $\mathcal{Q}(\delta) = \mathcal{Q}$, it is not hard to see that

$$\inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})] \geqslant 0 \iff \inf_{\mathbb{Q}\in\mathcal{Q}(\epsilon)} \mathbb{E}_{\mathbb{Q}}[g(\mathbf{x})] \geqslant -\epsilon \quad \forall \epsilon \in [0, \delta],$$

i.e., standard robustness of a function plus bounded, additive noise is equivalent to soft robustness of the function with the family $\{\mathcal{Q}(\epsilon)\}_{\epsilon\in[0,\delta]}$.

Example 2.2 (Divergence Measures and Optimized Certainty Equivalents). Fix a reference measure $\mathbb{P}$ and let $\mathcal{P}$ represent all measures absolutely continuous with respect to $\mathbb{P}$. Let $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+$ be a convex function that attains a minimum value of 0 at 1. The penalty

$$\alpha(\mathbb{Q}) = \mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right]$$

is known as the *ϕ-divergence* from $\mathbb{Q}$ to $\mathbb{P}$. It is shown by Ben-Tal and Teboulle (2007) that the induced convex risk measure can be written as

$$\rho(Y) = \inf_{\nu\in\mathbb{R}} \{\nu + \mathbb{E}_{\mathbb{P}}[\phi^*(-Y - \nu)]\},$$

where ϕ^* is the conjugate of ϕ, i.e.,

$$\phi^*(y) = \sup_{w\in\mathbb{R}_+} (yw - \phi(w)).$$

We refer to this risk measure as the *optimized certainty equivalent* (OCE) under ϕ. Because ϕ^* by definition is convex, computation of ρ is itself a convex optimization problem over a single parameter ν. This computation can be interpreted as minimizing the net present value of an uncertain debt to be paid over two periods, with ϕ^* playing the role of a loss function (negative of a concave utility function) that captures the present value of tomorrow's (uncertain) payment.

The robust sets have the form

$$\mathcal{Q}(\epsilon) = \left\{\mathbb{Q} \in \mathcal{P}\colon \mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] \leqslant \epsilon\right\},$$

and the problem can be interpreted as enforcing weaker constraints at measures with greater ϕ-divergence from $\mathbb{P}$.

For instance, consider the family of functions

$$\phi_\pi(w) = \frac{1}{\pi} w + \frac{1}{1-\pi} - \frac{1}{\pi(1-\pi)} w^{1-\pi},$$

parameterized by $\pi \in (0, 1]$. The corresponding conjugate function is given by

$$\phi^*_\pi(y) = \begin{cases} \dfrac{1}{1-\pi}[(1-\pi y)^{(\pi-1)/\pi} - 1] & \text{if } y < 1/\pi, \\ +\infty & \text{otherwise.} \end{cases}$$

For $\pi = 1/2$, the associated divergence function is $\phi_{1/2}(w) = 2(\sqrt{w} - 1)^2$, which is the divergence function generating the *Hellinger distance*

$$\sqrt{\tfrac{1}{2}\int \left(\sqrt{d\mathbb{Q}} - \sqrt{d\mathbb{P}}\right)^2}$$

between measures $\mathbb{Q}$ and $\mathbb{P}$. In the limiting case when $\pi = 0$, we recover the relative entropy divergence; indeed, we have

$$\begin{aligned} \lim_{\pi\to 0} \phi^*_\pi(y) &= \lim_{\pi\to 0} \frac{1}{1-\pi}[(1-\pi y)^{(\pi-1)/\pi} - 1] \\ &= \lim_{\pi\to 0} (1-\pi y)^{-1/\pi} - 1 \\ &= e^y - 1, \end{aligned}$$

which corresponds to $\phi(w) = w \log w - w + 1$. By scaling this by a factor $1/\gamma \geqslant 0$, we obtain

$$\alpha(\mathbb{Q}) = \frac{1}{\gamma}\mathbb{P}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] = \frac{1}{\gamma}\int_{\omega\in\Omega} \log\left(\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega)\right) d\mathbb{Q}(\omega),$$

i.e., α is the relative entropy from $\mathbb{Q}$ to the reference measure $\mathbb{P}$ scaled by γ. In this case, using the duality representation for OCEs, we obtain $\rho_\gamma(Y) = (1/\gamma)\log \mathbb{E}_{\mathbb{P}}[e^{-\gamma Y}]$. Following Föllmer and Schied (2004), we refer to this as the *entropic risk measure at level* $\gamma \geqslant 0$.

We can use relative entropy as the basis of a soft robust constraint. For instance, suppose we wish to protect against all distributions $\mathbb{Q} \in \mathcal{P}$ contained within δ-relative entropy of $\mathbb{P}$ in a soft way. Then we have $\mathbf{x} \in X^Q(\delta)$ if and only if

$$\begin{aligned} &\min_{\lambda\geqslant 0}\{\lambda\delta + (\lambda+1)\rho_\gamma(f(\mathbf{x})/(\lambda+1))\} \leqslant 0 \\ &\iff \min_{\lambda\geqslant 0}\left\{\lambda\delta + \frac{(\lambda+1)}{\gamma}\log \mathbb{E}_{\mathbb{P}}[e^{-\gamma f(\mathbf{x})/(\lambda+1)}]\right\} \leqslant 0 \\ &\iff \min_{\lambda\geqslant 0}\{\lambda\delta + \rho_{\gamma/(\lambda+1)}(f(\mathbf{x}))\} \leqslant 0, \end{aligned}$$

where ρ_γ denotes the entropic risk measure with parameter γ. This means $\mathbf{x} \in X^Q(\delta)$ if and only if there exists a $\lambda \geqslant 0$ such that

$$\mathbb{E}_{\mathbb{P}}[e^{-\gamma f(\mathbf{x})/(\lambda+1)}] \leqslant e^{-\lambda\gamma\delta/(\lambda+1)}.$$

When f is linear in $\mathbf{x}$, for fixed $\lambda \geqslant 0$, this is a constraint of a geometric program. One solution approach here would be to solve a sequence (bisecting over $\lambda \geqslant 0$) of such geometric programs to obtain a soft robust solution.

Another OCE of widespread interest is the *conditional value-at-risk* (*CVaR*) risk measure (see, for instance, Rockafellar and Uryasev 2000). This family of risk measures is indexed by a parameter $\beta \in (0, 1]$ and defined as

$$\text{CVaR}_\beta(Y) = \inf_{\nu\in\mathbb{R}}\left\{\nu + \frac{1}{\beta}\mathbb{E}_{\mathbb{P}}[\max(-Y - \nu, 0)]\right\}.$$

CVaR is an OCE with the function

$$\phi^*(y) = \frac{1}{\beta}\max(y, 0).$$

Because it is also a coherent risk measure, it is described by a single set of distributions, given in this case by $\mathcal{Q}(1/\beta)$, where

$$\mathcal{Q}(\epsilon) \triangleq \left\{\mathbb{Q} \in \mathcal{P}\colon \frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) \leqslant \epsilon \;\forall\, \omega \in \Omega\right\}.$$

OCEs seem useful for soft robustness because they encompass a wide range of convex risk measures and provide a natural ambiguity interpretation in terms of divergences between measures.

