European Journal of Operational Research 241 (2015) 771–782

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

Restricted risk measures and robust optimization

Guido Lagos [a], Daniel Espinoza [b], Eduardo Moreno [c], Juan Pablo Vielma [d,*]

[a] H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA
[b] Department of Industrial Engineering, Universidad de Chile, Santiago, Chile
[c] Faculty of Engineering and Sciences, Universidad Adolfo Ibañez, Santiago, Chile
[d] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA

ARTICLE INFO

Article history:
Received 3 October 2012
Accepted 20 September 2014
Available online 7 October 2014

Keywords:
Risk management
Stochastic programming
Uncertainty modeling

ABSTRACT

In this paper we consider characterizations of the robust uncertainty sets associated with coherent and distortion risk measures. In this context we show that if we are willing to enforce the coherent or distortion axioms only on random variables that are affine or linear functions of the vector of random parameters, we may consider some new variants of the uncertainty sets determined by the classical characterizations. We also show that in the finite probability case these variants are simple transformations of the classical sets. Finally we present results of computational experiments that suggest that the risk measures associated with these new uncertainty sets can help mitigate estimation errors of the Conditional Value-at-Risk.

1. Introduction

Coherent risk measures and their relation to robust optimization have received significant attention in the literature (Artzner, Delbaen, Eber, & Heath, 1999; Bertsimas & Brown, 2009; Natarajan, Pachamanova, & Sim, 2009; Shapiro, Dentcheva, & Ruszczyński, 2009; Ben-Tal, Ghaoui, & Nemirovski, 2009; Wächter & Mazzoni, 2013). It is known that every coherent risk measure is associated with a precisely determined convex uncertainty set with properties that are strongly tied to the axioms characterizing coherent risk measures (e.g. Bertsimas and Brown (2009); Natarajan et al. (2009)). Similar results have also been given for a special class of coherent risk measures known as distortion risk measures, which include the widely used Conditional Value-at-Risk (Bertsimas & Brown, 2009; Pichler & Shapiro, 2013; Shapiro, 2013). All these characterizations are based on the restrictions imposed by the coherence or distortion axioms on the actions of the coherent risk measure over all possible random variables. However, in many settings, the random variables considered are either an affine or linear function of a, potentially correlated, vector of random parameters. A classical example is portfolio optimization (see for example Markowitz (1952); Konno and Yamazaki (1991); Black and Litterman (1992); Cvitanić and Karatzas (1992); Krokhmal, Palmquist, and Uryasev (2002); Zymler, Rustem, and Kuhn (2011); Lim, Shanthikumar, and Vahn (2011); Kawas and Thiele (2011); Fertis, Baes, and Lüthi (2012); Kolm, Tütüncü, and Fabozzi (2014)) where the random return of a portfolio is usually modeled as a weighted linear combination of the random returns of individual assets (with weights equal to the fraction invested in a given asset) plus a possibly null constant representing investment in a riskless asset. In this paper we show that imposing the coherence and distortion axioms *only* on random variables that are a linear, or affine linear function of a vector of random variables allows the inclusion of uncertainty sets that are deemed invalid by the classical characterizations. In particular, we show that in the finite probability case these additional sets at least include certain expansions of the classical sets. We also show that such expansions are in turn related to the common practice of taking the convex combination of a risk measure with the expected value. More specifically, we show that risk measures associated to these expansions are affine combinations of a risk measure with the expected value.

Finally we present computational experiments that suggest that the risk measures associated with these uncertainty sets can help mitigate estimation errors of the Conditional Value-at-Risk.

The rest of this paper is organized as follows. In Section 2 we give some notation and background on risk measures and robust optimization. In Section 3 we show the existence of uncertainty sets that do not fall in the classical characterizations, but do yield distortion risk measures on the subspace of random variables that are either affine or linear functions of a fixed random vector. In Section 4 we show that the risk measures associated to these uncertainty sets are affine combinations of a risk measure with the expected value. Then, in Section 5 we present some results of computational experiments showing that these uncertainty sets could be useful to mitigate estimation errors. Finally, in Section 6 we present some final remarks.

* Corresponding author. Tel.: +1 617 324 1204.
E-mail addresses: glagos@gatech.edu (G. Lagos), daespino@dii.uchile.cl (D. Espinoza), eduardo.moreno@uai.cl (E. Moreno), jvielma@mit.edu (J. P. Vielma).

http://dx.doi.org/10.1016/j.ejor.2014.09.024
0377-2217/

2. Notation and background on risk measure and robust optimization

2.1. Notation

Throughout the paper we will use bold letters to denote column vectors, and we will use an apostrophe to denote the transposition operation. Thus, $\boldsymbol{x} \in \mathbb{R}^d$ is a column vector and $\boldsymbol{x}'$ its transpose. We also note $\boldsymbol{e}$ as the vector with a 1 in every component and $\boldsymbol{e}_N := \frac{1}{N}\boldsymbol{e}$. For a given set $S \subseteq \mathbb{R}^n$ we denote by $\text{aff}(S)$, $\text{conv}(S)$ and $\overline{\text{conv}}(S)$ its affine, convex and closed convex hull respectively. We also let $\text{lin}(S)$ be the linear space spanned by S and $\text{ri}(S)$ the relative interior of S. For a given convex set C we denote by $\text{ext}(C)$ the set of its extreme points. To denote index sets, we use $[m] := \{1, \ldots, m\}$.

2.2. Coherent risk measures

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $L_1(\Omega, \mathcal{F}, \mathbb{P})$ be the set of integrable random variables that are an outcome of the uncertain parameter in Ω. We use a tilde to identify random variables as in $\tilde{g} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2.1. A function $\rho : L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ is a coherent risk measure if it satisfies the following properties.

(C1) Convexity: $\rho(t\tilde{g}_1 + (1-t)\tilde{g}_2) \le t\rho(\tilde{g}_1) + (1-t)\rho(\tilde{g}_2)$ for all $\tilde{g}_1, \tilde{g}_2 \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ and $t \in [0, 1]$.
(C2) Positive homogeneity: $\rho(t\tilde{g}) = t\rho(\tilde{g})$ for all $\tilde{g} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ and $t > 0$.
(C3) Translation equivariance: $\rho(t + \tilde{g}) = t + \rho(\tilde{g})$ for all $\tilde{g} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ and $t \in \mathbb{R}$.
(C4) Monotonicity: $\rho(\tilde{g}_1) \le \rho(\tilde{g}_2)$ for all $\tilde{g}_1, \tilde{g}_2 \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ such that $\tilde{g}_1 \le \tilde{g}_2$ *a.s.*

The following theorem gives another characterization of coherent risk measures (Shapiro et al., 2009, Theorem 6.4).

Theorem 2.2. *Let*

$$\Delta := \left\{ f \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}) \ : \ \int_\Omega f(\omega) d\mathbb{P}(\omega) = 1 \right\} \text{ and} \tag{1a}$$

$$\Delta_+ := \left\{ f \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}) \ : \ \int_\Omega f(\omega) d\mathbb{P}(\omega) = 1, \ f(\omega) \ge 0 \ a.s. \right\}. \tag{1b}$$

Then a function $\rho : L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ satisfies (C1)–(C3) if and only if there exists $\mathcal{J} \subseteq \Delta$ such that

$$\rho(\tilde{g}) = \sup_{f \in \mathcal{J}} \int_\Omega \tilde{g}(\omega) f(\omega) d\mathbb{P}(\omega). \tag{2}$$

The function additionally satisfies (C4) if and only if $\mathcal{J} \subseteq \Delta_+$. Finally, if ρ satisfies (C1)–(C3), then it is additionally continuous. In that case, we have that $\mathcal{J}$ is convex and weakly compact.*

A relation between risk measures and robust uncertainty sets emerges when we focus on random variables that are affine or linear functions of a fixed d-dimensional random vector $\tilde{\boldsymbol{u}} \in L_1^d(\Omega, \mathcal{F}, \mathbb{P})$ (i.e. $\tilde{\boldsymbol{u}}_i \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ for each $i \in [d]$). For instance $\tilde{\boldsymbol{u}}$ could be the random returns on d assets and we may be interested in analyzing random portfolio returns of the form $\tilde{g}_{\boldsymbol{x}}(\tilde{\boldsymbol{u}}(\omega)) := \sum_{i=1}^d \boldsymbol{x}_i \tilde{\boldsymbol{u}}_i(\omega)$ where $\boldsymbol{x} \in \mathbb{R}^d$ indicates the fractions invested in each asset (i.e. $\boldsymbol{x} \in [0, 1]^d$ and $\sum_{i=1}^d \boldsymbol{x}_i = 1$). In general, this corresponds to restricting attention to the subspaces of $L_1(\Omega, \mathcal{F}, \mathbb{P})$ given by

$$\begin{aligned}
\mathcal{V}(\tilde{\boldsymbol{u}}) &:= \{\tilde{g} \in L_1(\Omega, \mathcal{F}, \mathbb{P}) : \exists (\boldsymbol{x}, x_0) \in \mathbb{R}^d \times \mathbb{R} \text{ such that } \tilde{g}(\omega) \\
&= \tilde{g}_{\boldsymbol{x}, x_0}(\omega) := \boldsymbol{x}'\tilde{\boldsymbol{u}}(\omega) + x_0\} \text{ and} \\
\mathcal{V}_0(\tilde{\boldsymbol{u}}) &:= \{\tilde{g} \in L_1(\Omega, \mathcal{F}, \mathbb{P}) : \exists \boldsymbol{x} \in \mathbb{R}^d \text{ such that } \tilde{g}(\omega) = \tilde{g}_{\boldsymbol{x}}(\omega) \\
&:= \boldsymbol{x}'\tilde{\boldsymbol{u}}(\omega)\}.
\end{aligned}$$

From now on we assume that the random vector $\tilde{\boldsymbol{u}} \in L_1^d(\Omega, \mathcal{F}, \mathbb{P})$ is fixed and we simplify the notation to $\mathcal{V}$ and $\mathcal{V}_0$, to which we colloquially refer to as the spaces of *affine* and *linear random variables*.

An advantage of restricting our attention to $\mathcal{V}$ or to $\mathcal{V}_o$ is that the effect of a coherent risk measure on such random variables can be interpreted using the language of robust optimization as follows. Let ρ be a risk measure satisfying (C1)–(C3) and let $\mathcal{J} \subseteq \Delta$ be a convex and weakly* compact set satisfying (2). Then, for any $\tilde{g}_{\boldsymbol{x}, x_0} \in \mathcal{V}$ we have

$$\rho(\tilde{g}_{\boldsymbol{x}, x_0}) = \sup_{f \in \mathcal{J}} \int_\Omega (\boldsymbol{x}'\tilde{\boldsymbol{u}}(\omega) + x_0) f(\omega) d\mathbb{P}(\omega) = x_0 + \sup_{\boldsymbol{u} \in \mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{u}, \tag{3}$$

where

$$\mathcal{U}(\rho) := \left\{ \int_\Omega \tilde{\boldsymbol{u}}(\omega) f(\omega) d\mathbb{P}(\omega) : f \in \mathcal{J} \right\} \subseteq \mathbb{R}^d.$$

We have that $\mathcal{U}(\rho)$ is the image of convex and weakly* compact set $\mathcal{J}$ under $M : L_\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^d$ given by $M_i(f) := \int_\Omega \tilde{\boldsymbol{u}}_i(\omega) f(\omega) d\mathbb{P}(\omega)$. Because $\tilde{\boldsymbol{u}}_i \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ we have that M is linear and weakly* continuous and hence $\mathcal{U}(\rho)$ is a compact and convex set satisfying $\mathcal{U}(\rho) \subseteq \text{aff}(\text{supp}(\tilde{\boldsymbol{u}}))$, where $\text{supp}(\tilde{\boldsymbol{u}})$ is the support of $\tilde{\boldsymbol{u}}$. If ρ additionally satisfies (C4) then $\mathcal{J} \subseteq \Delta_+$, and we have that $\mathcal{U}(\rho) \subseteq \overline{\text{conv}}(\text{supp}(\tilde{\boldsymbol{u}}))$. In the robust optimization literature this set $\mathcal{U}(\rho)$ is usually denoted the robust uncertainty set and the following well known theorem (e.g. Theorem 4 of Natarajan et al. (2009)) states that its existence essentially characterizes coherent risk measures over $\mathcal{V}$.

Theorem 2.3. *$\rho : \mathcal{V} \to \mathbb{R}$ satisfies properties (C1)–(C3) of Definition 2.1 over $\mathcal{V}$ if and only if there exists a closed convex set $\mathcal{U} \subseteq \text{aff}(\text{supp}(\tilde{\boldsymbol{u}}))$ such that*

$$\rho(\tilde{g}_{\boldsymbol{x}, x_0}) = x_0 + \sup_{\boldsymbol{u} \in \mathcal{U}} \boldsymbol{x}'\boldsymbol{u} \tag{4}$$

for every $\tilde{g}_{\boldsymbol{x}, x_0} \in \mathcal{V}$. In such case we have that the set $\mathcal{U}$ satisfying (4) is unique and equal to

$$\mathcal{U}(\rho) := \left\{ \boldsymbol{u} \in \mathbb{R}^d \ : \ \boldsymbol{x}'\boldsymbol{u} \le \rho(\tilde{g}_{\boldsymbol{x}}) \quad \forall \boldsymbol{x} \in \mathbb{R}^n \right\}. \tag{5}$$

Furthermore, ρ additionally satisfies property (C4) if and only if $\mathcal{U}(\rho) \subseteq \overline{\text{conv}}(\text{supp}(\tilde{\boldsymbol{u}}))$.

