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]*Departamento de Ingenieria Industrial, 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.*

Abstract

In this paper we consider the restriction of coherent and distortion risk measures to the spaces of linear and affine linear random variables and the robust uncertainty sets associated to these risk measures. In this context we study two variants of the *permutahull* uncertainty set introduced by Bertsimas and Brown for finite probability spaces. The first variant considers the extension of the permutahull to non-atomic distributions and the second variant considers scalings of the permutahull. In particular, by studying expansions of the permutahull we are able to show that there are measures that are distortion risk measures over the space of linear random variables that cannot be extended to a distortion risk measures over all random variables. We also present preliminary computational results illustrating that using expansions of the permutahull could help mitigate estimation errors.

Keywords: Risk management; Stochastic programming; Uncertainty modelling

1. Introduction

Coherent risk measures and their relation to robust optimization have received significant attention recently (Bertsimas and Brown, 2009; Natarajan et al., 2009; Artzner et al., 1999; Shapiro et al., 2009; Ben-Tal et al.). We now know that every coherent risk measure is associated to 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)). In particular, in the context of finite probability spaces, Bertsimas and Brown (2009) gives a very interesting geometric characterization of a special class of coherent risk measures usually denoted distortion risk measures. This characterization associates every distortion risk measure to a robust uncertainty set referred to as a $\mathbf{q}$-permutahull. This paper studies two variants of the $\mathbf{q}$-permutahull and the risk measures associated to them.

The first variant is an extension of the $\mathbf{q}$-permutahull to non-atomic distributions through a characterization of distortion risk measures introduced in Shapiro (2011). The second variant considers risk measures associated to some scalings of $\mathbf{q}$-permutahulls. When the scaling is a contractions it automatically leads to a distortion risk measure, but when the scaling is an expansion, the associated risk measure can violate the axiom of monotonicity included in the definition of coherent risk measures. Fortunately, we show that in many cases the axiom of monotonicity is satisfied over the subspaces of linear and affine linear random variables even if it is not satisfied over all random variables. Hence, the expansion of a $\mathbf{q}$-permutahull could lead to a measure that is a distortion risk measures only over these subspaces. These measures can sometimes be repaired outside the subspaces of linear or affine random variables to become distortion risk measures over all random variables. However, we construct an expansion of a $\mathbf{q}$-permutahull that is not a $\mathbf{q}$-permutahull, but does lead to a distortion risk measure over the subspace of linear random variables. This shows that there are indeed measures that are distortion risk measures over the space of linear random variables that

Email addresses: `glagos@gatech.edu` (Guido Lagos), `daespino@dii.uchile.cl` (Daniel Espinoza), `eduardo.moreno@uai.cl` (Eduardo Moreno), `jvielma@mit.edu` (Juan Pablo Vielma)

cannot be extended to a distortion risk measure over all random variables. We further motivate the study of expansions of **q**-permutahulls by giving some preliminary computational results illustrating that the resulting risk measures behave different than the ones associated to the original **q**-permutahull and, in particular, seem to have more stable estimators.

The rest of this paper is organized as follows. In Section 2 we give some notation and background on risk measures and robust optimization. Our results begin in Section 3 where we study some simple properties of the natural extension of the **q**-permutahull to non-atomic distributions. Then in Section 4 we study coherent and distortion risk measures restricted to the subspace of linear or affine linear random variables and consider the modification of a risk measure we denote the *epsilon scaling*. In particular, we observe that under mild assumptions the epsilon scaling of a distortion risk measure is a distortion risk measure over the subspace of linear or affine linear random variables. Section 5 continues the study of restricted risk measures and the epsilon scaling on finite probability spaces and contains the main results of this paper. In particular we construct an epsilon scaling of a distortion risk measure that is a distortion risk measure over the subspace of linear random variables and is not associated to any **q**-permutahull. This gives an example of a distortion risk measure over the space of linear random variables that cannot be extended to a distortion risk measure over the space of all random variables. Finally, in Section 6 we present some preliminary computational results that suggest the *epsilon scaling* could be useful to useful to mitigate estimation errors.

2. Notation and Background on Risk Measure and Robust Optimization

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 $\mathrm{aff}(S)$, $\mathrm{conv}(S)$ and $\overline{\mathrm{conv}}(S)$ its affine, convex and closed convex hull respectively. We also let $\mathrm{lin}(S)$ be the linear space spanned by S and $\mathrm{ri}(S)$ the relative interior of S. For a given convex set C we denote by $\mathrm{ext}(C)$ the set of its extreme points.

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[1] that are an outcome of the uncertain parameter $\boldsymbol{\omega} \in \Omega$. We use a tilde to identify random variables and vectors as in $\tilde{v}$ and $\tilde{\boldsymbol{\omega}}$.