Example 2.3 (Comprehensive Robustness). As a variation of the theme, consider the function

$$\alpha(\mathbb{Q}) = \begin{cases} 0, & \text{if } \mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] \leqslant \delta \\ \mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] - \delta & \text{otherwise,} \end{cases}$$

for some $\delta > 0$ and where ϕ satisfies the properties from the previous example. This has the interpretation of enforcing break-even performance in expectation for all probability measures within δ of $\mathbb{P}$ according to the divergence measure generated by ϕ, as well as enforcing performance guarantees outside this set that weakens as the ϕ-divergence between $\mathbb{P}$ and $\mathbb{Q}$ grows. We can express the corresponding risk measure ρ_C compactly in terms of OCEs as follows:

$$\begin{aligned}\rho_C(Y) &= \sup_{\mathbb{Q}\in\mathcal{P}}\{-\mathbb{E}_{\mathbb{Q}}[Y] - \alpha(\mathbb{Q})\} \\ &= \sup_{\mathbb{Q}\in\mathcal{P}}\left\{-\mathbb{E}_{\mathbb{Q}}[Y] - \max\left(0, \mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] - \delta\right)\right\} \\ &= \sup_{\mathbb{Q}\in\mathcal{P},\, t\in\mathbb{R}}\left\{t - \mathbb{E}_{\mathbb{Q}}[Y]\colon t \leqslant 0,\, t \leqslant \delta - \mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right]\right\} \\ &= \min_{\lambda_1,\lambda_2\in\mathbb{R}_+^2}\sup_{\mathbb{Q}\in\mathcal{P},\, t\in\mathbb{R}}\left\{-\mathbb{E}_{\mathbb{Q}}[Y] + (1-\lambda_1-\lambda_2)t \right. \\ &\qquad\qquad \left. - \lambda_2\mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] + \lambda_2\delta\right\}.\end{aligned}$$

Clearly, for the supremum to be finite we must have $\lambda_1 + \lambda_2 = 1$, so it suffices to consider $\lambda_1 \in [0, 1]$. If the corresponding optimal λ_1 satisfies $\lambda_1 = 0$, then $\rho_C(Y) = -\min_{\omega\in\Omega} Y(\omega)$ (see the proof of Proposition 2.2). Otherwise, we have $\lambda_1 > 0$ and thus

$$\begin{aligned}\rho_C(Y) &= \min_{\lambda\in[0,1]}\sup_{\mathbb{Q}\in\mathcal{P}}\left\{-\mathbb{E}_{\mathbb{Q}}[Y] - \lambda\mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] + \lambda\delta\right\} \\ &= \inf_{\lambda\in(0,1]}\sup_{\mathbb{Q}\in\mathcal{P}}\left\{-\mathbb{E}_{\mathbb{Q}}[Y] - \lambda\mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] + \lambda\delta\right\} \\ &= \inf_{\lambda\in(0,1]}\lambda\sup_{\mathbb{Q}\in\mathcal{P}}\left\{-\mathbb{E}_{\mathbb{Q}}[Y/\lambda] - \mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] + \delta\right\} \\ &= \inf_{\lambda\in(0,1]}\{\lambda\rho(Y/\lambda) + \lambda\delta\},\end{aligned}$$

where ρ is the OCE generated by ϕ. Ben-Tal et al. (2006) explore a similar type of robust approach, although their methodology does not deal with probability distributions or ambiguity.

2.4. Complexity of the Soft Robust Optimization Problem

By definition, a soft robust problem contains an *infinite number* of standard robust constraints (one for every $\epsilon \in [0, \delta]$ and each corresponding set $\mathcal{Q}(\epsilon)$). One might therefore wonder what the computational expense is of imposing an infinite number of robust constraints. We show in this section that one can optimize in a soft robust way by solving a small sequence of standard robust optimization problems.

Indeed, consider the problem of testing membership within the soft robust set $X^{Q}(\delta)$. With the equivalence of separation and optimization in mind, we can do this efficiently if and only if we can also efficiently compute

$$\min_{\mathbf{x}\in X}\max_{\epsilon\in[0,\delta]}\left\{-\inf_{\mathbb{Q}\in\mathcal{Q}(\epsilon)}\mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})] - \epsilon\right\}.$$

This problem deals with an infinite number of standard robust problems (at all levels $\epsilon \in [0, \delta]$). We have the following.

Proposition 2.1. *We have*

$$\min_{\mathbf{x}\in X}\max_{\epsilon\in[0,\delta]}\left\{-\inf_{\mathbb{Q}\in\mathcal{Q}(\epsilon)}\mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})] - \epsilon\right\} = \max_{\epsilon\in[0,\delta]}\theta(\epsilon),$$

where

$$\theta(\epsilon) \triangleq \min_{\mathbf{x}\in X}\left\{-\inf_{\mathbb{Q}\in\mathcal{Q}(\epsilon)}\mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})] - \epsilon\right\}.$$

Moreover, θ is a concave function on $\epsilon \in [0, \delta]$.

Proof. See the appendix. □

The upshot of Proposition 2.1 is that the complexity of optimizing over a soft robust constraint is comparable to that of optimizing over a standard robust constraint. In particular, concavity of $\theta(\epsilon)$ means we can bisect on $\epsilon \in [0, \delta]$, at each step solving a standard robust optimization problem of the form

$$\min_{\mathbf{x}\in X}\left\{-\inf_{\mathbb{Q}\in\mathcal{Q}(\epsilon)}\mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})]\right\}. \tag{8}$$

Therefore, provided we have an oracle for solving the standard robust problem, we can solve the soft robust problem by solving a small number of standard robust problems. For more on the calculus of optimization of functions of the form $\rho(f(\mathbf{x}))$ when ρ is a coherent risk measure, see Ruszczyński and Shapiro (2006).

It is also worth noting that even if one does not solve a soft robust problem as a sequence of standard robust problems (indeed, in some cases, such as in the risky bond portfolio example in §4.1, we will simply formulate the soft robust problem as a single, convex problem), Proposition 2.1 also provides a simple way of interpreting the soft robust solution. In particular, if $\theta(\epsilon)$ is strictly concave, then there will be a single level $\epsilon^\star \in [0, \delta]$ at which the robustness constraint is active at the optimal solution, i.e.,

$$\inf_{\mathbb{Q}\in\mathcal{Q}(\epsilon^\star)} \mathbb{E}_{\mathbb{Q}}[f(\mathbf{x}^\star)] = -\epsilon^\star.$$

Thus, in this case, the soft robust approach is equivalent to a standard robust approach using the set $\mathcal{Q}(\epsilon^\star)$ at the level $-\epsilon^\star$. Of course, $\epsilon^\star$ is not known beforehand; it is only found after solving the soft robust problem.

If the underlying event space is large, naturally even the standard robust problem might be difficult to solve. In such situations, one may approximate the problem either with samples or with some type of convex approximation. (There are many papers on this topic; for example, in the case of partial moment knowledge and OCEs based on piecewise linear utility functions, see Natarajan et al. 2010.) In the applications we consider, we will either have a manageable event space (the risky bond example) or approximate an unknown distribution via samples (the asset allocation example).

2.5. Conservatism Benefits of Soft Robustness

One of the central motivations for soft robustness is to provide solutions that are less conservative than those generated by standard robust formulations. In this section, we provide one characterization of this phenomenon.

Consider a family of sets of distributions, $\mathbf{Q}$, as before, and fix a $\delta \geqslant 0$. Denote the standard robust feasible set by

$$X(\delta) \triangleq \left\{\mathbf{x}\in X\colon \inf_{\mathbb{Q}\in\mathcal{Q}(\delta)} \mathbb{E}_{\mathbb{Q}}[f(\mathbf{x})] \geqslant 0\right\}. \tag{9}$$

If we write our objective function as $r(\mathbf{x})$, we are interested in comparing problems of the form

$$v(\delta) \triangleq \max\{r(\mathbf{x})\colon \mathbf{x}\in X(\delta)\}$$

to soft robust problems of the form

$$v^S(\delta) \triangleq \max\{r(\mathbf{x})\colon \mathbf{x}\in X^Q(\delta)\}.$$

Clearly, for any fixed $\delta \geqslant 0$, $X(\delta) \subseteq X^Q(\delta)$, so we have $v^S(\delta) \geqslant v(\delta)$, as the soft robust approach imposes identical constraints at less restrictive levels. In this sense, the soft robust approach is clearly less conservative than the standard robust approach. One way to quantify the degree of this conservatism is to fix the size parameter δ, then ask how much larger a region the soft robust approach can "cover" without compromising objective value. We now make this explicit.