Proof. For the forward implication of the first equivalence note that because ρ is a real valued function that is convex and positive homogeneous over $\mathcal{V}_0 \subseteq \mathcal{V}$, we have that $\rho(\boldsymbol{x}'\tilde{\boldsymbol{u}})$ is a continuous sublinear function of $\boldsymbol{x}$. Then $\rho(\boldsymbol{x}'\tilde{\boldsymbol{u}}) = \sup_{\boldsymbol{u} \in \mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{u}$ for the closed convex set $\mathcal{U}(\rho)$ defined in (5) (Theorem C-3.1.1 of Hiriart-Urruty and Lemaréchal (2001)). Now, let $\boldsymbol{u}_0 \in \mathbb{R}^d$ and $L \subseteq \mathbb{R}^d$ be a linear subspace such that $\text{aff}(\text{supp}(\tilde{\boldsymbol{u}})) = L + \boldsymbol{u}_0$. If $x \in L^\perp$ then $\boldsymbol{x}'\tilde{\boldsymbol{u}} = \boldsymbol{x}'\boldsymbol{u}_0$ *a.s.* and hence $\rho(\boldsymbol{x}'\tilde{\boldsymbol{u}}) = \boldsymbol{x}'\boldsymbol{u}_0$. Then, by (5) we have that

$$\begin{aligned}
\mathcal{U}(\rho) &\subseteq \left\{ \boldsymbol{u} \in \mathbb{R}^d \ : \ \boldsymbol{x}'\boldsymbol{u} \le \boldsymbol{x}'\boldsymbol{u}_0 \quad \forall \boldsymbol{x} \in L^\perp \right\} \\
&= \left\{ \boldsymbol{u} \in \mathbb{R}^d \ : \ \boldsymbol{x}'\boldsymbol{u} = \boldsymbol{x}'\boldsymbol{u}_0 \quad \forall \boldsymbol{x} \in L^\perp \right\} \\
&= L + \boldsymbol{u}_0 = \text{aff}(\text{supp}(\tilde{\boldsymbol{u}})).
\end{aligned}$$

The implication then follows from the translation equivariance property. The reverse implication is straightforward.

For the forward implication of the second equivalence note that $\mathcal{U}(\rho) \subseteq \overline{\text{conv}}(\text{supp}(\tilde{\boldsymbol{u}}))$ is equivalent to $\sup_{\boldsymbol{u} \in \mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{u} \le \sup_{\boldsymbol{u} \in \text{supp}(\mathbb{P})} \boldsymbol{x}'\boldsymbol{u}$ for all $\boldsymbol{x}$. If $\sup_{\boldsymbol{u} \in \text{supp}(\tilde{\boldsymbol{u}})} \boldsymbol{x}'\boldsymbol{u} = \infty$ this last inequality holds automatically. If not, by translation equivariance and positive homogeneity of ρ we have $\rho(\sup_{\boldsymbol{u} \in \text{supp}(\tilde{\boldsymbol{u}})} \boldsymbol{x}'\boldsymbol{u}) = \sup_{\boldsymbol{u} \in \text{supp}(\tilde{\boldsymbol{u}})} \boldsymbol{x}'\boldsymbol{u}$. Then, because of $\boldsymbol{x}'\tilde{\boldsymbol{u}} \le \sup_{\boldsymbol{u} \in \text{supp}(\tilde{\boldsymbol{u}})} \boldsymbol{x}'\boldsymbol{u}$ and monotonicity of ρ we have

$$\sup_{\boldsymbol{u} \in \mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{u} = \rho(\boldsymbol{x}'\tilde{\boldsymbol{u}}) \le \rho\left(\sup_{\boldsymbol{u} \in \text{supp}(\tilde{\boldsymbol{u}})} \boldsymbol{x}'\boldsymbol{u} \right) = \sup_{\boldsymbol{u} \in \text{supp}(\tilde{\boldsymbol{u}})} \boldsymbol{x}'\boldsymbol{u}.$$

For the reverse implication note that if $\mathcal{U}(\rho) \subseteq \overline{\text{conv}}(\text{supp}(\tilde{\boldsymbol{u}}))$ and $\boldsymbol{x}'\tilde{\boldsymbol{u}}(\omega) + x_0 \le 0$ *a.s.* then $\rho(\boldsymbol{x}'\tilde{\boldsymbol{u}} + x_0) = x_0 + \sup_{\boldsymbol{u} \in \mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{u} \le x_0 +$

$\sup_{\boldsymbol{u}\in\text{supp}(\widetilde{\boldsymbol{u}})} \boldsymbol{x}'\boldsymbol{u} \le 0$. Together with sub-additivity of ρ this implies that if $\widetilde{g}_{\boldsymbol{x},x_0} \le \widetilde{g}_{\boldsymbol{y},y_0}$ then $\rho(\widetilde{g}_{\boldsymbol{x},x_0}) \le \rho(\widetilde{g}_{\boldsymbol{x}-\boldsymbol{y},x_0-y_0}) + \rho(\widetilde{g}_{\boldsymbol{y},y_0}) \le \rho(\widetilde{g}_{\boldsymbol{y},y_0})$. □

Note that in the proof of Theorem 2.3 necessity of $\mathcal{U}(\rho) \subseteq \overline{\text{conv}}(\text{supp}(\widetilde{\boldsymbol{u}}))$ was because of dominance between a constant $(\sup_{\boldsymbol{u}\in\text{supp}(\widetilde{\boldsymbol{u}})} \boldsymbol{x}'\boldsymbol{u})$ and a linear $(\boldsymbol{x}'\widetilde{\boldsymbol{u}})$ random variable. In Section 4 we will see that this condition can sometimes be eliminated when we only consider linear random variables (i.e. if we restrict ourselves to $\mathcal{V}_0$).

Remark 2.1. It is also interesting to note the difference between the characterization of coherent risk measures over $L_1(\Omega,\mathcal{F},\mathbb{P})$ given by Theorem 2.2 and the characterization of coherent risk measures over subspace $\mathcal{V}$ of $L_1(\Omega,\mathcal{F},\mathbb{P})$ given by Theorem 2.3. While any closed convex set $\mathcal{J} \subseteq \Delta_+$ induces a convex uncertainty set $\mathcal{U}(\mathcal{J}) := \{\int_\Omega \widetilde{\boldsymbol{u}}(\omega) f(\omega) d\mathbb{P}(\omega) : f \in \mathcal{J}\} \subseteq \overline{\text{conv}}(\text{supp}(\widetilde{\boldsymbol{u}}))$, the converse does not always hold. For instance, if we let $\widetilde{\boldsymbol{u}}$ be uniformly distributed on a compact convex set C and $\boldsymbol{u}^0$ be an extreme point of C, we have that $\mathcal{U} = \{\boldsymbol{u}^0\}$ is a convex uncertainty set that will induce a coherent risk measure over $\mathcal{V}$ through (4). However, there is no $f \in \Delta_+$ such that $\int_\Omega \widetilde{\boldsymbol{u}}(\omega) f(\omega) d\mathbb{P}(\omega) = \boldsymbol{\omega}^0$ and hence by Theorem 2.2 and (3) there cannot be a coherent risk measure over $L_1(\Omega,\mathcal{F},\mathbb{P})$ that coincides with this measure in $\mathcal{V}$. Therefore the set of coherent risk measures over $\mathcal{V}$ is larger than those over $L_1(\Omega,\mathcal{F},\mathbb{P})$. Note that, if we restrict ourselves to finite probability distributions, it is not too hard to prove that this difference vanishes. However, in Section 3, we show that this difference no longer vanishes for the so-called *distortion risk measures*.

2.3. Distortion risk measures

Definition 2.4. A coherent risk measure $\rho : L_1(\Omega,\mathcal{F},\mathbb{P}) \to \mathbb{R}$ is a distortion or spectral risk measure if it satisfies the following additional properties.

(D1) Comonotonicity: $\rho(\widetilde{g}_1 + \widetilde{g}_2) = \rho(\widetilde{g}_1) + \rho(\widetilde{g}_2)$ for all $\widetilde{g}_1, \widetilde{g}_2$ such that $(\widetilde{g}_1(\omega_1) - \widetilde{g}_1(\omega_2))(\widetilde{g}_2(\omega_1) - \widetilde{g}_2(\omega_2)) \ge 0, \quad \forall \omega_1, \omega_2 \in \Omega$.

(D2) Law invariance: $\rho(\widetilde{g}_1) = \rho(\widetilde{g}_2)$ for all $\widetilde{g}_1, \widetilde{g}_2$ that have the same distribution.

Example 2.1. One of the most well known distortion risk measures is the *Conditional Value-at-Risk* which is given by $CVaR_\delta(\widetilde{g}) := \inf_{t\in\mathbb{R}} \left\{ t + \frac{1}{\delta}\mathbb{E}[(\widetilde{g} - t)^+] \right\}$.

While some characterizations of distortion risk measures are given for more general probability distributions (e.g. see Shapiro 2013 and Pichler & Shapiro, 2013), we now concentrate on the uniform probability distribution with finite support. Results in this section can be found in, or are direct corollaries of results in Bertsimas and Brown (2009).

For uniform discrete distributions we let $\text{supp}(\mathbb{P}) = \{\omega_i : i \in [N]\} \subseteq \Omega$ for which $\mathbb{P}(\{\omega_i\}) = \frac{1}{N}$ for all $i \in [N]$. In this setting we assume $\Omega = \{\omega_i : i \in [N]\}$ and that $\mathcal{F}$ is the σ-algebra of all subsets of Ω. Furthermore, under these assumptions sets Δ and Δ_+ defined in (1) become

$$\Delta^N := \left\{ \boldsymbol{q} \in \mathbb{R}^N : \sum_{i=1}^N q_i = 1 \right\} \quad \text{and}$$

$$\Delta^N_+ := \left\{ \boldsymbol{q} \in \mathbb{R}^N : \sum_{i=1}^N q_i = 1, \quad q_i \ge 0 \quad \forall i \in [N] \right\}.$$

With this notation, every random variable $\widetilde{g} \in L_1(\Omega,\mathcal{F},\mathbb{P})$ is representable by means of a vector $\boldsymbol{g} \in \mathbb{R}^N$ where $g_i := \widetilde{g}(\omega_i)$ for all $i \in [N]$. Indeed, for finite probability spaces it is somewhat meaningless to consider $L_1(\Omega,\mathcal{F},\mathbb{P})$ as all $L_p(\Omega,\mathcal{F},\mathbb{P})$ are trivially equal to space of functions from Ω to $\mathbb{R}$. However, we continue using this notation to have a consistent way of distinguishing risk measures that are defined over arbitrary functions of Ω from those that are only defined over $\mathcal{V}$ or $\mathcal{V}_o$.

Theorem 2.5. *If $\mathbb{P}$ is a finite uniform distribution over $\Omega = \{\omega_i : i \in [N]\}$, ρ satisfies (C1)–(C3), (D1)–(D2) over $L_1(\Omega,\mathcal{F},\mathbb{P})$ if and only if there exists $\boldsymbol{q} \in \Delta^N$ such that*

$$\rho(\widetilde{g}) = \max_{\sigma\in S_N} \sum_{i=1}^N q_{\sigma(i)} \boldsymbol{g}_i, \tag{6}$$

where S_N is the group of permutations of N elements. Furthermore, in this representation we can additionally choose $\boldsymbol{q} \in \widehat{\Delta}^N := \{\boldsymbol{q} \in \Delta^N : q_1 \ge \ldots \ge q_N\}$. Finally, ρ further satisfies (C4) over $L_1(\Omega,\mathcal{F},\mathbb{P})$ if and only if $\boldsymbol{q}$ additionally belongs to Δ^N_+ or $\widehat{\Delta}^N_+ := \{\boldsymbol{q} \in \widehat{\Delta}^N : q_N \ge 0\}$. In both cases we have that

$$\mathcal{U}(\rho) = \Pi_{\boldsymbol{q}}(\widetilde{\boldsymbol{u}}) := \text{conv}\left(\left\{ \sum_{i=1}^N q_{\sigma(i)} \boldsymbol{u}^i : \sigma \in S_N \right\} \right) \tag{7}$$

where $\boldsymbol{u}^i = \widetilde{\boldsymbol{u}}(\omega_i)$ for each $i \in [N]$.

For notational convenience we again drop the dependence of $\Pi_{\boldsymbol{q}}$ on $\widetilde{\boldsymbol{u}}$.

Example 2.2. Let $\delta \in [0,1]$ be such that $\delta N \in \mathbb{Z}_+$. Then $\mathcal{U}(CVaR_\delta) = \Pi_{\boldsymbol{h}^\delta}(\Omega)$ where

$$\boldsymbol{h}^\delta_j := \begin{cases} \frac{1}{\delta N} & j \le \delta N \\ 0 & \text{otherwise} \end{cases}. \tag{8}$$

3. Distortion risk measures for uniform, discrete random variables in $\mathcal{V}$ and $\mathcal{V}_o$

In this section we will prove that, even in the case of $\mathbb{P}$ being a finite uniform distribution, there exists distortion risk measures $\rho : \mathcal{V} \to \mathbb{R}$ that are not induced by any distortion risk measure $\rho' : L_1(\Omega,\mathcal{F},\mathbb{P}) \to \mathbb{R}$. For this, we will need some previous technical lemmas.

Lemma 3.1. *If $\mathbf{0} \in \text{ri}(\text{conv}(\text{supp}(\widetilde{\boldsymbol{u}})))$ then for any $\widetilde{v} \in \mathcal{V}_o$, $\widetilde{v} \ge 0$ a.s. implies $\widetilde{v} = 0$ a.s.*

Proof. Since $\widetilde{v} \in \mathcal{V}_o$, then $\exists \boldsymbol{x} \in \mathbb{R}^d$ such that $\widetilde{v}(\omega) = \boldsymbol{x}'\widetilde{\boldsymbol{u}}(\omega)$, $\forall \omega \in \Omega$. From this, $\widetilde{v} \ge 0$ a.s. implies that $\boldsymbol{x}'\widetilde{\boldsymbol{u}}(\omega) \ge 0$, $\forall \omega \in \text{supp}(\widetilde{\boldsymbol{u}})$. If $\boldsymbol{x} = \mathbf{0}$ the result is direct. By contradiction, assume that there exists $\boldsymbol{u}_o \in \text{supp}(\widetilde{\boldsymbol{u}})$ such that $\boldsymbol{x}'\boldsymbol{u}_o > 0$. Now, since $\mathbf{0} \in \text{ri}(\text{conv}(\text{supp}(\widetilde{\boldsymbol{u}})))$, there exists U a relatively open neighborhood of $\mathbf{0}$ within $\text{conv}(\text{supp}(\widetilde{\boldsymbol{u}}))$. However, because $\boldsymbol{u}_o \in \text{aff}(\text{conv}(\text{supp}(\widetilde{\boldsymbol{u}})))$, there exists $\varepsilon > 0$ such that $\varepsilon\boldsymbol{u}_o, -\varepsilon\boldsymbol{u}_o \in U \subseteq \text{conv}(\text{supp}(\widetilde{\boldsymbol{u}}))$. Then, there must exists $\boldsymbol{u}_1 \in \text{supp}(\widetilde{\boldsymbol{u}})$ such that $\boldsymbol{x}'\boldsymbol{u}_1$ has the same sign as $\boldsymbol{x}'(-\varepsilon\boldsymbol{u}_o) < 0$ which contradicts $\widetilde{v} \ge 0$ a.s.. □

Lemma 3.1 implies that if $\mathbf{0} \in \text{ri}(\text{conv}(\text{supp}(\widetilde{\boldsymbol{u}})))$, then, condition (C4) is moot for $\mathcal{V}_o$.