Proposition 2.2. *Let $\mathbf{x}_\delta$ be an optimal solution to the standard robust problem for some $\delta > 0$ and let $\lambda^\star$ be a minimizer of the function $\lambda\rho^Q(f(\mathbf{x}_\delta)/\lambda) + \lambda\delta$ over $\lambda \geqslant 0$. If $\lambda^\star > 1$, then $v^S(\delta') \geqslant v(\delta)$ for all $\delta' \leqslant [\lambda^\star/(\lambda^\star - 1)]\delta$. Otherwise, $v^S(\delta') \geqslant v(\delta)$ for all $\delta' \geqslant 0$.*

Proof. See the appendix. □

Proposition 2.2 provides a post-solution mechanism for understanding one view on how conservative a soft robust solution will be relative to a standard robust solution. With $\mathbf{x}_\delta$ as the optimal robust solution at some level $\delta > 0$, we have

$$\inf_{\mathbb{Q}\in\mathcal{Q}(\delta)} \mathbb{E}_{\mathbb{Q}}[f(\mathbf{x}_\delta)] \geqslant 0.$$

Such a solution must also be soft robust feasible at some level $\delta' \geqslant \delta$, i.e., for some $\delta' \geqslant \delta$, we must have

$$\inf_{\mathbb{Q}\in\mathcal{Q}(\epsilon)} \mathbb{E}_{\mathbb{Q}}[f(\mathbf{x}_\delta)] \geqslant -\epsilon \;\forall\, \epsilon \in [0, \delta'].$$

Proposition 2.2 says we can find such a δ' by looking at the optimal $\lambda^\star$ associated with the dual, risk measure representation of the standard robust solution. To interpret this, note that $\lambda^\star$ varies inversely with δ; thus, large $\lambda^\star$ will be associated with relatively small values of δ for which robustness is a fairly weak requirement. In such cases, the solutions will be soft robust but only up to the level stated in the result. $\lambda^\star$ small, on the other hand, correspond to relatively high values of δ for which robustness is a very stringent requirement. Here, the robustness condition is so strict that such solutions are automatically soft robust at any level $\delta' \geqslant 0$. The level $\lambda^\star = 1$ is the threshold level between these two types of behaviors for the robust solutions.

We note that Proposition 2.2 provides a generic result relating the standard robust and soft robust approaches, but it does not provide any explicit bounds on the difference in optimality. Such statements are certainly of interest (for instance, we might want to know the difference in optimality between a solution that ignores ambiguity (i.e., $\delta = 0$) and the soft robust approach at some $\delta > 0$) but to make any useful statements, one would likely need to have more specific knowledge about the problem structure.

3. Probability Guarantees for Robust and Soft Robust Solutions

Proposition 2.2 provides one analysis of how much less conservative the soft robust approach can be compared to the standard robust approach. Having less conservative solutions is a benefit of the soft robust approach. This benefit comes at the price of being less robust. From a managerial perspective (particularly in financial settings, such as asset management), one often desires an understanding of the likelihood of a particular "downside" performance. Computing probability guarantees that hold under ambiguity for both the soft robust and standard robust approaches is the subject of this section.

We would like to understand what membership in $X^Q(\delta)$ or $X(\delta)$ implies about downside performance of the solutions. In particular, and with slight abuse of notation, we let $\mathcal{P}(X^Q(\delta))$ (resp., $\mathcal{P}(X(\delta))$) be the set of admissible $\mathbb{P} \subseteq \mathcal{P}$ such that $\mathbf{x} \in X^Q(\delta)$ (resp., $\mathbf{x} \in X(\delta)$). Then we would like to understand the implications

$$\mathbf{x} \in X^Q(\delta) \implies \sup_{\mathbb{P} \in \mathcal{P}(X^Q(\delta))} \mathbb{P}\{f(\mathbf{x}) \leqslant -z\},$$

$$\mathbf{x} \in X(\delta) \implies \sup_{\mathbb{P} \in \mathcal{P}(X(\delta))} \mathbb{P}\{f(\mathbf{x}) \leqslant -z\}$$

for some downside performance level z. Notice that this will be useful in situations for which we can enforce soft robustness (or standard robustness), but lack an analytical description of the full reference distribution. It is not hard to conceive of such situations. As an example, let Y be a random variable that is the sum of m independent random variables $Y = Y_1 + \cdots + Y_m$, and let ϕ be relative entropy. Noting the dual connection to the entropic risk measure, enforcing soft robustness requires constraints involving terms related to

$$\log \mathbb{E}_{\mathbb{P}}[e^{-Y}] = \sum_{i=1}^{m} \log \mathbb{E}_{\mathbb{P}_i}[e^{-Y_i}],$$

which involves dealing only with the marginal distributions $\mathbb{P}_i$. If we have a tractable description of these marginal distributions, we can enforce robustness. On the other hand, the full distribution $\mathbb{P}$ requires m-dimensional convolution, which is computationally challenging; thus, implied bounds from robustness considerations could be useful in such a situation because the full distribution might be difficult to compute.

At the same time, because our focus is ambiguity, guarantees under only a reference distribution might not suffice (for instance, in the example just mentioned, perhaps the marginal distributions are not known with full precision, or the random variables might be somewhat correlated). In such situations, we might also want guarantees that are also valid under distributions that are within some distance of $\mathbb{P}$. To this end, it is helpful to have some pre-specified notion of what we mean by distance between distributions; we will find it convenient to use sets of the following form:

$$\mathcal{Q}^{\phi}(\Delta) = \left\{ \mathbb{Q} \in \mathcal{P}\colon \mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}} \right) \right] \leqslant \Delta \right\}, \tag{10}$$

for $\Delta \geqslant 0$, where, as introduced in Example 2.2, $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+$ is a convex function that attains a minimum value of 0 at 1. Notice that $\mathcal{Q}^{\phi}(\Delta)$ is the set of distributions within a ϕ-divergence of Δ from the reference distribution $\mathbb{P}$.

In sum, given $\mathbf{x} \in X^Q(\delta)$, we seek a bound on the quantity

$$\sup_{\{\mathbb{Q} \in \mathcal{Q}^{\phi}(\Delta),\, \mathbb{P} \in \mathcal{P}(X^Q(\delta))\}} \mathbb{Q}\{f(\mathbf{x}) \leqslant -z\},$$

and analogously for the standard robust formulation. Here, z is the downside performance level of interest and Δ is an ambiguity level. Note that $\mathcal{Q}^{\phi}(0) = \{\mathbb{P}\}$ here, so for $\Delta = 0$, any bounds will correspond simply to all admissible reference distributions (i.e., without ambiguity taken into account).

We first present a lemma that will be useful in establishing the main result of this section; these simple results might be of some independent interest outside the context of soft robustness.

Lemma 3.1. *Let $Y \in \mathcal{Y}$ be a random variable and ρ be a convex risk measure. The following hold*:

(a) *If ρ is an OCE generated by ϕ, then, for any z,*

$$\rho(Y) \leqslant 0 \implies \mathbb{P}\{Y < -z\} \leqslant \inf \left\{ p \in (0, 1]\colon (1-p)\phi(0) + p\phi\left(\frac{1}{p}\right) \leqslant z \right\}.$$

(b) *Let $\mathcal{P}_{\rho}(Y)$ be the set of distributions $\mathbb{P} \in \mathcal{P}$ such that $\rho(Y) \leqslant 0$ under $\mathbb{P}$. If $Y \leqslant \bar{\epsilon}$, then for any $z \geqslant -\bar{\epsilon}$,*

$$\sup_{\mathbb{P} \in \mathcal{P}_{\rho}(Y)} \mathbb{P}\{Y \leqslant -z\} = \sup\{p \in [0, 1]\colon \rho(Y_z) \leqslant 0\},$$

where Y_z is the random variable, such that $Y_z(\omega) = -z$ if $Y(\omega) \leqslant -z$ and $Y_z(\omega) = \bar{\epsilon}$ otherwise.

(c) *If Y satisfies $\mathbb{P}\{Y \leqslant -z\} \leqslant p$ for some z, then for any $\Delta > 0$,*

$$\sup_{\mathbb{Q} \in \mathcal{Q}^{\phi}(\Delta)} \mathbb{Q}\{Y \leqslant -z\} \leqslant \inf_{\lambda > 0,\, \nu} \left\{ \lambda\Delta + \lambda\left[\nu + p\phi^*\left(\frac{1}{\lambda} - \nu\right) + (1-p)\phi^*(-\nu) \right] \right\},$$

with equality holding if $\mathbb{P}\{Y \leqslant -z\} = p$.

Proof. See the appendix. □

We will explicitly use parts (b) and (c) of Lemma 3.1 in calculating downside probability guarantees under ambiguity for robust solutions. We will not use (a) directly but present it as it an auxiliary result.

To illustrate some of the statements of the lemma, consider the case of the entropic risk measure at level $\gamma \geqslant 0$. Here, ϕ is a relative entropy function: $\phi(z) = \gamma^{-1}(z \log z - z + 1)$, and $\phi(0) = \gamma^{-1}$, so (a) in the lemma gives

$$\frac{1}{\gamma}(1-p) + \frac{p}{\gamma}\left(-\frac{1}{p}\log p - \frac{1}{p} + 1 \right) \leqslant z$$
$$\Updownarrow$$
$$p \geqslant e^{-\gamma z},$$

i.e., $p(z) = e^{-\gamma z}$. Thus, with this choice of a convex risk measure, the probability that a feasible solution performs worse than $-z$ decreases exponentially with z.

Unfortunately, the bound $\mathrm{CVaR}_\beta(Y) \geqslant \mathrm{VaR}_\beta(Y)$ used in the proof of (a) in the lemma is often loose. Thus, even though these bounds can be exponential, they might nonetheless be weak. One valuable piece of information that is ignored is support information, which we have here in the case of bounded random variables. Exploiting the support information alone, as is done in part (b) of the lemma, we can substantially improve the bounds, as we now briefly discuss.

For instance, again with ρ being the entropic risk measure, we have $\phi^*(y) = \gamma^{-1}(e^{\gamma y} - 1)$, and the bound from (b) simplifies:

$$\begin{aligned}
&\sup_{\mathbb{P}\in\mathcal{P}_\rho(Y)} \mathbb{P}\{Y \leqslant -z\} \\
&\quad = \sup\{p \in [0,1]\colon\ \gamma^{-1}\log(pe^{\gamma z} + (1-p)e^{-\gamma\bar{\epsilon}}) \leqslant 0\} \\
&\quad = \frac{1 - e^{-\gamma\bar{\epsilon}}}{e^{\gamma z} - e^{-\gamma\bar{\epsilon}}}.
\end{aligned}$$

In the limit when $\bar{\epsilon} \to \infty$, this bound reduces to $e^{-\gamma z}$, which is the same as the bound from (a). For smaller values of the right support $\bar{\epsilon}$, however, this bound can be significantly tighter than the naive, exponential bound. Therefore, exploiting even crude information like the support of the distribution can be quite valuable in strengthening probability guarantees. This is illustrated in Figure 1. Of course, if one has additional information on the random variable (e.g., mean, variance, etc.), the bounds could be further tightened.

Remark 3.1. As a side note, the method in the proof of Lemma 3.1(a) also provides a means for approximating CVaR (which can be difficult to compute in general if the probability space is not compactly described) via convex risk measures. In particular, given the development in the proof of Proposition 2.2, we have

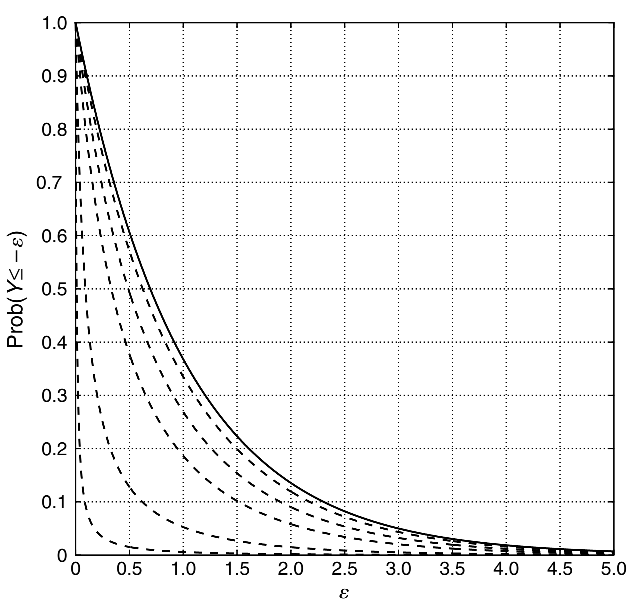
$$\inf_{\mathbb{Q}\in\mathcal{Q}(z)} \mathbb{E}_{\mathbb{Q}}[Y] = \inf_{\lambda>0}\left\{\lambda\rho\left(\frac{Y}{\lambda}\right) + z\lambda\right\} \geqslant \mathrm{CVaR}_\beta(Y),$$

provided that $\beta \geqslant \beta(z)$. If we know enough about the distribution to evaluate ρ, we can then bound $\mathrm{CVaR}_\beta(Y)$ by choosing z accordingly. Nemirovski and Shapiro (2006) use a special case of this in tractably bounding CVaR when one has information about the moment generating function. Specifically, using the entropic risk measure and choosing $z = -\log\beta$, we obtain the bound

$$\inf_{\lambda>0}\{\lambda\log\mathbb{E}_{\mathbb{P}}[e^{-Y/\lambda}] - (\log\beta)\lambda\} \geqslant \mathrm{CVaR}_\beta(Y).$$

This idea, proven in a very different fashion, is the basis for the bound used in the "Bernstein approximation" of Nemirovski and Shapiro (2006).

Figure 1. Bounds for the entropic risk measure for various values of the right support $\bar{\epsilon}$ and $\gamma = 1$.



Notes. From smallest to largest, $\bar{\epsilon} = 0.01, 0.1, 0.5, 1, 2$. The largest bound is $e^{-\epsilon}$.

We are now ready to compare downside probability guarantees for soft robust solutions to those of standard robust solutions.

Theorem 3.1. *Let* $\mathbf{x}_\delta^R \in X(\delta)$ *and* $\mathbf{x}_\delta^S \in X^Q(\delta)$ *be standard robust and soft robust feasible under the family of sets* $\mathbf{Q}$, *respectively, for some* $\delta \geqslant 0$, *and let* $\bar{\epsilon}$ *be an upper bound on* $f(\mathbf{x}_\delta^R)$ *and* $f(\mathbf{x}_\delta^S)$. *Let* $\lambda^R > 0$ *and* $\lambda^S \geqslant 0$ *be such that* $\lambda^R\delta + \lambda^R\rho^Q(f(\mathbf{x}_\delta^R)/\lambda^R) \leqslant 0$ *and* $\lambda^S\delta + (\lambda^S+1)\rho^Q(f(\mathbf{x}_\delta^S)/(\lambda^S+1)) \leqslant 0$ *and* Y_z *as in Lemma* 3.1(b). *Then for any* $\Delta \geqslant 0$ *and any* $z \geqslant -\bar{\epsilon}$,

$$\sup_{\{\mathbb{Q}\in\mathcal{Q}^\phi(\Delta),\,\mathbb{P}\in\mathcal{P}(X^Q(\delta))\}} \mathbb{Q}\{f(\mathbf{x}_\delta^S) \leqslant -z\} = g(\Delta, p^S(z)),$$

$$\sup_{\{\mathbb{Q}\in\mathcal{Q}^\phi(\Delta),\,\mathbb{P}\in\mathcal{P}(X(\delta))\}} \mathbb{Q}\{f(\mathbf{x}_\delta^R) \leqslant -z\} = g(\Delta, p^R(z)),$$

where

$$p^R(z) \triangleq \sup\left\{p \in [0,1]\colon\ \lambda^R\delta + \lambda^R\rho^Q\left(\frac{Y_z}{\lambda^R}\right) \leqslant 0\right\}$$

$$p^S(z) \triangleq \sup\left\{p \in [0,1]\colon\ \lambda^S\delta + (\lambda^S+1)\rho^Q\left(\frac{Y_z}{\lambda^S+1}\right) \leqslant 0\right\},$$

and

$$\begin{aligned}
g(\Delta, p) = \inf_{\lambda>0,\,\nu}\Bigg\{\lambda\Delta + \lambda\Bigg[\nu + p\phi^*\left(\frac{1}{\lambda} - \nu\right) \\
+ (1-p)\phi^*(-\nu)\Bigg]\Bigg\}
\end{aligned}$$

if $\Delta > 0$ *and* $g(0, p) = p$.