We then get the following refinement of Theorems 2.3 and 2.5 for linear random variables.

Corollary 3.2. *If $\mathbf{0} \in \text{ri}(\text{conv}(\text{supp}(\widetilde{\boldsymbol{u}})))$, then*

1. *$\rho : \mathcal{V}_0 \to \mathbb{R}$ satisfies (C1)–(C4) over $\mathcal{V}_0$ if and only if $\mathcal{U}(\rho) \subseteq \text{aff}(\text{supp}(\widetilde{\boldsymbol{u}}))$ and $\rho(\widetilde{g}_{\boldsymbol{x}}) = \sup_{\boldsymbol{u}\in\mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{u}$.*
2. *If $\mathcal{U}(\rho) = \Pi_{\boldsymbol{q}}$ for $\boldsymbol{q} \in \widehat{\Delta}^N$ and $\rho(\widetilde{g}_{\boldsymbol{x}}) = \sup_{\boldsymbol{u}\in\mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{u}$, then ρ satisfies (C1)–(C4), (D1)–(D2) over $\mathcal{V}_o$.*

Proof. For 1 note that the first part of the proof of Theorem 2.3 shows that $\rho : \mathcal{V} \to \mathbb{R}$ satisfies properties (C1)–(C3) over $\mathcal{V}_0$ if and only if $\mathcal{U}(\rho) \subseteq \text{aff}(\text{supp}(\widetilde{\boldsymbol{u}}))$ and $\rho(\widetilde{g}_{\boldsymbol{x}}) = \sup_{\boldsymbol{u}\in\mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{u}$. To obtain the first equivalence it only remains to show that if $\mathcal{U}(\rho) \subseteq \text{aff}(\text{supp}(\widetilde{\boldsymbol{u}}))$, then ρ satisfies property (C4) over $\mathcal{V}_0$. For this let $\widetilde{g}_i := \boldsymbol{x}^{i\prime}\widetilde{\boldsymbol{u}}(\omega)$ for $i \in \{1,2\}$ be such that $\widetilde{g}_1(\omega) \le \widetilde{g}_2(\omega)$ a.s. Then $(\boldsymbol{x}^2 - \boldsymbol{x}^1)'\widetilde{\boldsymbol{u}}(\omega) \ge 0$ a.s. and hence

by Lemma 3.1 we have $(\boldsymbol{x}^2 - \boldsymbol{x}^1)'\tilde{\boldsymbol{u}}(\omega) = 0$ *a.s.* Hence $\tilde{g}_1(\omega) = \tilde{g}_2(\omega)$ *a.s* which implies $\rho(\tilde{g}_1) = \rho(\tilde{g}_2)$ and condition (C4) holds.

Statement 2 follows similarly from Theorem 2.5 and Lemma 3.1. □

Using this corollary we can characterize inclusion relations between family of sets inducing coherent, or distortion risk measures for the case $\mathbf{0} \in \text{ri}\,(\text{conv}\,(\text{supp}\,(\tilde{\boldsymbol{u}})))$. For this, we introduce the following definitions:

Definition 3.3. Let $\mathbb{W}, \mathbb{V}, \mathbb{V}_o$ the set of coherent risk measures defined over $L_1(\Omega, \mathcal{F}, \mathbb{P}), \mathcal{V}, \mathcal{V}_o$ respectively; and let $\mathbb{W}^*, \mathbb{V}^*, \mathbb{V}_o^*$ the set of distortion risk measures defined over $L_1(\Omega, \mathcal{F}, \mathbb{P}), \mathcal{V}, \mathcal{V}_o$ respectively. We will denote $\mathcal{U}(\cdot)$ as the family of sets that induce all risk measures in a given set.

Note that, form the definitions above, we always have that $\mathcal{U}(\mathbb{H}^*) \subseteq \mathcal{U}(\mathbb{H})$ for any $\mathbb{H} \in \{\mathbb{W}, \mathbb{V}, \mathbb{V}_o\}$. With these definitions, we can write the following result:

Corollary 3.4. *If* $\mathbf{0} \in \text{ri}\,(\text{conv}\,(\text{supp}\,(\tilde{\boldsymbol{u}})))$, *then*

1. $\mathcal{U}(\mathbb{V}_o) = \{\mathcal{U} \subseteq \text{aff}(\text{supp}(\tilde{\boldsymbol{u}})) : \mathcal{U}$ *is closed and convex*$\}$ *and*
$$\mathcal{U}(\mathbb{W}) \subseteq \mathcal{U}(\mathbb{V}) = \left\{\mathcal{U} \in \mathcal{U}(\mathbb{V}_o) \,:\, \mathcal{U} \subseteq \overline{\text{conv}}\,(\text{supp}\,(\tilde{\boldsymbol{u}}))\right\}.$$
Hence, $\mathcal{U}(\mathbb{V})$ *and* $\mathcal{U}(\mathbb{W})$ *can be strictly contained in* $\mathcal{U}(\mathbb{V}_o)$.
2. *If* $\mathbb{P}$ *is a finite uniform distribution then* $\{\Pi_{\boldsymbol{q}} \,:\, \boldsymbol{q} \in \widehat{\Delta}^N\} \subseteq \mathcal{U}(\mathbb{V}_o^*)$ *and*
$$\mathcal{U}(\mathbb{W}^*) = \left\{\Pi_{\boldsymbol{q}} \,:\, \boldsymbol{q} \in \widehat{\Delta}_+^N\right\} \subseteq \mathcal{U}(\mathbb{V}^*) \subseteq \mathcal{U}(\mathbb{V}).$$
Hence, $\mathcal{U}(\mathbb{V}^*)$ *and* $\mathcal{U}(\mathbb{W}^*)$ *can be strictly contained in* $\mathcal{U}(\mathbb{V}_o^*)$.

Proof. The characterizations are direct from Corollary 3.2 and Theorems 2.2, 2.3 and 2.5. In particular, the potential lack of equality between $\mathcal{U}(\mathbb{W})$ and $\mathcal{U}(\mathbb{V})$ comes from Remark 2.1. For the first potential strict containment it suffices to find a closed convex set $\mathcal{U} \subseteq \text{aff}(\text{supp}(\tilde{\boldsymbol{u}}))$ such that $\mathcal{U} \nsubseteq \overline{\text{conv}}\,(\text{supp}\,(\tilde{\boldsymbol{u}}))$. For the second it suffices to find $\boldsymbol{q} \in \widehat{\Delta}^N$ such that $\Pi_{\boldsymbol{q}} \nsubseteq \overline{\text{conv}}\,(\text{supp}\,(\tilde{\boldsymbol{u}}))$. □

Corollary 3.4 shows that, when $\mathbf{0} \in \text{ri}\,(\text{conv}\,(\text{supp}\,(\tilde{\boldsymbol{u}})))$, there are *somewhat reasonable* uncertainty sets for random variables in $\mathcal{V}_o$ that are not induced by coherent and distortion risk measures over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ or $\mathcal{V}$. However, those sets include points outside $\overline{\text{conv}}\,(\text{supp}\,(\tilde{\boldsymbol{u}}))$. Remembering that the risk measure with $\mathcal{U}(\rho) = \overline{\text{conv}}\,(\text{supp}\,(\tilde{\boldsymbol{u}}))$ corresponds to the worst case over all possible realizations of the random variable, we conclude that a risk measure with ρ' with $\overline{\text{conv}}\,(\text{supp}\,(\tilde{\boldsymbol{u}})) \subsetneq \mathcal{U}(\rho)$ would be clearly over-conservative. Considering an uncertainty set that neither contains nor is contained in $\overline{\text{conv}}\,(\text{supp}\,(\tilde{\boldsymbol{u}}))$ is a bit more reasonable, but it still somewhat strange to include points outside $\overline{\text{conv}}\,(\text{supp}\,(\tilde{\boldsymbol{u}}))$ in the risk evaluation. To avoid this philosophical issue we now concentrate on the following result, which holds irrespective of the assumption $\mathbf{0} \in \text{ri}\,(\text{conv}\,(\text{supp}\,(\tilde{\boldsymbol{u}})))$.

Corollary 3.5. *Let* $\mathbb{P}$ *be a finite uniform distribution for which* $\mathbf{0}$ *is not necessarily contained in* $\text{ri}\,(\text{conv}\,(\text{supp}\,(\tilde{\boldsymbol{u}})))$. *Then*

$$\begin{aligned}\mathcal{U}(\mathbb{W}^*) = \left\{\Pi_{\boldsymbol{q}} \,:\, \boldsymbol{q} \in \widehat{\Delta}_+^N\right\} &\subseteq \{\Pi_{\boldsymbol{q}} \,:\, \boldsymbol{q} \in \widehat{\Delta}^N,\ \Pi_{\boldsymbol{q}} \subseteq \overline{\text{conv}}\,(\text{supp}\,(\tilde{\boldsymbol{u}}))\} \\ &\subseteq \mathcal{U}(\mathbb{V}_o^*), \mathcal{U}(\mathbb{V}^*) \end{aligned} \tag{9}$$

Proof. Direct from Theorems 2.3 and 2.5. □

From this corollary, the existence of an actually reasonable uncertainty set for random variables in $\mathcal{V}_o$ and $\mathcal{V}$ (i.e. one that is contained in $\overline{\text{conv}}\,(\text{supp}\,(\tilde{\boldsymbol{u}}))$ and induces measures that satisfy (C1)–(C4), (D1)–(D2)) that is not induced by coherent and distortion risk measures over $L_1(\Omega, \mathcal{F}, \mathbb{P})$, reduces to the possibility of a strict containment in (9). We now show that the first containment can indeed be strict.

To show this strict containment we need to find $\boldsymbol{q} \in \widehat{\Delta}^N$ such that $\Pi_{\boldsymbol{q}} \subseteq \overline{\text{conv}}\,(\text{supp}\,(\tilde{\boldsymbol{u}}))$ and for which there is no $\boldsymbol{r} \in \widehat{\Delta}_+^N$ such that $\Pi_{\boldsymbol{q}} = \Pi_{\boldsymbol{r}}$. For this, note that, for any $\boldsymbol{r}$, by definition, we have that $\Pi_{\boldsymbol{r}} = \text{conv}(\{\sum_{i=1}^N \boldsymbol{r}_{\sigma(i)}\boldsymbol{u}^i : \sigma \in S_N\})$; but each term $\sum_{i=1}^N \boldsymbol{r}_{\sigma(i)}\boldsymbol{u}^i$ can be rewritten as $M \cdot P_\sigma \cdot \boldsymbol{r}$; where $M = (\boldsymbol{u}^1 | \ldots | \boldsymbol{u}^N)$ and P_σ is a permutation matrix (i.e. $\boldsymbol{e}'P_\sigma = \boldsymbol{e}'$ and $P_\sigma \boldsymbol{e} = \boldsymbol{e}$) that depends on σ. Also, note that if we know all vertices $\{\boldsymbol{v}^k\}_{k\in[m]}$ of $\Pi_{\boldsymbol{q}}$, then, for $\Pi_{\boldsymbol{q}}$ to be equal to $\Pi_{\boldsymbol{r}}$ for some $\boldsymbol{r}$, we need at least that each $\boldsymbol{v}^k$ is the image of some permutation matrix P_σ, i.e. $\exists P^k$ permutation matrix such that $\boldsymbol{v}^k = M \cdot P^k \cdot \boldsymbol{r}$. Proposition 3.6 expresses this idea as an optimization problem.

Proposition 3.6. *For a given* $\boldsymbol{q}$, *let* $\{\boldsymbol{v}^k\}_{k\in[m]}$ *be the set of extreme points of* $\Pi_{\boldsymbol{q}}$. *If there exists* $\boldsymbol{r} \in \widehat{\Delta}_+^N$ *such that* $\Pi_{\boldsymbol{q}} = \Pi_{\boldsymbol{r}}$, *then, the following optimization problem has optimal value zero.*

$$\min \sum_{k\in[m]} ||\boldsymbol{s}^k||_1 \tag{10a}$$
$$s.t.\ \boldsymbol{v}^k = M \cdot P^k \cdot \boldsymbol{r} + \boldsymbol{s}^k \quad \forall k \in [m] \tag{10b}$$
$$\boldsymbol{e}'P^k = \boldsymbol{e}' \quad \forall k \in [m] \tag{10c}$$
$$P^k\boldsymbol{e} = \boldsymbol{e} \quad \forall k \in [m] \tag{10d}$$
$$\boldsymbol{r}_i \geq \boldsymbol{r}_{i+1} \quad \forall i \in [N] \tag{10e}$$
$$\boldsymbol{e}'\boldsymbol{r} = 1 \tag{10f}$$
$$P^k \in \{0,1\}^{N\times N} \quad \forall k \in [m] \tag{10g}$$
$$s^k \in \mathbb{R}^d \quad \forall k \in [m] \tag{10h}$$
$$r \in \mathbb{R}_+^N \tag{10i}$$

Proof. Problem (10) has as variables the vectors $\boldsymbol{s}^k$, $\boldsymbol{r}$ and binary matrices P^k for $k \in [m]$. Conditions (10c), (10d) and (10g), ensure that P^k is a permutation matrix; conditions (10e), (10f) and (10i) ensure that $\boldsymbol{r} \in \widehat{\Delta}_+^N$, while condition (10b) just says that each $\boldsymbol{v}^k = M \cdot P^k \cdot \boldsymbol{r} + \boldsymbol{s}^k$. To finish, just note that the objective function (10a) can only be zero when all $\boldsymbol{s}^k$ are zero, and thus ensuring that each $\boldsymbol{v}^k$ corresponds to one of the points generating the set $\Pi_{\boldsymbol{r}}$. □

Note that in the previous result, (10) having optimal value 0 is not a sufficient condition for $\Pi_{\boldsymbol{q}} = \Pi_{\boldsymbol{r}}$, since it only ensures that $\Pi_{\boldsymbol{q}} \subseteq \Pi_{\boldsymbol{r}}$. However, we show that for a particular $\boldsymbol{q}$, problem (10) has non-zero optimal value.