Proof. First, assume $\Delta > 0$. We first focus on the soft robust solution. We have, using Lemma 3.1(c),

$$\begin{aligned}
&\sup_{\{\mathbb{Q}^\phi\in\mathcal{Q}(\Delta),\,\mathbb{P}\in\mathcal{P}(X^Q(\delta))\}} \mathbb{Q}\{f(\mathbf{x}_\delta^S) \leqslant -z\}, \\
&\quad = \sup_{\{\mathbb{P}\in\mathcal{P}(X^Q(\delta))\}} g(\Delta, \mathbb{P}\{f(\mathbf{x}_\delta^S) \leqslant -z\}),
\end{aligned}$$

with $g(\Delta, p)$ as defined above. It is not hard to see that $g(\Delta, p)$ is nondecreasing in p, so the problem reduces to computing the maximum value of $\mathbb{P}\{f(\mathbf{x}_\delta^S) \leqslant -z\}$ over all admissible $\mathbb{P}$. Now, from Theorem 2.1, $\mathbf{x}_\delta^S \in X^Q(\delta)$ implies existence of a $\lambda^S \geqslant 0$ such that

$$\underbrace{\lambda^S\delta + (\lambda^S+1)\rho^Q(f(\mathbf{x}_\delta^S)/(\lambda^S+1))}_{\triangleq \tilde{\rho}^Q(f(\mathbf{x}_\delta^S))} \leqslant 0.$$

Notice that $\tilde{\rho}^Q$ is a convex risk measure, so we can apply Lemma 3.1(b). Namely, $f(\mathbf{x}_\delta^S)$ is a random variable bounded above by $\bar{\epsilon}$, so using the notation from the lemma,

$$\begin{aligned}
&\sup_{\{\mathbb{P}\in\mathcal{P}_{\tilde{\rho}^Q}(f(\mathbf{x}_\delta^S))\}} \mathbb{P}\{f(\mathbf{x}_\delta^S)\leqslant -z\}\\
&\quad=\sup\{p\in[0,1]\colon \tilde{\rho}^Q(Y_z)\leqslant 0\}\\
&\quad=\sup\left\{p\in[0,1]\colon \lambda^S\delta+(\lambda^S+1)\rho^Q\left(\frac{Y_z}{\lambda^S+1}\right)\leqslant 0\right\}\\
&\quad=p^S(z),
\end{aligned}$$

where in the second equality we are using the definition of $\tilde{\rho}^Q$. This establishes the claim for positive Δ for the soft robust case.

For the standard robust case, the arguments are nearly identical, except that we have $\mathbf{x}_\delta^R \in X(\delta)$ and $\lambda^R > 0$ satisfying

$$\underbrace{\lambda^R\delta + \lambda^R\rho(f(\mathbf{x}_\delta^R)/\lambda^R)}_{\triangleq \hat{\rho}^Q(f(\mathbf{x}_\delta^R))} \leqslant 0.$$

Now applying Lemma 3.1(b) in the same manner as above with the convex (in fact, coherent) risk measure $\hat{\rho}^Q$, we obtain $p^R(z)$.

For $\Delta = 0$, there is no ambiguity, and $\mathcal{Q}(0) = \{\mathbb{P}\}$, so it is simply a matter of maximizing under the reference measure $\mathbb{P}$. The arguments then go through as above simply on $p^R(z)$ and $p^S(z)$. $\square$

Theorem 3.1 provides bounds on ambiguous distributions for soft and standard robust solutions. Although the bounds look complicated, they are fairly simple to compute because they can be found via one-dimensional convex problems. To connect to our running example of entropic risk measure, the robust bound in this case with $\gamma = 1$ is

$$\begin{aligned}
p^R(z)&=\sup\big\{p\in[0,1]\colon\\
&\qquad \lambda^R\delta+\lambda^R\log[pe^{z/\lambda^R}+(1-p)e^{-\bar{\epsilon}/\lambda^R}]\leqslant 0\big\}\\
&=\frac{e^{-\delta}-e^{-\bar{\epsilon}/\lambda^R}}{e^{z/\lambda^R}-e^{-\bar{\epsilon}/\lambda^R}}.
\end{aligned}$$

Similarly, we obtain for the soft robust solution

$$p^S(z)=\frac{e^{-\delta\lambda^S/(\lambda^S+1)}-e^{-\bar{\epsilon}/(\lambda^S+1)}}{e^{z/(\lambda^S+1)}-e^{-\bar{\epsilon}/(\lambda^S+1)}}.$$

For $\bar{\epsilon}\to\infty$, these bounds reduce to

$$\begin{aligned}
p^R(z)&=e^{-\delta}e^{-z/\lambda^R},\\
p^S(z)&=e^{-\delta\lambda^S/(\lambda^S+1)}e^{-z/(\lambda^S+1)}.
\end{aligned}$$

Roughly speaking, for some parameter λ, the standard robust bound scales as $\exp(-z/\lambda)$, and the soft robust bound scales as $\exp(-z/(\lambda+1))$. When λ is large relative to z, the bounds will not be very different. For λ much smaller than z, however, the soft robust bound can be considerably larger. As noted, the optimal λ from the robust problems is nonincreasing with δ; we can therefore interpret small enough values of λ to correspond to appreciably high δ. Because δ is large in these situations, standard robustness is costly, and we would expect the soft robust solution to have weaker bounds as softening is a significant relaxation in such cases.

In this example, we also note that the bounds in both cases are exponential and will dominate a sharp Chebyshev bound, which scales roughly as z^{-2}, for large enough deviations.

To accommodate ambiguity in the bounds, we need to compute (with, for example, ϕ as relative entropy) the function

$$g(\Delta, p)=\inf_{\lambda>0}\big\{\lambda\Delta+\lambda\log[pe^{1/\lambda}+(1-p)]\big\}.$$

This is done easily because it is a one-dimensional convex problem. We point out that if we have bounds that are tighter than $p^R(z)$ or $p^S(z)$ (e.g., from a sharp Chebyshev inequality or another moment-based bound), we can always take p as the best available bound and supply it to $g(\Delta, p)$ to get improved bounds that are still valid under ambiguity.

4. Applications

In this section, we illustrate the soft robust approach on two examples: first, allocation of a portfolio of bonds with default probabilities that might be uncertain; second, in a portfolio optimization problem with rebalancing and historical market data.

4.1. Risky Bond Portfolio

We first illustrate our approach, including the bounds from Theorem 3.1, with an example involving optimization across a set of risky bonds. We have $n = 49$ bonds and a single, riskless asset that pays a risk-free rate of $r_f = 0.005$. Each of the bonds may default and we assume in the reference model that defaults are independent. A bond defaults with probability p_i and becomes worthless; otherwise, with probability $1 - p_i$, the bond pays off $r_i > r_f$. We set $r_i = (r_f + 2p_i^2 + p_i)/(1 - p_i)$; this ensures that the expected payoff of each bond is greater than r_f and strictly increasing with p_i, so that bonds with higher default probabilities pay out more. For the purposes of this example, we randomly generate default probabilities in the range $[0.001, 0.05]$.