Lemma 3.7. *There exist* $\boldsymbol{q} \in \widehat{\Delta}^N$ *such that* $\Pi_{\boldsymbol{q}} \subseteq \overline{\text{conv}}(\Omega)$ *and* $\Pi_{\boldsymbol{q}} \neq \Pi_{\boldsymbol{q}'}$ *for all* $q \in \widehat{\Delta}_+^N$.

Proof. Let $d = 2$, $N = 5$, $\text{supp}\,(\tilde{\boldsymbol{u}}) = \{(8600, 5000), (5700, 8100), (1300, 9900), (-9600, 3000), (8500, -5200)\}$ and $\boldsymbol{q} = (27/100, 27/100, 27/100, 27/100, -2/25) \in \widehat{\Delta}^N$. Using a symbolic computation software it is checked that $\text{ext}(\Pi_{\boldsymbol{q}}) = \{(905, 3866), (1920, 2781), (3460, 2151), (7275, 4566), (940, 7436)\}$ and $\Pi_{\boldsymbol{q}} \subseteq \overline{\text{conv}}(\Omega)$. Furthermore, using the exact MIP solvers developed in Espinoza (2006); Cook, Koch, Steffy, and Wolter (2011), we were able to computationally prove that the optimal objective value of (10), for this data, is greater than or equal to 1000.

Note that the exact MIP solvers from Espinoza (2006); Cook et al. (2011) can only solve linear MIP problems and problem (10) is a nonlinear MIP problem. However, (10) can easily be transformed into a linear MIP as follows. The first step is to linearize the products between P^k and $\boldsymbol{r}$ in (10b) using a standard technique (e.g. Adams and Sherali (1986)). For this we introduce matrix variables $\boldsymbol{G}^k \in [0,1]^{N\times N}$ for all $k \in [m]$. This matrix will be such that $\boldsymbol{G}^k_{i,j} = P^k_{i,j} \cdot \boldsymbol{r}_j$ for all i, j and k, $\boldsymbol{r}_j \in [0,1]$ and $P^k_{i,j} \in \{0,1\}$. To achieve this we add the set of linear inequalities given by

$$\begin{aligned}
&\boldsymbol{G}^k \leq \boldsymbol{P}^k && \forall k \in [m] \\
&\boldsymbol{G}^k_{i,j} \leq \boldsymbol{r}_j && \forall i, j \in [N],\ k \in [m] \\
&\boldsymbol{P}^k_{i,j} + \boldsymbol{r}_j \leq \boldsymbol{G}^k_{i,j} + 1 && \forall i, j \in [N], k \in [m].
\end{aligned}$$

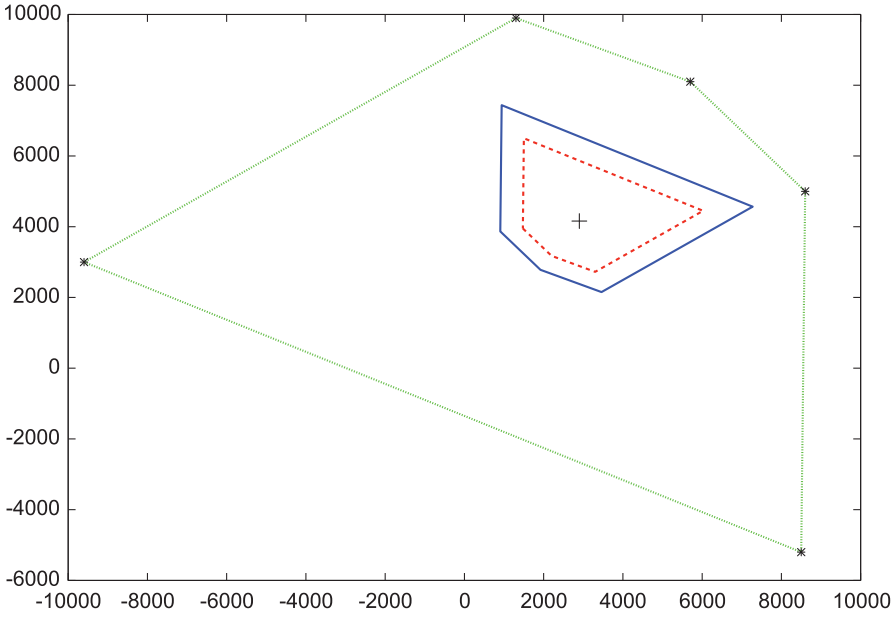

Fig. 1. Uncertainty sets from Lemma 3.7.

We then simply replace $\boldsymbol{P}^k \cdot \boldsymbol{r}$ in (10b) with $\boldsymbol{G}^k \boldsymbol{e}$. Finally, to linearize the objective function we introduce variables $\boldsymbol{sp}^k, \boldsymbol{sm}^k \in \mathbb{R}^d_+$ for all $k \in [m]$, replace $\boldsymbol{s}^k$ in (10b) with $\boldsymbol{sp}^k - \boldsymbol{sm}^k$ and replace the objective function with $\sum_{k\in[m]} \boldsymbol{sp}^k + \boldsymbol{sm}^k$. □

Corollary 3.5 and Lemma 3.7 show that there are indeed reasonable uncertainty sets (distortion risk measures) for random variables in $\mathcal{V}$ and $\mathcal{V}_o$ that are not induced by coherent and distortion risk measures over $L_1(\Omega, \mathcal{F}, \mathbb{P})$. However, while theoretically interesting, the conditions for constructing or detecting these sets can be highly intractable. For this reason, in the next section we present a more practical representation of the uncertainty sets $\Pi_{\boldsymbol{q}}$ for $\boldsymbol{q} \in \widehat{\Delta}^N$.

4. Epsilon scaling of a risk measure

From Lemma 3.7 we know that there exists risk measures represented by $\boldsymbol{q} \in \widehat{\Delta}^N$ whose uncertainty sets do not coincide with any risk measure in $\widehat{\Delta}^N_+$. However, it is possible to give a different characterization of these uncertainty sets, providing a natural geometrical interpretation of these measures.

Consider for example the finite uniform probability over the $N = 5$ points in $\operatorname{supp}(\widetilde{\boldsymbol{u}}) = \{\boldsymbol{u}^i\}_{i=1}^n = \{(8600, 5000), (5700, 8100), (1300, 9900), (-9600, 3000), (8500, -5200)\}$ and

$$\boldsymbol{q} = (27/100,\ 27/100,\ 27/100,\ 27/100,\ -2/25) \in \widehat{\Delta}^N$$

used in Lemma 3.7. We can check that $\boldsymbol{q} = \varepsilon \boldsymbol{q}' + (1-\varepsilon)\boldsymbol{e}_N$ for $\boldsymbol{q}' = (1/4,\ 1/4,\ 1/4,\ 1/4,\ 0) \in \widehat{\Delta}^N_+$ and $\varepsilon = 7/5$. Note that it is not a convex combination but an affine combination, because $\varepsilon > 1$. Fig. 1 shows $\Pi_{\boldsymbol{q}}$ in solid blue, $\Pi_{\boldsymbol{q}'}$ in dashed red and $\overline{\operatorname{conv}}(\operatorname{supp}(\widetilde{\boldsymbol{u}}))$ in dotted green. The figure also shows $\operatorname{supp}(\widetilde{\boldsymbol{u}})$ as asterisks and $\overline{\boldsymbol{u}} := \frac{1}{N}\sum_{i=1}^n \boldsymbol{u}^i$ as a plus sign. We can see from the figure that $\Pi_{\boldsymbol{q}}$ is an expansion of $\Pi_{\boldsymbol{q}'}$ around the mean $\overline{\boldsymbol{u}}$ that is still contained in $\overline{\operatorname{conv}}(\operatorname{supp}(\widetilde{\boldsymbol{u}}))$. In this section we show that this figure is representative of all $\boldsymbol{q} \in \widehat{\Delta}^N$ in that for such vectors $\Pi_{\boldsymbol{q}}$ is always an expansion of $\Pi_{\boldsymbol{q}'}$ for some $\boldsymbol{q}' \in \widehat{\Delta}^N_+$. This implies that the risk measures associated with elements in $\widehat{\Delta}^N$ are always an affine combination of a distortion risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ and the expected value. When $\varepsilon \in [0, 1]$, this convex combination is a well known modification of a risk measure (e.g. see Lagos, Espinoza, Moreno, and Amaya (2011) and Eq. (6.68) in Shapiro et al. (2009)), note however that in this case, ε is not restricted to be within $[0, 1]$, as it can take values above 1. Hence the associated measure is an affine combination of a distortion risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ and the expected value. Because the uncertainty sets associated to these measures are scalings of the traditional sets, we denote these new measures as *epsilon scalings*.

Definition 4.1. For a given risk measure $\rho : L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ and $\varepsilon \geq 0$ let the *epsilon scaling* of the measure be $\widehat{\rho}_\varepsilon(\widetilde{v}) := \varepsilon\, \rho(\widetilde{v}) + (1-\varepsilon)\, \mathbb{E}[\widetilde{v}]$.

It is straightforward to show that if ρ is a distortion risk measure over $\mathcal{S} \subseteq L_1(\Omega, \mathcal{F}, \mathbb{P})$ then, for any $\varepsilon \in [0, 1]$, $\widehat{\rho}_\varepsilon$ is also a distortion risk measure over $\mathcal{S}$. It is also easy to see that for $\varepsilon \geq 0$ the only property that $\widehat{\rho}_\varepsilon$ may fail to inherit is monotonicity. Fortunately, for $\mathcal{S} = \mathcal{V}$ or $\mathcal{S} = \mathcal{V}_o$ we can give simple conditions for $\widehat{\rho}_\varepsilon$ to be a distortion risk measure. To give these conditions note that the uncertainty set associated with $\widehat{\rho}_\varepsilon$ is $\mathcal{U}(\widehat{\rho}_\varepsilon) = \overline{\boldsymbol{u}} + \varepsilon(\mathcal{U}(\rho) - \overline{\boldsymbol{u}})$ where $\overline{\boldsymbol{u}} := \mathbb{E}[\widetilde{\boldsymbol{u}}]$.

Proposition 4.2. *Let $\varepsilon \geq 0$, ρ be a distortion risk measure over $\mathcal{V}$ and $\mathbb{P}$ be an arbitrary distribution.*

1. *If $\overline{\boldsymbol{u}} + \varepsilon(\mathcal{U}(\rho) - \overline{\boldsymbol{u}}) \subseteq \overline{\operatorname{conv}}(\operatorname{supp}(\widetilde{\boldsymbol{u}}))$, then $\widehat{\rho}_\varepsilon$ is a distortion risk measure over $\mathcal{V}$ and $\mathcal{V}_o$.*
2. *If $\boldsymbol{0} \in \operatorname{ri}(\operatorname{conv}(\operatorname{supp}(\widetilde{\boldsymbol{u}})))$, then $\widehat{\rho}_\varepsilon$ is a distortion risk measure over $\mathcal{V}_o$ even if $\overline{\boldsymbol{u}} + \varepsilon(\mathcal{U}(\rho) - \overline{\boldsymbol{u}}) \nsubseteq \overline{\operatorname{conv}}(\operatorname{supp}(\widetilde{\boldsymbol{u}}))$.*

Proof. Direct from Lemma 3.1, Theorem 2.3 and the preservation of (D1) and (D2) under linear combinations. □

If we restrict to finite uniform distributions we can show that epsilon scalings precisely correspond to the uncertainty sets associated with elements in $\widehat{\Delta}^N$.

Proposition 4.3. *If $\mathbb{P}$ is a finite uniform distributions, then $\{\Pi_{\boldsymbol{q}} : \boldsymbol{q} \in \widehat{\Delta}^N\} = \{\mathcal{U}(\widehat{\rho}_\varepsilon) : \varepsilon \geq 0 \text{ and } \rho \in \mathbb{W}^*\}$.*

Proof. For $\boldsymbol{q} \in \widehat{\Delta}^N \setminus \widehat{\Delta}^N_+$ let $\varepsilon := 1 - N q_N > 0$ and $\boldsymbol{q}' := \frac{1}{\varepsilon}(\boldsymbol{q} + (\varepsilon - 1)\boldsymbol{e}_N)$. Then $\boldsymbol{q} = \varepsilon \boldsymbol{q}' + (1-\varepsilon)\boldsymbol{e}_N$, $\varepsilon \geq 0$ and $\boldsymbol{q}' \in \widehat{\Delta}^N_+$. The result then follows from Corollary 3.4. □

5. Computational stability of epsilon scalings

In this section we present a computational example that shows that epsilon scalings seem to be less susceptible to estimation errors when approximated using samples. The need for such estimations is

common in applications (e.g. Lagos et al. (2011); Vielma, Espinoza, and Moreno (2009)) and, unfortunately, risk measures such as the *Conditional Value-at-Risk* (*CVaR*) measure have been shown to be highly susceptible to estimation errors in this setting (Lim et al., 2011). For this reason we study how using the epsilon scaling of *CVaR* could help alleviate these estimation errors. Following an approach similar to that in Lim et al. (2011) we consider a simple portfolio optimization problem, in which we have d possible assets we want to invest over a single time period, and we have to decide what proportion of our capital we will invest in each of the assets. Every asset i has a return $r_i \in [-1, \infty)$, such that if we initially invested C_i on i then at the end of the period we will have $C_i(1 + r_i)$. When the vector $\boldsymbol{r} := (r_1, \ldots, r_d)'$ of returns is known this problem is formulated as $\max\{\boldsymbol{x}'\boldsymbol{r} : \boldsymbol{x}'\boldsymbol{e} = 1, \boldsymbol{x} \geq \boldsymbol{0}\}$. Naturally the vector of returns $\boldsymbol{r}$ is subject to uncertainty, hence it is necessary to adopt some decision scheme that considers the risk inherent to the problem. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\widetilde{\boldsymbol{r}} \in L_1^d(\Omega, \mathcal{F}, \mathbb{P})$ be the random vector of returns. Interpreting $-\boldsymbol{x}'\widetilde{\boldsymbol{r}}$ as the random losses of the portfolio, a classic and well studied approach to this problem is to minimize the Conditional Value-at-Risk of the losses:

$$z_\delta^* := \min_{\boldsymbol{x}} \{CVaR_\delta(-\boldsymbol{x}'\widetilde{\boldsymbol{r}}) : \boldsymbol{x}'\boldsymbol{e} = 1, \boldsymbol{x} \geq \boldsymbol{0}\}, \tag{11}$$

where $CVaR_\delta(\widetilde{v}) := \min_t \left\{t + \frac{1}{\delta}\mathbb{E}[(\widetilde{v} - t)^+]\right\}$. If the distribution of $\widetilde{\boldsymbol{r}}$ is known, then (11) is a well defined convex optimization problem which can be solved in theory. However, evaluating *CVaR* requires multidimensional integration and hence solving (11) is, in general, intractable. Furthermore, more often than not, the distribution of $\widetilde{\boldsymbol{r}}$ can only be accessed through a finite number of samples. A common data-driven approach for this issue is to use this finite number of samples to approximate the integrals in the definition of *CVaR* with the sample mean. This approximation technique is known as *Sample Average Approximation* (SAA) for stochastic programming and its convergence is assured under very broad settings, see e.g. Shapiro et al. (2009, Section 5.1.1). Assume then that we have a finite i.i.d. sample $\boldsymbol{r}^1, \ldots, \boldsymbol{r}^N \in \mathbb{R}^d$ of the vector of returns $\widetilde{\boldsymbol{r}}$ (e.g. from past observed returns or simulations). The SAA version of (11) is given by

$$z_{\delta,N}^*\left(\left\{\boldsymbol{r}^i\right\}_{i=1}^N\right) := \min_{\boldsymbol{x}} \left\{CVaR_\delta^N\left(-\boldsymbol{x}'\widetilde{\boldsymbol{r}}, \left\{\boldsymbol{r}^i\right\}_{i=1}^N\right) : \boldsymbol{x}'\boldsymbol{e} = 1, \boldsymbol{x} \geq \boldsymbol{0}\right\}, \tag{12}$$

where $CVaR_\delta^N(-\boldsymbol{x}'\widetilde{\boldsymbol{r}}, \{\boldsymbol{r}^i\}_{i=1}^N) := \min_t\{t + \frac{1}{\delta N}\sum_{i=1}^N[-\boldsymbol{x}'\boldsymbol{r}^i - t]^+\}$ is $CVaR_\delta$ for the case in which $\widetilde{\boldsymbol{r}}$ is uniformly distributed in $\{\boldsymbol{r}^i\}_{i=1}^N$. For notational convenience we drop the dependence of $z_{\delta,N}^*$, $CVaR_\delta^N$ and related values, on $\{\boldsymbol{r}^i\}_{i=1}^N$, while noting that any value or solution derived from (12) is dependent on the N samples of $\widetilde{\boldsymbol{r}}$ and hence is random unless the sample is fixed. With this in mind, it is well known that, under mild conditions, $z_{\delta,N}^*$ converges to z_δ^* w.p. 1 as N grows to infinity and that, under slightly stronger conditions, the optimal set of (12) also converges w.p. 1 to the optimal set of (11) (e.g. Shapiro et al. (2009, Section 5.1)). Furthermore, from Rockafellar and Uryasev (2002, 2000) we have that (12) is equivalent to

$$\min_{\boldsymbol{x},t}\left\{t + \frac{1}{\delta N}\sum_{i=1}^N[-\boldsymbol{x}'\boldsymbol{r}^i - t]^+ : t \in \mathbb{R}, \boldsymbol{x}'\boldsymbol{e} = 1, \boldsymbol{x} \geq \boldsymbol{0}\right\}. \tag{13}$$

Note that this problem can be formulated as linear programming problem, which can be easily solved. Unfortunately, as noted in Lim et al. (2011), for moderate values of N and small values of δ, the optimal solutions of (12)/(13) can have a significant difference between their sampled $CVaR_\delta^N$ and their real $CVaR_\delta$. Furthermore, the real $CVaR_\delta$ of these solutions can be far from z_δ^*. More specifically, if x_N^* is an optimal solution to (12)/(13) it is common to have $CVaR_\delta^N(x_N^*) \ll z_\delta^* \ll CVaR_\delta(x_N^*)$. We aim to use $\widehat{CVaR}_{\gamma,\varepsilon}$ (i.e., an epsilon scaling of $CVaR_\gamma$ with γ not necessarily equal to δ) to construct a variant of (12)/(13) with optimal solutions that reduce both these gaps. Our motivation for this construction can be best illustrated if we consider elliptical distributions, which have the following convenient characterization of $\mathcal{U}(\rho)$ that we prove in Appendices A and B. The use of this characterization will come from the equivalence between approximating $CVaR_\delta$ with $CVaR_\delta^N$ and approximating $\mathcal{U}(CVaR_\delta)$ with $\mathcal{U}(CVaR_\delta^N)$

Lemma 5.1. *Let $\mu \in \mathbb{R}^d$, $B \in \mathbb{R}^{d\times d}$ be a non-singular matrix and let $\widetilde{\boldsymbol{r}} \in L_1^d(\Omega, \mathcal{F}, \mathbb{P})$ be such that $\widetilde{u}_{\boldsymbol{x}} := \boldsymbol{x}'B^{-1}(\widetilde{\boldsymbol{r}} - \mu)$ has the same continuous probability distribution for every $\boldsymbol{x} \in S^{d-1} := \left\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\right\}$ (e.g. $\widetilde{\boldsymbol{r}}$ is the uniformly distribution over the ellipsoid $\left\{\boldsymbol{r} \in \mathbb{R}^d : \left\|B(\boldsymbol{r} - \mu)\right\|_2 \leq 1\right\}$ or $\boldsymbol{r} \sim \mathcal{N}(\mu, BB')$). Then, for any distortion risk measure ρ we have*

$$\mathcal{U}(\rho) = \left\{\boldsymbol{r} \in \mathbb{R}^d : \left\|B^{-1}(\boldsymbol{r} - \mu)\right\|_2 \leq \rho(\widetilde{u}_{\boldsymbol{x}_0})\right\} \tag{14}$$

where $\boldsymbol{x}_0$ is an arbitrary element of S^{d-1}.

If $\widetilde{\boldsymbol{r}}$ is distributed as in Lemma 5.1 with $B = I$ and $\mu = 0$, then $\mathcal{U}(CVaR_\delta)$ is an Euclidean ball for any δ. In turn, the characterization from Example 2.2 shows that, if $\delta N \in \mathbb{Z}_+$, then $\mathcal{U}(CVaR_\delta^N)$ is the convex hull of the $\binom{N}{\delta N}$ points in $\widehat{\Omega}_{\boldsymbol{h}^\delta} := \{\sum_{i=1}^N \boldsymbol{h}_{\sigma(i)}^\delta \boldsymbol{r}^i : \sigma \in S_N\}$ for $\boldsymbol{h}^\delta$ defined in (8) ($\widehat{\Omega}_{\boldsymbol{h}^\delta}$ corresponds to all averages of δN points from $\{\boldsymbol{r}^i\}_{i=1}^N$). Now, it is well known that to obtain a good approximations of the Euclidean ball by a set of the form $\mathrm{conv}(\widehat{\Omega}_{\boldsymbol{h}^\delta})$ we need the number of extreme points of this set to be quite large (see Ball (1997)). While it is hard to predict the number of extreme points of $\mathrm{conv}(\widehat{\Omega}_{\boldsymbol{h}^\delta})$, it is likely to be a non-decreasing function of $|\widehat{\Omega}_{\boldsymbol{h}^\delta}| = \binom{N}{\delta N}$. Hence, we would then expect the approximation of $\mathcal{U}(CVaR_{0.5})$ by $\mathcal{U}(CVaR_{0.5}^N)$ to be much better that the approximation of $\mathcal{U}(CVaR_\delta)$ by $\mathcal{U}(CVaR_\delta^N)$ for small δ. This aligns with the SAA approximation issues of $CVaR_\delta$ being worse for small δ. Unfortunately, small values of δ are precisely the ones needed to incorporate appropriate levels of risk aversion and it is unlikely that $\mathcal{U}(CVaR_{0.5}^N)$ will provide a good approximation of $\mathcal{U}(CVaR_\delta)$ for $\delta \ll 0.5$. However, by noting that both $\mathcal{U}(CVaR_{0.5})$ and $\mathcal{U}(CVaR_\delta)$ are Euclidean balls (just with different radii), we have that $\mathcal{U}(CVaR_{0.5}^N)$ is indeed a good approximation of a scaling of $\mathcal{U}(CVaR_\delta)$ for $\delta \ll 0.5$. Conversely, for any δ an appropriate scaling of $\mathcal{U}(CVaR_{0.5}^N)$ will be a good approximation of $\mathcal{U}(CVaR_\delta)$. More precisely, if $r_{0.5}$ is the radius of $\mathcal{U}(CVaR_{0.5})$ and r_δ is the radius of $\mathcal{U}(CVaR_\delta)$, then $\mathcal{U}(CVaR_\delta) = (r_\delta/r_{0.5})\mathcal{U}(CVaR_{0.5})$ and hence we expect $(r_\delta/r_{0.5})\mathcal{U}(CVaR_{0.5}^N)$ to be a better approximation of $\mathcal{U}(CVaR_\delta)$ than $\mathcal{U}(CVaR_\delta^N)$ (at least for small δ). The potential advantage of using $\widehat{CVaR}_{\gamma,\varepsilon}$ emerges by noting that scalings of $\mathcal{U}(CVaR_{0.5}^N)$ are precisely the uncertainty sets $\mathcal{U}(\widehat{CVaR}_{0.5,\varepsilon}^N)$ of $\widehat{CVaR}_{0.5,\varepsilon}$ for an appropriately chosen ε. We formalize this in the following corollary that shows how to calculate the appropriate ε for elliptical distributions and all values of δ. Note that the proposition can be directly extended to $\widehat{CVaR}_{\gamma,\varepsilon}$ for values of γ other than 0.5.

Corollary 5.2. *Let $\mu \in \mathbb{R}^d$, $B \in \mathbb{R}^{d\times d}$ be a non-singular matrix and let $\widetilde{\boldsymbol{r}} \in L_1^d(\Omega, \mathcal{F}, \mathbb{P})$ be such that $\widetilde{u}_{\boldsymbol{x}} := \boldsymbol{x}'B^{-1}(\widetilde{\boldsymbol{r}} - \mu)$ has the same continuous probability distribution for every $\boldsymbol{x} \in S^{d-1} := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}$. Then, $\mathcal{U}(CVaR_\delta) = \mathcal{U}(\widehat{CVaR}_{0.5,\varepsilon})$ for $\varepsilon = \frac{CVaR_\delta(\widetilde{u}_{\boldsymbol{x}_0})}{CVaR_{0.5}(\widetilde{u}_{\boldsymbol{x}_0})}$, where $\boldsymbol{x}_0$ is an arbitrary element of S^{d-1}.*

Proof. Note that $\widehat{CVaR}_{0.5,\varepsilon}(\widetilde{u}_{\boldsymbol{x}_0}) = \varepsilon CVaR_{0.5}(\widetilde{u}_{\boldsymbol{x}_0}) + (1 - \varepsilon)\mathbb{E}(\widetilde{u}_{\boldsymbol{x}_0}) = \varepsilon CVaR_{0.5}(\widetilde{u}_{\boldsymbol{x}_0}) = CVaR_\delta(\widetilde{u}_{\boldsymbol{x}_0})$. □

The following example provides a graphical illustration of the advantage of using $\mathcal{U}(\widehat{CVaR}_{0.5,\varepsilon}^N)$ over $\mathcal{U}(CVaR_\delta^N)$ to approximate $\mathcal{U}(CVaR_\delta)$.