Note that the dimension of the state space is 2^{49}, so it will likely be difficult to exactly evaluate the full probability distribution of an allocation in these bonds.

As default probabilities are invariably not known with full precision, and the independence assumption is likely to be incorrect, ambiguity is a concern. We therefore consider the optimal allocations according to the standard and soft robust approaches and compare the results. Specifically, we solve the two problems:

$$\max\{\hat{\mathbf{r}}'\mathbf{x}\colon \mathbf{x} \in X(\delta)\}, \quad \text{and} \quad \max\{\hat{\mathbf{r}}'\mathbf{x}\colon \mathbf{x} \in X^{Q}(\delta)\},$$

where $\hat{r}_i$ is the expected return of asset i, i.e., $\hat{r}_i = -p_i + (1-p_i)r_i$ for $i < 50$ and $\hat{r}_{50} = r_f$. Here we will use the family of sets $\mathbf{Q}$ generated by relative entropy as in (10), so the resulting optimization problems will involve transformations of the entropic risk measure. The standard robust problem, for instance, can be written as

$$\max_{\mathbf{x}\in X,\,\lambda>0}\left\{\hat{\mathbf{r}}'\mathbf{x}\colon \lambda\delta + \lambda\sum_{i=1}^{49}\log\left[p_i\exp\left(\frac{x_i}{\lambda}\right) + (1-p_i)\exp\left(\frac{-r_i x_i}{\lambda}\right)\right] \leqslant r_f x_{50}\right\},$$

which is a convex optimization problem in $(\mathbf{x}, \lambda)$. The soft robust problem is similar:

$$\max_{\mathbf{x}\in X,\,\lambda\geqslant 0}\left\{\hat{\mathbf{r}}'\mathbf{x}\colon \lambda\delta + (\lambda+1)\sum_{i=1}^{49}\log\left[p_i\exp\left(\frac{x_i}{\lambda+1}\right) + (1-p_i)\exp\left(\frac{-r_i x_i}{\lambda+1}\right)\right] \leqslant r_f x_{50}\right\}.$$

Here, we will use the set $X = \{\mathbf{x}\colon \mathbf{e}'\mathbf{x} = 1, \mathbf{x}_i \geqslant -0.2\ \forall i\}$, i.e., we allow a limit of 20% in any leveraged asset.

We solve both problems for various choices of the parameter δ. To motivate the choice of δ, we consider an error factor of $\kappa \geqslant 1$ on the default probabilities. If we restrict ourselves to default probabilities within this factor, the relative entropy $\delta(\kappa)$ to the perturbed distribution is given by

$$\delta(\kappa) = \sum_{i=1}^{n}\left[\kappa p_i \log(\kappa) + (1-\kappa p_i)\log\left(\frac{1-\kappa p_i}{1-p_i}\right)\right].$$

We solve the standard robust and soft robust problems for $\kappa \in \{1.005, 1.01, 1.1, 10\}$ and compare the results. To solve these problems, we use the software package ROME (Goh and Sim 2009) in a MATLAB environment. On a personal computer with an Intel Pentium® 3.0 GHz processor and 2.0 GB of RAM, solution of each problem for a given $\delta(\kappa)$ takes about 30 seconds.

The results are shown in Table 1. Each row of the table shows the maximum return of a position (with no defaults), the expected return under the nominal distribution, and bounds on a 10% and 20% loss in value both under the nominal distribution and for all distributions within a relative entropy of Δ from $\mathbb{P}$. Here, the relative entropy on the ambiguous distribution bounds was chosen because δ was above with a factor of $\kappa = 1.5$.[5]

We make a few points on the results. First, the standard robust solutions quickly become highly conservative, even for small κ (and hence small δ). For instance, even for sets corresponding to a perturbation of the default probabilities by 10% (which seems very small), the standard robust solution recommends an 80%+ investment in the risk-free asset and generates only an additional 0.05% in expected return over the risk-free rate. Of course, such a position also is very unlikely to experience large losses, and the downside guarantees reflect this.

By contrast, the soft robust solutions have higher expected return and more upside and downside risk. In some cases, the difference from the standard robust approach seems dramatic. For instance, for $\kappa = 1.1$, the soft robust approach results in about an extra 1.2% in expected return and over 14% more in maximum return. The downside guarantees are of course somewhat worse (still, for this position, the chance under $\mathbb{P}$ of a 10% loss in portfolio

Table 1. The table shows the following, from left to right, for both the standard robust and the soft robust approaches in the risky bond example of §4.1: maximum return (no defaults), expected return (under $\mathbb{P}$), and downside guarantees at −10% and −20% under both $\mathbb{P}$ and ambiguous distributions within a relative entropy of Δ from $\mathbb{P}$; this parameter is selected as $\Delta = \delta(1.5)$.

κ	$\mathbf{r}'\mathbf{x}^\star$	$\mathbb{E}_{\mathbb{P}}[\hat{\mathbf{r}}'\mathbf{x}^\star]$	$\mathbb{P}\{\hat{\mathbf{r}}'\mathbf{x}^\star \leqslant -0.1\}$	$\sup_{\mathbb{Q}\in\mathcal{Q}(\Delta)}\mathbb{Q}\{\hat{\mathbf{r}}'\mathbf{x}^\star \leqslant -0.1\}$	$\mathbb{P}\{\hat{\mathbf{r}}'\mathbf{x}^\star \leqslant -0.2\}$	$\sup_{\mathbb{Q}\in\mathcal{Q}(\Delta)}\mathbb{Q}\{\hat{\mathbf{r}}'\mathbf{x}^\star \leqslant -0.2\}$
			Standard robust			
1.005	0.2899	0.0334	0.3025	0.5704	0.1208	0.3304
1.010	0.1722	0.0188	0.0346	0.1747	0.0127	0.1136
1.100	0.0123	0.0055	6.6e−6	0.0216	5.8e−11	0.0091
10.00	0.0053	0.0050	0	0	0	0
			Soft robust			
1.005	0.2933	0.0338	0.3248	0.5954	0.1299	0.3471
1.010	0.2000	0.0220	0.0578	0.2244	0.0022	0.1417
1.100	0.1557	0.0169	0.0248	0.1502	0.0089	0.0995
10.00	0.1557	0.0169	0.0248	0.1502	0.0089	0.0995

value is less than 3%). Also of note is that in this example, soft robustness at fairly small levels (e.g., $\kappa = 1.1$) implies soft robustness at larger levels (e.g., $\kappa = 10$). This is not surprising in light of the discussion in §2.4.

The final point worth noting here is the dramatic impact ambiguity can have in inflating downside guarantees. For instance, if $\mathbb{P}$ is an accurate probabilistic model, the standard robust solution has an extremely small chance (0.00066%) of a 10% loss. Yet even for fairly small perturbations of the distribution, this guarantee increases in the worst-case to a nonnegligible value (about 2%). While we motivated the choice of the ambiguity sets through errors in default assessments, ambiguity is a concern for other reasons (such as defaults are likely to be somewhat correlated in reality), and ignoring it in analyzing downside risks can lead to assessments that might be grossly understated.

4.2. Asset Allocation

Here we apply the soft robust approach to an asset allocation problem using real-world financial data. In particular, we investigate the out-of-sample performance of using a soft robust approach as opposed to a standard robust approach. We will also compare these approaches to an approach based on CVaR.

We consider an investor who wishes to allocate wealth among n assets. The decision vector $\mathbf{x} \in \mathbb{R}^n$ denotes the vector of weights the investor allocates to each asset for a current time period. We will use the feasible set $\mathbf{x} \in X := \{\mathbf{x} \in \mathbb{R}^n\colon \mathbf{e}'\mathbf{x} = 1,\ \mathbf{x} \geqslant \mathbf{0}\}$. In a real-world setting, of course, there might be many constraints in addition to this.