Example 5.1. The uncertainty set associated with $CVaR_{1/8}$ for a three-dimensional standard normal distributed $\widetilde{\boldsymbol{r}}$ corresponds to a sphere of radius 1.6468 centered at the origin. Fig. 2 shows the uncertainty

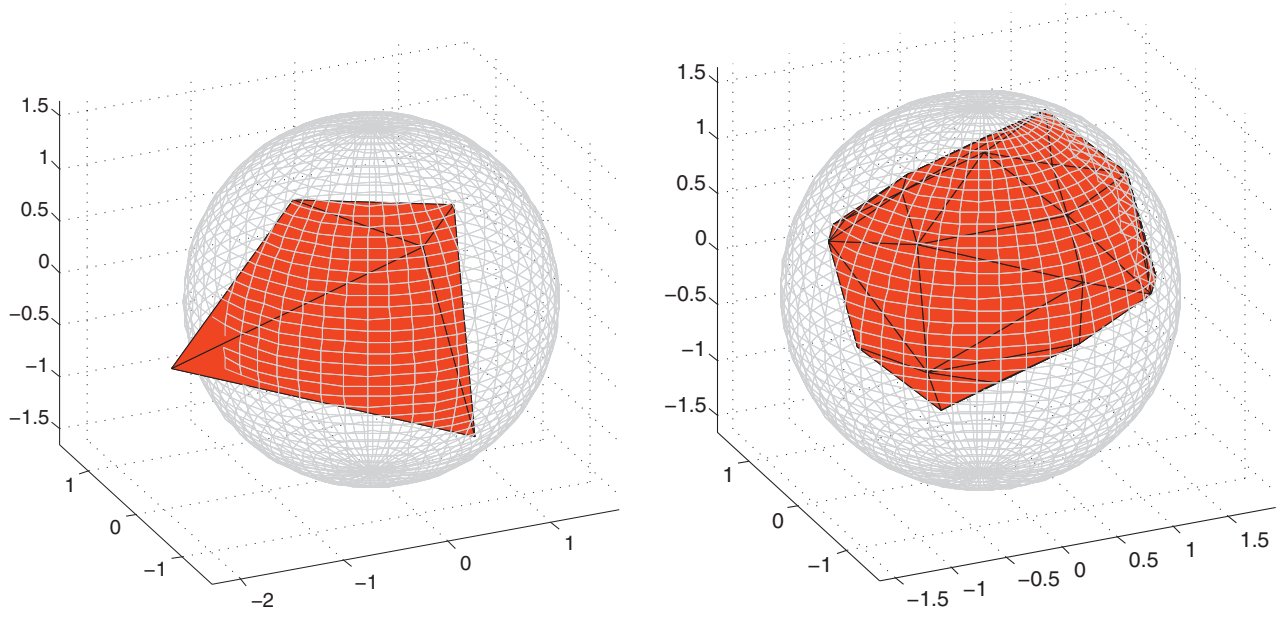

Fig. 2. Approximation of $\mathcal{U}\,(CVaR_{1/8})$ by $\mathcal{U}\left(CVaR^N_{1/8}\right)$ and $\mathcal{U}\left(\widehat{CVaR}^N_{0.5,\varepsilon}\right)$ for $N = 8$.

sets associated with $CVaR^N_{1/8}$ (left) and $\widehat{CVaR}^N_{0.5,\varepsilon}$ (right) for a sample of $N = 8$ random points (with ε selected as in Corollary 5.2). The uncertainty set associated with $CVaR^N_{1/8}$ has six vertices and eight faces. In contrast, the uncertainty set associated with $\widehat{CVaR}^N_{0.5,\varepsilon}$ has 30 vertices and 56 faces and seems to give a closer approximation of the ball.

Corollary 5.2 shows that

$$z^*_{0.5,\varepsilon} := \min\left\{\widehat{CVaR}_{0.5,\varepsilon}\left(-\boldsymbol{x}'\widetilde{\boldsymbol{r}}\right) : \boldsymbol{x}'\boldsymbol{e} = 1,\ \boldsymbol{x} \geq \boldsymbol{0}\right\} \tag{15}$$

is equivalent to (11) if $\widetilde{\boldsymbol{r}}$ is elliptically distributed and ε is chosen as in the corollary (in particular $z^*_{0.5,\varepsilon} = z^*_\delta$). Furthermore, Example 5.1 illustrates how the SAA version (12)/(13) of (11) is not equivalent to the SAA version of (15) given by

$$z^*_{0.5,\varepsilon,N} := \min_{\boldsymbol{x}}\left\{\widehat{CVaR}^N_{0.5,\varepsilon}\left(-\boldsymbol{x}'\widetilde{\boldsymbol{r}}\right) : \boldsymbol{x}'\boldsymbol{e} = 1,\ \boldsymbol{x} \geq \boldsymbol{0}\right\}, \tag{16}$$

where $\widehat{CVaR}^N_{0.5,\varepsilon}(-\boldsymbol{x}'\widetilde{\boldsymbol{r}}) := \varepsilon \min_t\{t + \frac{1}{0.5N}\sum_{i=1}^N[-\boldsymbol{x}'\boldsymbol{r}^i - t]^+\} - (1-\varepsilon)\ \frac{1}{N}\sum_{i=1}^N \boldsymbol{x}'\boldsymbol{r}^i$. However, Example 5.1 also suggests that (16) is likely to provide a better approximation of (11) than (12)/(13). While the equivalence between (11) and (15) no longer holds for general distributions, (16) might still provide a better approximation than (12)/(13). In particular, it is still reasonable to expect that $\mathcal{U}(\widehat{CVaR}^N_{0.5,\varepsilon})$ has a richer structure than $\mathcal{U}(CVaR^N_\delta)$ as the former is constructed by taking a larger number of partial averages of the sample points. This could have a smoothing effect similar to the one depicted in Fig. 2, which could provide more stability for small sample sizes. Still, as N grows, we can only guarantee that $\widehat{CVaR}^N_{0.5,\varepsilon}$ converges to $\widehat{CVaR}_{0.5,\varepsilon}$, and this last risk measure may not be equivalent to $CVaR_\delta$ for any δ. However, Corollary 3.4 shows that $\widehat{CVaR}_{0.5,\varepsilon}$ is a valid risk measure on its own right, which validates the use of $\widehat{CVaR}^N_{0.5,\varepsilon}$ independent of its potential approximation of $CVaR_\delta$. Nonetheless, in the next two subsections we test quality of this potential approximation on both elliptical and non-elliptical distributions. We end this section with two observations. The first one concerns the calculation of the scaling factor ε for non-elliptical distributions. While Corollary 5.2 no longer provides a precise formula we could still follow its general idea and choose

$$\varepsilon \approx \frac{CVaR_\delta(-\boldsymbol{x}'\widetilde{\boldsymbol{r}}) - \mathbb{E}[-\boldsymbol{x}'\widetilde{\boldsymbol{r}}]}{CVaR_{0.5}(-\boldsymbol{x}'\widetilde{\boldsymbol{r}}) - \mathbb{E}[-\boldsymbol{x}'\widetilde{\boldsymbol{r}}]} \tag{17}$$

for some fixed $\boldsymbol{x} \in S^{d-1}$. Our approach will be to select a SAA approximation of this ratio.

Our final observation is that, similar to (13), (16) is also equivalent to the convenient problem given by

$$\min\left\{\varepsilon\left(t + \frac{1}{0.5N}\sum_{i=1}^N[-\boldsymbol{x}'\boldsymbol{r}^i - t]^+\right) - (1-\varepsilon)\,\boldsymbol{x}'\bar{\boldsymbol{r}} : t \in \mathbb{R},\ \boldsymbol{x}'\boldsymbol{e} = 1,\ \boldsymbol{x} \geq \boldsymbol{0}\right\} \tag{18}$$

where $\bar{\boldsymbol{r}} := \frac{1}{N}\sum_{i=1}^N \boldsymbol{r}^i$, which can be also easily formulated as a linear programming problem.

5.1. Results for Gaussian distribution

We begin our experiments with a Gaussian distribution as it satisfies the conditions of Corollary 5.2 and it also allows for the exact solution of (11). To generate the data for our experiments we utilize the same historical data for 200 stocks listed in SP-500 used in Vielma, Ahmed, and Nemhauser (2008) to estimate the mean vector μ and covariance matrix Σ of these assets. We then assume that the real distribution of the assets is Gaussian with this mean and covariance. Hence, by Lemma 5.1, we have that (11) is equivalent to the second-order conic problem given by

$$z^*_\delta = \min_{\boldsymbol{x},\boldsymbol{t}}\left\{CVaR_\delta\left(\widetilde{\nu}\right)\cdot t - \boldsymbol{x}'\bar{\boldsymbol{r}} : \boldsymbol{x}'\boldsymbol{e} = 1,\ \left\|\Sigma^{1/2}\boldsymbol{x}\right\|_2 \leq \boldsymbol{t},\quad \boldsymbol{x},\boldsymbol{t} \geq \boldsymbol{0}\right\} \tag{19}$$

where $\widetilde{\nu} \sim \mathcal{N}(0,1)$.

Our objective is to compare the approximation effectiveness of $CVaR^N_\delta$ and $\widehat{CVaR}^N_{0.5,\varepsilon}$ for this problem, with a particular emphasis on the quality of the obtained feasible portfolios. For this we proceed as follows for each $\delta \in \{0.01, 0.1\}$.

1. Generate N i.i.d. samples from our real distribution $\mathcal{N}(\mu, \Sigma)$.

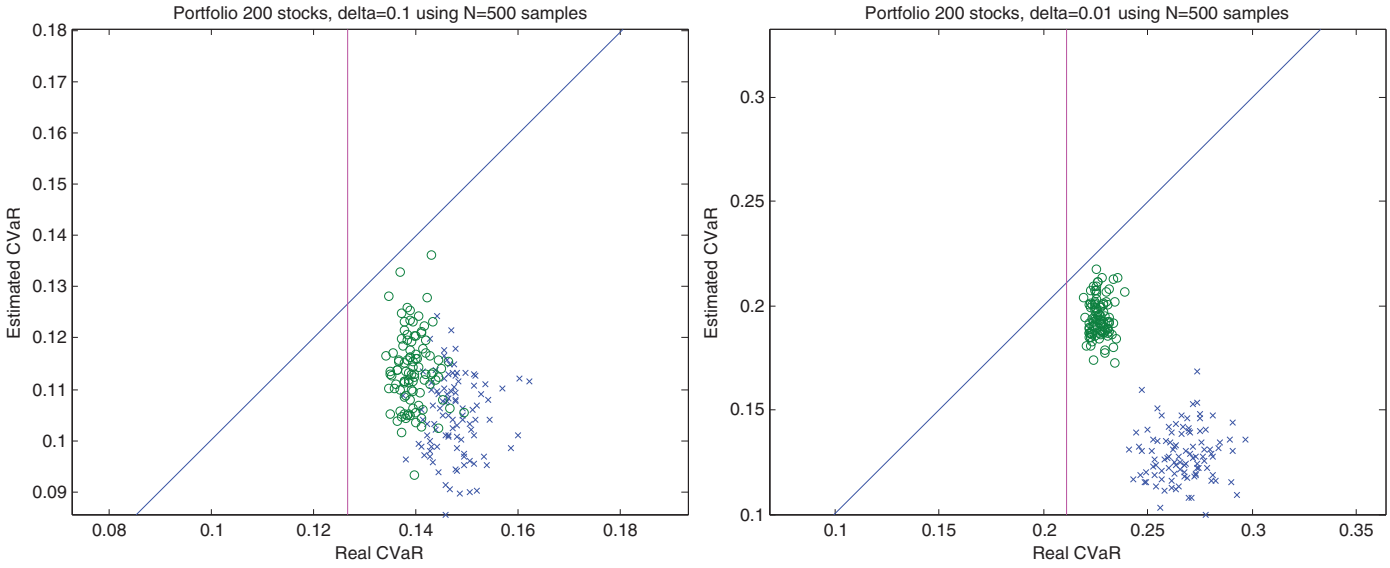

Fig. 3. Mitigating estimation errors of *CVaR* (Gaussian distribution).

Table 1
Solution quality for portfolio of 20 and 200 stocks for different values of δ and N (Gaussian distribution).

Portfolio size	δ	z^*_δ	Type	Best solution			Average solution		
				$N=100$	$N=500$	$N=10,000$	$N=100$	$N=500$	$N=10,000$
20	0.01	0.3502	$CVaR^N_\delta$	0.3709	0.3610	0.3509	0.4398	0.3932	0.3535
			$\widehat{CVaR}^N_{0.5,\varepsilon(\delta)}$	0.3577	0.3521	0.3503	0.3846	0.3576	0.3506
20	0.1	0.2195	$CVaR^N_\delta$	0.2333	0.2206	0.2197	0.2595	0.2288	0.2200
			$\widehat{CVaR}^N_{0.5,\varepsilon(\delta)}$	0.2265	0.2208	0.2196	0.2461	0.2253	0.2198
200	0.01	0.2107	$CVaR^N_\delta$	0.2866	0.2412	0.2140	0.3643	0.2665	0.2171
			$\widehat{CVaR}^N_{0.5,\varepsilon(\delta)}$	0.2488	0.2190	0.2112	0.2829	0.2267	0.2116
200	0.1	0.1266	$CVaR^N_\delta$	0.1741	0.1375	0.1272	0.2179	0.1474	0.1277
			$\widehat{CVaR}^N_{0.5,\varepsilon(\delta)}$	0.1545	0.1342	0.1269	0.1916	0.1395	0.1272

2. Solve the sampled *CVaR* problem (13) and save the optimal solution $\boldsymbol{x}^*_{CVaR^N_\delta}$.
3. Compute $\varepsilon(\delta) = \frac{CVaR_\delta(\tilde{\nu})}{CVaR_{0.5}(\tilde{\nu})}$ for $\tilde{\nu} \sim \mathcal{N}(0,1)$.
4. Solve the sampled $\widehat{CVaR}^N_{0.5,\varepsilon(\delta)}$ problem (18) and save the optimal solution $\boldsymbol{x}^*_{\widehat{CVaR}^N_{0.5,\varepsilon(\delta)}}$.
5. Plot $CVaR_\delta(-\boldsymbol{x}'\tilde{\boldsymbol{r}})$ versus $CVaR^N_\delta(-\boldsymbol{x}'\tilde{\boldsymbol{r}})$ for $\boldsymbol{x} \in \left\{\boldsymbol{x}^*_{CVaR^N_\delta}, \boldsymbol{x}^*_{\widehat{CVaR}^N_{0.5,\varepsilon(\delta)}}\right\}$.
6. Repeat steps 1–5 100 times.

Fig. 3 shows the results for this experiment. Blue *x*s correspond to $\boldsymbol{x}^*_{CVaR^N_\delta}$ and green circles correspond to $\boldsymbol{x}^*_{\widehat{CVaR}^N_{0.5,\varepsilon(\delta)}}$. The vertical magenta line shows the exact z^*_δ as computed by (19), and the diagonal blue line corresponds to equal values for the real and sampled *CVaR*. As expected (e.g. Shapiro et al. (2009, Proposition 5.6)), the sampled *CVaR* consistently underestimates the real *CVaR* and this effect is more significant for $\delta = 0.01$. However, the epsilon scaling tends to reduce this downward bias. More importantly, the epsilon scaling reduces variability of both the sampled and real *CVaR* of the optimal solutions and tends to provide better solutions to the original problem.