The n assets have an associated, random return vector $\tilde{\mathbf{r}}$ over the time period, with $\mathbb{P}\{\tilde{\mathbf{r}} \geqslant -\mathbf{1}\} = 1$. The portfolio return after a single period is therefore $\tilde{\mathbf{r}}'\mathbf{x}$. Denote $\mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{r}}]$ by $\hat{\mathbf{r}}$. For various types of robust formulations, we will find an optimal portfolio in a given period. We will then implement this choice on a new return, based on past market data, and store the result. We will then re-optimize.

For the robust formulations, let $\mathbf{Q}$ denote the family of sets. For our purposes, we will use the sets from (10) generated by the entropic risk measure with parameter $\gamma \geqslant 0$. In particular, we have

$$\mathscr{Q}(\epsilon) = \left\{\mathbb{Q} \in \mathscr{P}\colon \frac{1}{\gamma}\mathbb{E}_{\mathbb{P}}\left[\phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] \leqslant \epsilon\right\},$$

with ϕ being relative entropy. In the example that follows, we will use the parameter value $\gamma = 50$.[6]

During each time period, we will first solve, as a benchmark, the problem

$$\max\{\hat{\mathbf{r}}'\mathbf{x}\colon \mathbf{x} \in X, \mathrm{CVaR}_\beta(\tilde{\mathbf{r}}'\mathbf{x}) \leqslant 0\},$$

for the values $\beta = 0.1, 0.2, \ldots, 0.9$. Notice that as β increases, the CVaR constraint becomes less restrictive and therefore the (in-sample) expected return should increase and the portfolio choices will become more aggressive. Note that CVaR can be viewed as a standard robust constraint (with a different uncertainty set), as CVaR is a coherent risk measure (see the discussion in the examples following Theorem 2.1).

For each value of β and during each time period, we will also solve the standard robust and soft robust problems:

$$\max\{\hat{\mathbf{r}}'\mathbf{x}\colon \mathbf{x} \in X^{Q}(\delta)\},$$

$$\max\{\hat{\mathbf{r}}'\mathbf{x}\colon \mathbf{x} \in X(\delta)\},$$

for some value δ that we will vary with β in the CVaR formulation. We will choose δ to be as small as possible, such that the standard robust formulation is no less robust than the CVaR constraint, i.e., such that for any random variable Y,

$$\inf_{\lambda > 0}\left\{\lambda\delta + \lambda\rho^{Q}\left(\frac{Y}{\lambda}\right)\right\} \leqslant \mathrm{CVaR}_\beta(Y).$$

Following the reasoning in the proof of Lemma 3.1(a), the choice $\delta = \gamma^{-1}\log\beta^{-1}$ is the smallest such choice that guarantees this. In general, the standard robust approach with this choice of δ will be more conservative than the corresponding CVaR approach, perhaps considerably so; there is no reason to expect this approach to ever outperform CVaR in expected return with this parameter choice.

Finally, for each of these cases we will solve a "comprehensive robust" version of the problem:

$$\max\{\hat{\mathbf{r}}'\mathbf{x}\colon \mathbf{x} \in X_C^{Q}(\delta)\},$$

where

$$X_C^{Q}(\delta) = \left\{\mathbf{x} \in X\colon \min_{\lambda \in (0,1]}\left\{\lambda\delta + \lambda\rho^{Q}\left(\frac{\tilde{\mathbf{r}}'\mathbf{x}}{\lambda}\right)\right\} \leqslant 0\right\}.$$

The robustness interpretation here is that standard robustness is ensured for all $\mathbb{Q} \in \mathscr{Q}(\delta)$, and for measures outside this set, feasibility is ensured at levels that weaken with the relative entropy divergence (see Example 2.3). This is the most conservative of the four formulations.

It is not obvious that these problems will necessarily be feasible for particular parameter choices. If, however, we assume the presence of a risk-free asset with constant return $r_f \geqslant 0$, then the problem is always feasible. Indeed, if the investor invests all his or her wealth in the risk-free asset, we have, for any normalized convex risk measure ρ,

$$\rho(r_f) = \rho(0) - r_f = -r_f \leqslant 0,$$

where we are using translation invariance and the fact that ρ is normalized. Alternatively, if we assume the presence of a nearly risk-free asset whose returns are not constant but are always nonnegative (e.g., cash), the problem is also ensured to be feasible. In this case, denoting the nearly risk-free returns by $\tilde{r}_f \geqslant 0$, we have

$$\rho(\tilde{r}_f) \leqslant \rho(0) = 0,$$

where we are using monotonicity and the fact that ρ is normalized. Because all four approaches are equivalent to a constraint based on some normalized convex risk measure, feasibility is ensured.

For our empirical study, we use monthly historical returns for $n = 11$ publicly traded asset classes over the period of April, 1981 through February, 2006. The asset classes are listed in Table 2. In Table 3, we list the realized CVaR and VaR for the various assets based on the data from this time period. Note again that $\text{CVaR}_\beta(R)$ can, roughly speaking, be interpreted as that the (negative of the) expected value of the asset's return, given that the return is in the lower β-tail of its distribution; in particular, $\text{CVaR}_1(R)$ is the negative of the expected return of the asset.

Using the data described in the previous section, we solved the four formulations for each of the ten choices of β and compared the results. In this setup, we used a sliding window of the past three years of returns as the sample data for solving the problems: i.e., the previous $N = 36$ returns formed the discrete reference distribution $\mathbb{P}$ that we used to solve the problem. For each set of parameters and each formulation, we then implemented the optimal portfolio over the following year's worth of new returns, then re-balanced.[7] We repeated this process over each year within the entire data range and tabulated the out-of-sample performance statistics for each of the formulations.

In Table 4, we see a comparison of the out-of-sample performance in terms of CVaR_β and expected return for the various risk measures. Table 5 shows the probability that the out-of-sample return (annualized) drops below a pre-specified threshold for the different risk measures.

Some key observations from these empirical results are the following:

1. The standard robust and comprehensive robust approaches generally did the the best in terms of closeness to achieving a realized CVaR_β less than 0. This matches intuition, because they are the most conservative risk measures of the four. Note that there is no guarantee that an in-sample CVaR constraint will ensure satisfaction of the CVaR constraint out-of-sample (as we see in Table 4, in fact this is generally not the case). As β increases, and the optimal solution becomes more aggressive, it seems that the robust approaches tended to become better in out-of-sample CVaR_β; interestingly, the CVaR-based approach seemed to violate the constraint out-of-sample worse as β increased up to $\beta = 0.7$, then improved again for larger β. There is a trade-off in the out-of-sample CVaR value as β increases: the portfolios are becoming more aggressive, but at the same time, larger β means the CVaR level is less restrictive. Evidently, for the robust formulations, the latter effect tended to dominate more as β increased, but this was not the case for the CVaR-based approach.

2. Somewhat surprisingly, the soft robust approach significantly outperformed CVaR in terms of realized risk (measured in terms of CVaR_β for $\beta = 0.3, \ldots, 0.8$). This risk reduction was not always offset by a decrease in rate of return. For example, for $\beta = 0.8$ and $\beta = 0.9$, the soft robust measure significantly outperformed CVaR; this is clearly demonstrated in Figure 2 (in the electronic companion).

3. The comprehensive robust solutions were quite similar to the standard robust solutions for $\beta \leqslant 0.5$, as was the corresponding performance and risk. For $\beta > 0.5$, however, the standard robust approach offered an average of $+3.68\%$ expected return over the comprehensive robust approach. The probability of bad performance, however, was significantly higher for large β for the standard robust (e.g., 19.8% vs. 2.0% of the (annualized) monthly return dropping below -20% for the case $\beta = 0.9$).