The increased concentration along the real *CVaR* axis of the epsilon scalings solutions can be particularly advantageous when considering hard-to-solve optimization problems. Estimating the real *CVaR* of a particular solution can be significantly easier than approximating the whole *CVaR* function. Hence, if we can generate a relatively large number[1] of potentially good solutions, it is reasonable to estimate the real *CVaR* and pick the best one. For instance, if we look at the best among the traditional solutions (the blue x further to the left) we can see that it is a relatively good solution. However, generating enough solutions to guarantee we find such best solution may not always be computationally feasible. For example, if we consider portfolio optimization problems with limited diversification or cardinality constraints problem (11) becomes a mixed integer problem that can be very hard to solve (Vielma et al., 2008). Hence, in some cases, a more realistic comparison may be to simulate the effect of solving a single instance of the appropriate optimization problem by randomly selecting one of the traditional solutions (blue *x*'s) and one of the epsilon scaling solutions (green circles). We explore this evaluation in Table 1 where we also study the effect on the results of the number of samples and the number stocks.

Table 1 shows results for portfolio sized of 20 and 200 stocks, $\delta \in \{0.1, 0.01\}$ and sample sizes of $N = 100, 500$ and $10,000$. Column *Best Solution* shows the smallest value of $CVaR_\delta(-\boldsymbol{x}'\tilde{\boldsymbol{r}})$ over the 100 repetitions for each $\boldsymbol{x} \in \{\boldsymbol{x}^*_{CVaR^N_\delta}, \boldsymbol{x}^*_{\widehat{CVaR}^N_{0.5,\varepsilon(\delta)}}\}$. This is intended to illustrate the case in which the optimization problem is easy to use and we can generate several candidate solutions, evaluate them and select the best. In contrast, column *Average Solution* shows the average value of $CVaR_\delta(-\boldsymbol{x}'\tilde{\boldsymbol{r}})$ over the 100 repetitions. This is intended to illustrate the case in which the optimization problem is hard to

[1] Large enough to have variety, but still significantly smaller than all feasible solutions.

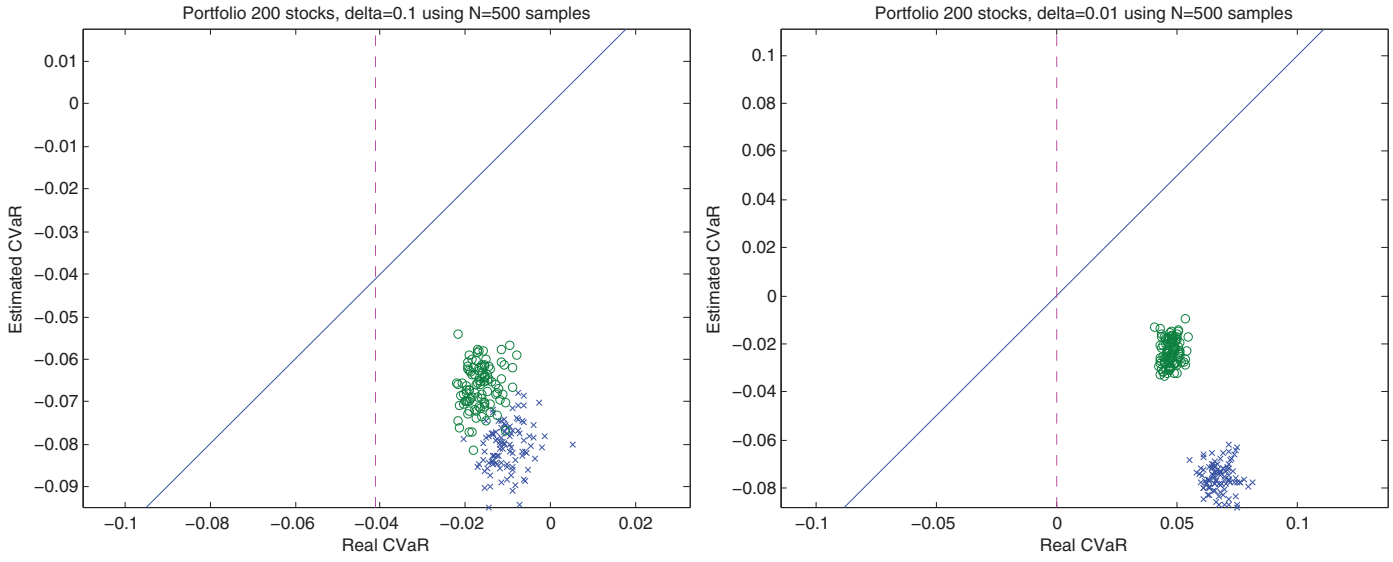

Fig. 4. Mitigating estimation errors of *CVaR* (uniform distribution).

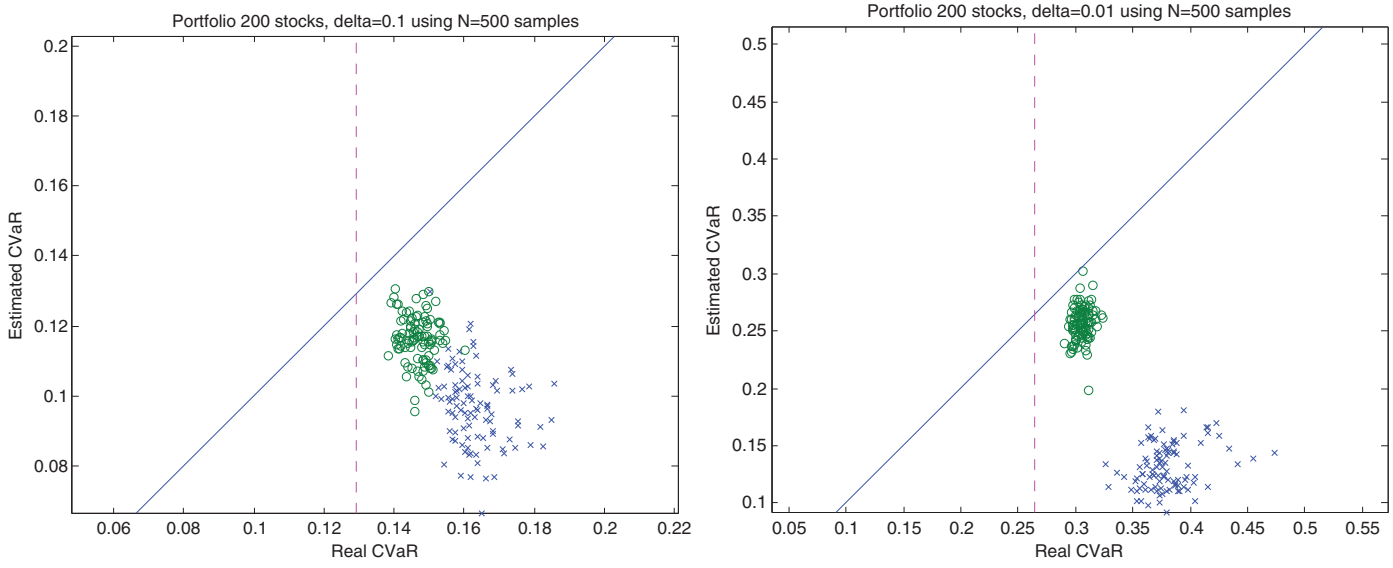

Fig. 5. Mitigating estimation errors of *CVaR* (normal-inverse Gaussian distribution).

solve and only one or very few solutions can be generated (i.e. we expect this average to be representative of a typical single solution). Finally, column z^*_δ shows the exact optimal value obtained through (19). We can see that the epsilon scaling yields better solutions for all parameters and metrics. This advantage is particularly strong for the metric of average solution and small number of samples and δ. In Appendix B we show how this advantage is increased further when we allow short-selling in the portfolio problem (i.e. when we remove the non-negativity constraint on x variables).

5.2. Results for other non-gaussian distribution

To study a case in which the conditions of Corollary 5.2 do not hold we repeat the previous experiment assuming that returns follow a uniform and a normal-inverse gaussian distribution. In the first case, each stock has a return $r_i = \mu_i + \tilde{\eta}_i$, where $\tilde{\eta}_i$ are independent random variables uniformly distributed in $[-1, 1]$. Note that in this case $\mathcal{U}(CVaR_\delta)$ for different δs are not scalings of one another. In the second case, we assume that r_i follows a multivariate normal-inverse Gaussian distribution, which is a heavy-tailed distribution commonly used on finance. In this latter case, we assume that stocks have a return $r = \mu + \sqrt{\tau}\upsilon$ (Aas, Haff, & Dimakos, 2006) where $\upsilon \sim \mathcal{N}(0, \Sigma)$ and τ follows a generalized inverse Gaussian distribution of parameters $\lambda = -0.5$, $\chi = 1$ and $\psi = 1$ (following the notation of Prause (1999)).

On both cases, even evaluating $CVaR_\delta$ requires multidimensional integration. For this reason we compute the scaling factor ε in step 3 as the sampled estimation of (17), given by

$$\varepsilon = \frac{CVaR^M_\delta(-\boldsymbol{x}'\tilde{\boldsymbol{r}}) + \frac{1}{M}\sum_{i=1}^M \boldsymbol{x}'r^i}{CVaR^M_{0.5}(-\boldsymbol{x}'\tilde{\boldsymbol{r}}) + \frac{1}{M}\sum_{i=1}^M \boldsymbol{x}'r^i}$$

where $\boldsymbol{x}$ is the solution obtained in step 2, and $M = 100,000$.

As explained in Shapiro et al. (2009, Section 5.6.1), it is possible to estimate a lower bound of z^*_δ using the law of large numbers computing the average and the variance of $CVaR^N_\delta(\boldsymbol{x}^*_{CVaR^N_\delta})$ over the 100 repetitions. We use this bound in Figs. 4 and 5 to replace the exact value calculated with (19), which is not applicable here. We now use a vertical dashed magenta line to emphasize that it is only a lower bound that holds with high probability and not the exact value of z^*_δ. Similarly, this bound in Tables 2 and 3 is labeled as $\underline{z}^*_\delta$.

We again see that the epsilon scaling provides an advantage, particularly for small number of samples and δ. Furthermore, while the

Table 2
Solution quality for portfolio of 20 and 200 stocks for different values of δ and N (uniform distribution).

Portfolio size	δ	z^*_δ	Type	Best solution			Average solution		
				$N = 100$	$N = 500$	$N = 10,000$	$N = 100$	$N = 500$	$N = 10,000$
20	0.01	0.2044	$CVaR^N_\delta$	0.2541	0.2288	0.2123	0.2987	0.2619	0.2144
			$\widehat{CVaR}^N_{\gamma,\varepsilon}$	0.2293	0.2173	0.2130	0.2628	0.2243	0.2139
20	0.1	0.0877	$CVaR^N_\delta$	0.1117	0.0970	0.0894	0.1427	0.1046	0.0902
			$\widehat{CVaR}^N_{\gamma,\varepsilon}$	0.1111	0.0931	0.0904	0.1297	0.1000	0.0909
200	0.01	0.0003	$CVaR^N_\delta$	0.1068	0.0591	0.0271	0.1326	0.0677	0.0294
			$\widehat{CVaR}^N_{\gamma,\varepsilon}$	0.0899	0.0425	0.0182	0.1031	0.0475	0.0191
200	0.1	−0.0410	$CVaR^N_\delta$	0.0092	−0.0183	−0.0359	0.03188	−0.0102	−0.0350
			$\widehat{CVaR}^N_{\gamma,\varepsilon}$	0.0058	−0.0222	−0.0363	0.0246	−0.0165	−0.0360

Table 3
Solution quality for portfolio of 20 and 200 stocks for different values of δ and N (normal-inverse Gaussian distribution).

Portfolio size	δ	z^*_δ	Type	Best solution			Average solution		
				$N = 100$	$N = 500$	$N = 10,000$	$N = 100$	$N = 500$	$N = 10,000$
20	0.01	0.4519	$CVaR^N_\delta$	0.4836	0.4802	0.4654	0.6074	0.5476	0.4700
			$\widehat{CVaR}^N_{\gamma,\varepsilon}$	0.4852	0.4620	0.4627	0.5118	0.4730	0.4637
20	0.1	0.2272	$CVaR^N_\delta$	0.2438	0.2303	0.2282	0.2832	0.2432	0.2287
			$\widehat{CVaR}^N_{\gamma,\varepsilon}$	0.2387	0.2306	0.2278	0.2561	0.2348	0.2282
200	0.01	0.2642	$CVaR^N_\delta$	0.4238	0.3328	0.2901	0.5388	0.3796	0.2969
			$\widehat{CVaR}^N_{\gamma,\varepsilon}$	0.3452	0.3004	0.2795	0.3823	0.3057	0.2833
200	0.1	0.1293	$CVaR^N_\delta$	0.1901	0.1517	0.1333	0.2472	0.1632	0.1337
			$\widehat{CVaR}^N_{\gamma,\varepsilon}$	0.1627	0.1417	0.1324	0.1997	0.1469	0.1326

gap between the traditional $CVaR$ and the epsilon scaling is virtually eliminated for very large number of samples ($N = 10,000$), the epsilon scaling still provides better solutions in both metrics. Again, results for problems where we allow short-selling are included in Figs. B.1–B.3 of Appendix B.

6. Conclusions

We have shown that, at least for finite uniform distributions, the family of uncertainty sets associated with distortion risk measures over affine or linear random variables is strictly larger that those associated with distortion risk measures over arbitrary random variables. In particular, we have shown that certain expansions of uncertainty sets associated with distortion risk measures also yield distortion risk measures over affine or linear random variables. This effectively expands the family of uncertainty sets with favorable theoretical properties. To study the potential advantage of these additional uncertainty sets we have included some preliminary experiments that suggest that these expansions could be useful to mitigate estimation errors.

We finally note that the additional uncertainty sets we have considered still do not give a precise characterization of the family of uncertainty sets associated with distortion risk measures over affine or linear random variables. In particular, it is easy to find examples where the law invariance property is also moot for linear random variables. For example, consider $\Omega = \{-1, 0, 2\}$ with the uniform probability. In this space a linear random variable is represented by a scalar x and its realizations are given by $\{-x, 0, 2x\}$. It is easy to see that the random variables associated with x and y have the same distribution only if $-x = 2y$ and $-y = 2x$. The only solution to this system is $x = y = 0$ and hence there are no non-trivial linear random variables with the same distribution. More general settings might not completely eliminate the possibility of non-trivial linear random variables with the same distribution. However, a significant limitation of such random variables could validate the use of additional uncertainty sets. Still, it is likely that any characterization of these yet additional sets will be highly dependent on the specific structure of Ω.

Acknowledgments

Daniel Espinoza was supported by FONDECYT grant number 1110024 and ICM grant number P10-024-F and Eduardo Moreno was supported by Anillo grant number ACT88 and FONDECYT grant number 1130681. The authors also thank two anonymous referees for their helpful feedback.

Appendix A. Proof of Lemma 5.1

Proof.

$$\begin{aligned}
\mathcal{U}(\rho) &= \left\{\boldsymbol{u} \in \mathbb{R}^d \,:\, \boldsymbol{x}'\boldsymbol{u} \le \rho(\widetilde{\boldsymbol{u}}'\boldsymbol{x}) \text{ for all } x \in S^{n-1}\right\} \\
&= \left\{\boldsymbol{u} \in \mathbb{R}^d \,:\, \boldsymbol{x}'(\boldsymbol{u}-\mu) \le \rho((\widetilde{\boldsymbol{u}}-\mu)'\boldsymbol{x}) \text{ for all } x \in S^{n-1}\right\} \\
&= \left\{\boldsymbol{u} \in \mathbb{R}^d \,:\, \left(\frac{(B^{-1})'\boldsymbol{x}}{\left\|(B^{-1})'\boldsymbol{x}\right\|_2}\right)' \right. \\
&\qquad \left. \times (\boldsymbol{u}-\mu) \le \rho\left((\widetilde{\boldsymbol{u}}-\mu)'\left(\frac{(B^{-1})'\boldsymbol{x}}{\left\|(B^{-1})'\boldsymbol{x}\right\|_2}\right)\right) \text{ for all } x \in S^{n-1}\right\} \\
&= \{\boldsymbol{u} \in \mathbb{R}^d \,:\, \boldsymbol{x}'B^{-1}(\boldsymbol{u}-\mu) \le \rho(\boldsymbol{x}'B^{-1}(\widetilde{\boldsymbol{u}}-\mu)) \text{ for all } x \in S^{n-1}\} \\
&= \left\{\boldsymbol{u} \in \mathbb{R}^d \,:\, \sup_{x \in S^{n-1}} \boldsymbol{x}'B^{-1}(\boldsymbol{u}-\mu) \le \rho(\widetilde{u}_{\boldsymbol{x}_0})\right\} \\
&= \left\{\boldsymbol{u} \in \mathbb{R}^d \,:\, \left\|B^{-1}(\boldsymbol{u}-\mu)\right\|_2 \le \rho(\widetilde{u}_{\boldsymbol{x}_0})\right\}.
\end{aligned}$$

The first equality comes from translation equivariance of ρ, the second one comes from non-singularity of $\left(B^{-1}\right)'$, the third from positive homogeneity of ρ and the fourth comes from $\rho(\widetilde{u}_{\boldsymbol{x}}) = \rho(\widetilde{u}_{\boldsymbol{x}_0})$ for all $x \in S^{n-1}$ because ρ is law invariant and the assumption on the distribution of $\widetilde{u}_{\boldsymbol{x}}$. □

Appendix B. Results for the portfolio problem allowing short-selling

See Figs. B.1–B.3.

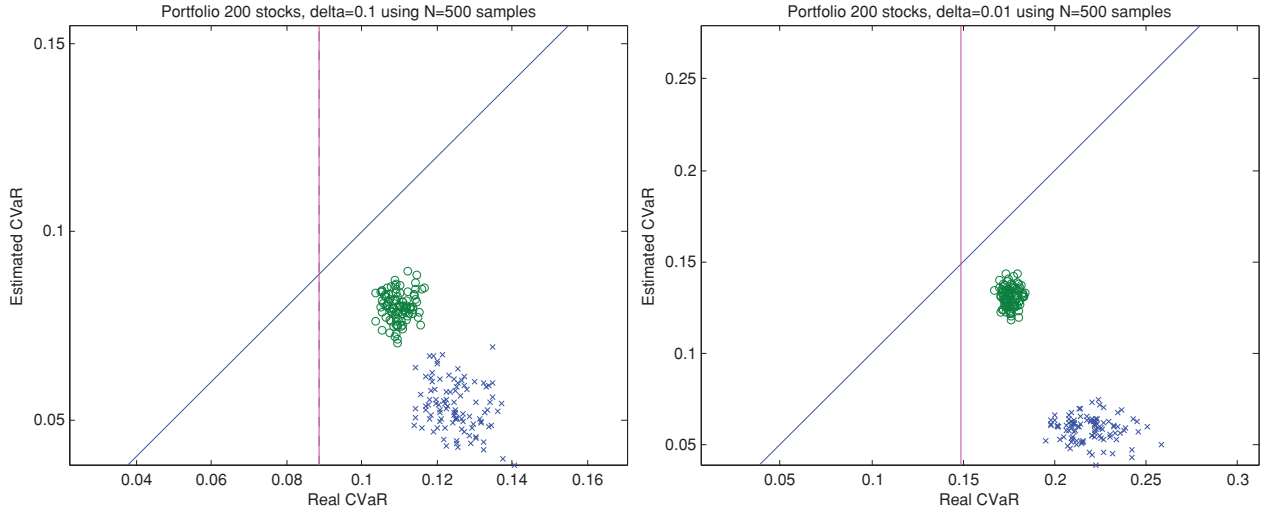

Fig. B.1. Mitigating estimation errors of *CVaR* allowing short-selling (Gaussian distribution).

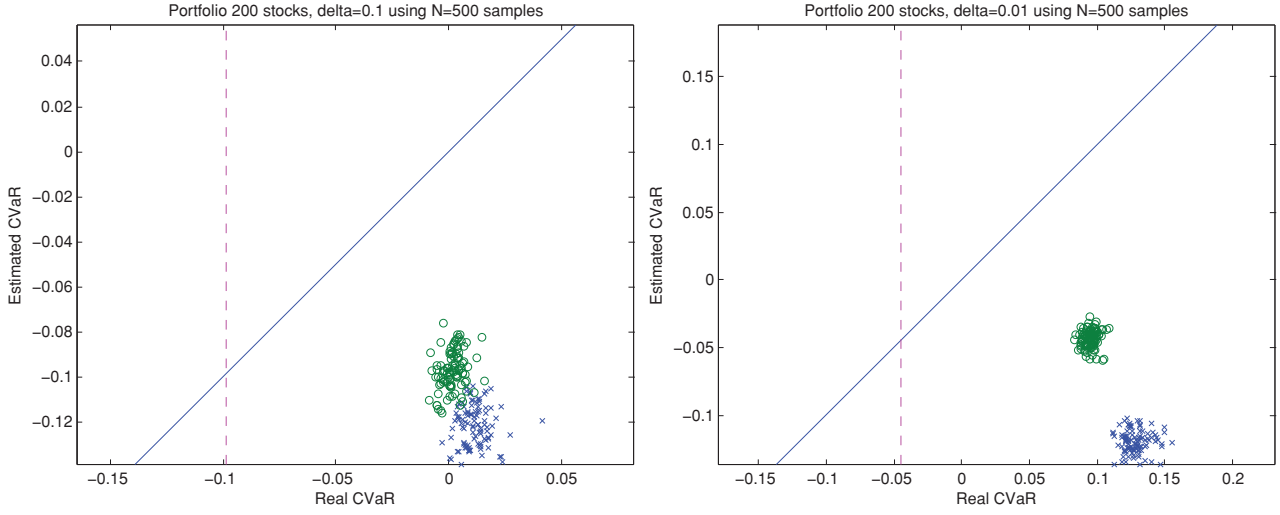

Fig. B.2. Mitigating estimation errors of *CVaR* allowing short-selling (uniform distribution).

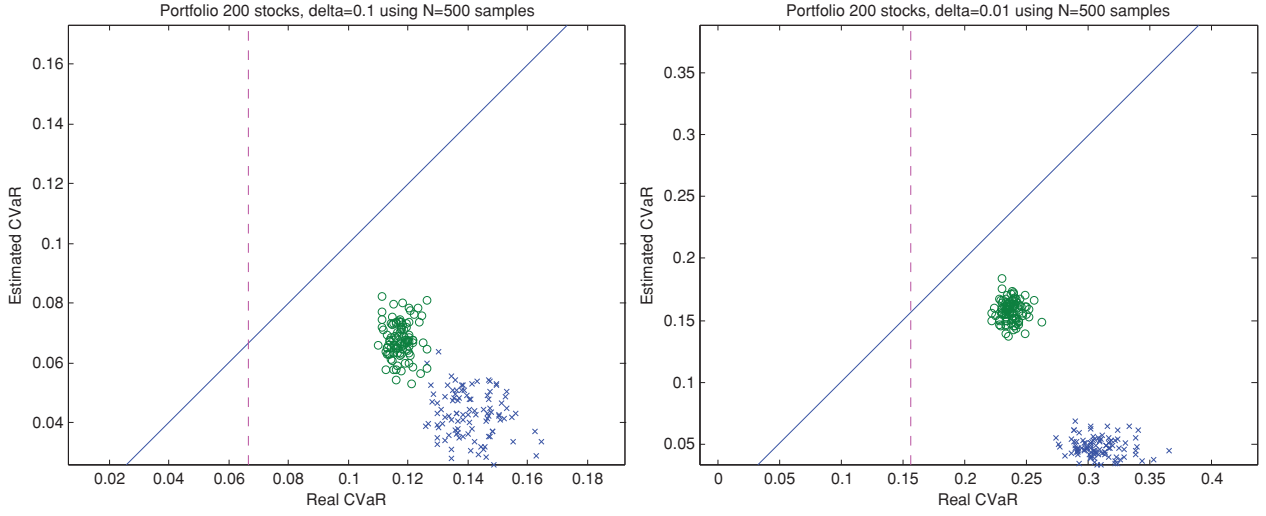

Fig. B.3. Mitigating estimation errors of *CVaR* allowing short-selling (normal-inverse Gaussian distribution).

References

Aas, K., Haff, I. H., & Dimakos, X. K. (2006). Risk estimation using the multivariate normal inverse Gaussian distribution. *Journal of Risk*, *8* (2), 39–60.

Adams, W., & Sherali, H. (1986). A tight linearization and an algorithm for zero-one quadratic programming problems. *Management Science*, *32*(10), 1274–1290.

Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228.

Ball, K. (1997). An elementary introduction to modern convex geometry. *Flavors of Geometry*, *31*, 1–58.

Ben-Tal, A., Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization*. Princeton series in applied mathematics. Princeton University Press.

Bertsimas, D., & Brown, D. (2009). Constructing uncertainty sets for robust linear optimization. *Operations Research*, *57*6, 1483–1495.

Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, *48*5, 28–43.

Cook, W., Koch, T., Steffy, D. E., & Wolter, K. (2011). An exact rational mixed-integer programming solver. In O. Günlük, & G. J. Woeginger (Eds.), *Integer programming and combinatorial optimization: vol. 6655. Lecture notes in computer science* (pp. 104–116). Springer.

Cvitanić, J., & Karatzas, I. (1992). Convex duality in constrained portfolio optimization. *The Annals of Applied Probability*, *2*(4), 767–818. 10.1214/aoap/1177005576.

Espinoza, D. G. (2006). *On linear programming, integer programming and cutting planes*. (Ph.D. thesis). Georgia Institute of Technology.

Fertis, A., Baes, M., & Lüthi, H.-J. (2012). Robust risk management. *European Journal of Operational Research*, *222*(3), 663–672.

Hiriart-Urruty, J.-B., & Lemaréchal, C. (2001). *Fundamentals of convex analysis*. Heidelberg: Springer Verlag.

Kawas, B., & Thiele, A. (2011). Short sales in log-robust portfolio management. *European Journal of Operational Research*, *215*(3), 651–661.

Kolm, P. N., Tütüncü, R., & Fabozzi, F. J. (2014). 60 years of portfolio optimization: practical challenges and current trends. *European Journal of Operational Research*, *234*(2), 356–371.

Konno, H., & Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management Science*, *37*(5), 519–531.

Krokhmal, P., Palmquist, J., & Uryasev, S. (2002). Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, *4*(3), 43–68.

Lagos, G., Espinoza, D., Moreno, E., & Amaya, J. (2011). Robust planning for an open-pit mining problem under ore-grade uncertainty. *Electronic Notes in Discrete Mathematics*, *37*(0), 15–20. LAGOS'11—VI Latin-American Algorithms, Graphs and Optimization Symposium. 10.1016/j.endm.2011.05.004.

Lim, A. E., Shanthikumar, J. G., & Vahn, G.-Y. (2011). Conditional Value-at-Risk in portfolio optimization: Coherent but fragile. *Operations Research Letters*, *39*(3), 163–171. 10.1016/j.orl.2011.03.004.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91.

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

Pichler, A., & Shapiro, A. (2013). Uniqueness of Kusuoka representations. Optimization Online. http://www.optimization-online.org/DB_HTML/2012/10/3660.html.

Prause, K. (1999). *The generalized hyperbolic model: Estimation, financial derivatives, and risk measures* (Ph.D. thesis). University of Freiburg.

Rockafellar, R., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, *2*(3), 21–42.

Rockafellar, R., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, *26*(7), 1443–1471.

Shapiro, A. (2013). On Kusuoka representation of law invariant risk measures. *Mathematics of Operations Research*, *38*(1), 142–152.

Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2009). *Lectures on stochastic programming: Modeling and theory*. Society for Industrial Mathematics.

Vielma, J. P., Ahmed, S., & Nemhauser, G. L. (2008). A lifted linear programming branch-and-bound algorithm for mixed-integer conic quadratic programs. *INFORMS Journal on Computing*, *20*(3), 438–450.

Vielma, J. P., Espinoza, D., & Moreno, E. (2009). Risk control in ultimate pits using conditional simulations. In *Proceedings of the 34th international symposium on application of computers and operations research in the mineral industry (APCOM 2009)* (pp. 107–114).

Wächter, H. P., & Mazzoni, T. (2013). Consistent modeling of risk averse behavior with spectral risk measures. *European Journal of Operational Research*, *229*(2), 487–495.

Zymler, S., Rustem, B., & Kuhn, D. (2011). Robust portfolio optimization with derivative insurance guarantees. *European Journal of Operational Research*, *210*(2), 410–424.