4. Although CVaR had the highest expected performance in many of the cases, this was not always so (see point 2 above), and in nearly every case listed in Table 5, the investment strategies using CVaR had the highest probability of bad performance (the only exceptions were against the soft robust for $\beta = 0.1, 0.2$). Figure 2 emphasizes this graphically for the case $\beta = 0.8$; note the large dips in cumulative return for the CVaR investment strategy. In this case, both the standard robust and the soft robust approaches outperform CVaR in final, realized wealth (with soft robust slightly better than standard robust, as in the other cases). Noteworthy here is the sizable downward spike in wealth for CVaR with about 3–4 years left. Neither the standard robust nor the soft robust approach suffers from this in this case. In general, these two robust approaches seem to yield more stable growths in wealth over time than the CVaR approach does.

5. Over the nine values of β, the average out-of-sample benefit of the soft robust measure over the standard robust approach was $+0.38\%$ of expected return. In fact, in every case, the soft robust approach performed better out-of-sample than the standard robust approach. This was traded off at a cost of a higher level of downside risk. The difference in downside performance levels (Table 5), however, often seemed marginal and was in fact very similar for $\beta > 0.5$ across the standard and soft robust approaches.

In summary, it seems that by relaxing the robustness requirements to soft robust requirements, we can potentially gain out-of-sample performance for not too high a price in increased downside risk. We emphasize that this was the case for all the computational experiments here, each of which was implemented over a 25-year investment horizon. The results here seem too one-sided and are based on data that are too expansive to be just a matter of happenstance. Perhaps using tools from statistical learning theory, investigating, the explicit benefits of soft robustness for out-of-sample performance in asset allocation is an interesting direction for further research.

5. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Endnotes

1. We sometimes refer to $\mathbb{P}$ as the *reference distribution*.
2. We can, of course, assume the right-hand side is zero without loss of generality simply be shifting the objective by a constant value.
3. Here, convex means if $\mathbb{Q}_1 \in \mathcal{Q}(\epsilon_1)$ and $\mathbb{Q}_2 \in \mathcal{Q}(\epsilon_2)$, then for all $\lambda \in [0, 1]$, $\lambda\mathbb{Q}_1 + (1-\lambda)\mathbb{Q}_2 \in \mathcal{Q}(\lambda\epsilon_1 + (1-\lambda)\epsilon_2)$.
4. We thank an anonymous referee for this example.
5. For the bounds under $\mathbb{P}$, we took the minimum of the bounds from Theorem 3.1 and those based on a sharp Chebyshev bound; although the Theorem 3.1 bounds scale better (exponentially) than Chebyshev bounds for large deviations, for these parameter values the Chebyshev bounds were tighter than the soft robust bounds and looser than the standard robust bounds. Regardless, as mentioned in §3, we are always free to use the better of the two bounds in the function $g(\Delta, p)$ when computing guarantees that account for ambiguity, as we have done in the adjacent columns.
6. A rough interpretation of γ is as the reciprocal of the risk tolerance for a CARA utility investor. We note that the standard robust formulation is unaffected by the choice of γ, as this formulation is equivalent to a coherent risk measure; for the soft robust formulation, we found $\gamma = 50$ convenient to illustrate the approach as it provided results in this example that varied smoothly across the ten cases.
7. We did not account for transactions costs in our results. We would expect with those included that the gap in out-of-sample performance between the more conservative solutions (e.g., comprehensive robustness) and the more aggressive solutions (e.g., CVaR) would narrow somewhat.

Acknowledgments

The authors are grateful to two anonymous referees for a number of helpful comments and insights that improved the paper.

References

Artzner, P., F. Delbaen, J. Eber, D. Heath. 1999. Coherent measures of risk. *Math. Finance* **9** 203–228.

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

Ben-Tal, A., M. Teboulle. 1991. Portfolio theory for the recourse certainty equivalent maximizing investor. *Ann. Oper. Res.* **31** 479–500.

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

Ben-Tal, A., S. Boyd, A. Nemirovski. 2006. Extending the scope of robust optimization. *Math. Programming, Ser. B* **107** 63–89.

Bertsimas, D., D. B. Brown. 2009. Constructing uncertainty sets for robust linear optimization. *Oper. Res.* **57**(6) 1483–1495.

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

Calafiore, G., L. El Ghaoui. 2006. Distributionally robust chance-constrained linear programs with applications. *J. Optim. Theory Appl.* **130**(1) 1–22.

Chen, X., M. Sim, P. Sun. 2007. A robust optimization perspective to stochastic programming. *Oper. Res.* **55**(6) 1058–1071.

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

El Ghaoui, L., M. Oks, F. Oustry. 2003. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Oper. Res.* **51**(4) 543–556.

El Ghaoui, L., F. Oustry, H. Lebret. 1998. Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.* **9**(1) 33–52.

Ellsberg, D. 1961. Risk, ambiguity, and the Savage axioms. *Quart. J. Econom.* **75** 643–669.

Epstein, L., T. Wang. 1994. Intertemporal asset pricing under Knightian uncertainty. *Econometrica* **62**(2) 283–322.

Erdogan, E., G. Iyengar. 2006. Ambiguous chance constrained problems and robust optimization. *Math. Programming* **107**(1–2) 37–61.

Fischetti, M., M. Monaci. 2008. Light robustness. Working paper, DEI, Università di Padova, Padua, Italy.

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

Föllmer, H., A. Schied. 2004. *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter, Berlin.

Gilboa, I., D. Schmeidler. 1989. Maxmin expected utility with a non-unique prior. *J. Math. Econom.* **18** 141–153.

Goh, J., M. Sim. 2009. ROME: Robust optimization made easy. Version 1.0.4 (beta), http://robustopt.com/.

Goldfarb, D., G. Iyengar. 2003. Robust portfolio selection problems. *Math. Oper. Res.* **28** 1–38.

Hansen, L., T. Sargent. 2001. Robust control and model uncertainty. *Amer. Econom. Rev.* **91** 60–66.

Klibanoff, P., M. Marinacci, S. Mukerji. 2005. A smooth model of decision making under ambiguity. *Econometrica* **73**(6) 1849–1892.

Knight, F. H. 1921. *Risk, Uncertainty, and Profit*. Houghton Mifflin, New York.

Lim, A. E. B., J. G. Shanthikumar, T. Watewai. 2010. Robust asset allocation with benchmarked objectives. *Math. Finance.* Forthcoming.

Luenberger, D. G. 1969. *Optimization by Vector Space Methods*. Wiley, New York.

Lüthi, H.-J., J. Doege. 2005. Convex risk measures for portfolio optimization and concepts of flexibility. *Math. Programming, Ser. B* **104**(2–3) 541–559.

Maccheroni, F., M. Marinacci, A. Rustichini. 2006. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica* **74**(6) 1447–1498.

Natarajan, K., D. Pachamanova, M. Sim. 2009. Constructing risk measures from uncertainty sets. *Oper. Res.* **57**(5) 1129–1141.

Natarajan, K., M. Sim, J. Uichanco. 2010. Tractable robust expected utility and risk models for portfolio optimization. *Math. Finance*. Forthcoming.

Nemirovski, A., A. Shapiro. 2006. Convex approximations of chance constrained programs. *SIAM J. Optim.* **17**(4) 969–996.

Rigotti, L., C. Shannon, T. Strzalecki. 2008. Subjective beliefs and ex ante trade. *Econometrica* **76**(5) 1167–1190.

Rockafellar, R. T., S. P. Uryasev. 2000. Optimization of conditional value-at-risk. *J. Risk* **2** 21–41.

Ruszczyński, A., A. Shapiro. 2006. Optimization of convex risk functions. *Math. Oper. Res.* **3**(31) 433–452.

Schmeidler, D. 1989. Subjective probability and expected utility without additivity. *Econometrica* **57** 571–587.