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*: $\rho(t\tilde{v} + (1-t)\tilde{u}) \leqslant t\rho(\tilde{v}) + (1-t)\rho(\tilde{u})$ for all $\tilde{u}, \tilde{v} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ and $t \in [0,1]$.*

(C2) Positive Homogeneity*: $\rho(t\tilde{v}) = t\rho(\tilde{v})$ for all $\tilde{u} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ and $t > 0$.*

(C3) Translation Invariance*: $\rho(t + \tilde{v}) = t + \rho(\tilde{v})$ for all $\tilde{u} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ and $t \in \mathbb{R}$.*

(C4) Monotonicity*: $\rho(\tilde{v}) \leqslant \rho(\tilde{u})$ for all $\tilde{u}, \tilde{v} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ such that $\tilde{v} \leqslant \tilde{u}$ a.s.*

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

Theorem 2.2. *A function $\rho : L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ is a* coherent risk measure *if and only if* [2] *there exists a family of probability measures $\mathcal{Q}$ such that $\mathbb{Q} \ll \mathbb{P}$ for all $\mathbb{Q} \in \mathcal{Q}$ and*

$$\rho(\tilde{v}) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}(\tilde{v}) \tag{1}$$

[1] Risk measures can be defined in $L^p(\Omega, \mathcal{F}, \mathbb{P})$ for other values of p. For simplicity we only consider $p = 1$ in this paper.

[2] More precisely, the representation in this Theorem is for proper lower semi-continuous coherent risk measures $\rho : L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$. For simplicity we will only consider real valued risk measures in this paper.

A relation between risk measures and robust uncertainty sets emerges when $\Omega \subseteq \mathbb{R}^d$ and when we focus on random variables that are affine-linear or linear functions of random vector $\tilde{\boldsymbol{\omega}}$. That is if we consider $\mathcal{V} := \left\{\tilde{v} : \exists(\boldsymbol{x}, x_0) \in \mathbb{R}^d \times \mathbb{R} \text{ such that } \tilde{v} = \tilde{v}_{\boldsymbol{x},x_0} := \boldsymbol{x}'\tilde{\boldsymbol{\omega}} + x_0\right\}$ and $\mathcal{V}_0 := \left\{\tilde{v} : \exists \boldsymbol{x} \in \mathbb{R}^d \text{ such that } \tilde{v} = \tilde{v}_{\boldsymbol{x}} := \boldsymbol{x}'\tilde{\boldsymbol{\omega}}\right\}$. For this we need to assure that such functions are at least $\mathcal{F}$-measurable. Hence, for simplicity, when dealing with general probability spaces we will assume $\Omega = \mathbb{R}^d$, that $\mathcal{F}$ is the σ-algebra of all Lebesque measurable sets of $\mathbb{R}^d$ and that $\mathbb{P}$ is such that $\mathcal{V}$, $\mathcal{V}_0 \subseteq L_1(\Omega, \mathcal{F}, \mathbb{P})$. This assumptions are not overly restrictive as they include most distributions of interest including most continuous and discrete distributions and their mixtures.

The effect of a risk measure on $\mathcal{V}$ and $\mathcal{V}_0$ can then be interpreted using the language of robust optimization as follows. Let ρ be a coherent risk measure and let $\mathcal{Q}$ be a family of probability measures satisfying (1). Then, for any $\tilde{v}_{\boldsymbol{x},x_0} \in \mathcal{V}$ we have

$$\rho(\tilde{v}_{\boldsymbol{x},x_0}) = \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[\tilde{v}_{\boldsymbol{x},x_0}] = x_0 + \sup_{\boldsymbol{\omega}\in\mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{\omega}. \tag{2}$$

where $\mathcal{U}(\rho) := \overline{\text{conv}}(\{\mathbb{E}_{\mathbb{Q}}[\boldsymbol{\omega}] : \mathbb{Q} \in \mathcal{Q}\})$ is a convex set satisfying $\mathcal{U}(\rho) \subseteq \overline{\text{conv}}(\text{supp}(\mathbb{P}))$, where $\text{supp}(\mathbb{P})$ is the support of $\mathbb{P}$. 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 convex set* $\mathcal{U}(\rho) \subseteq \mathbb{R}^d$ *such that*

$$\rho(\tilde{v}_{\boldsymbol{x},x_0}) = x_0 + \sup_{\boldsymbol{\omega}\in\mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{\omega}. \tag{3}$$

for every $\tilde{v}_{\boldsymbol{x},x_0} \in \mathcal{V}$. *Furthermore,* ρ *additionally satisfies property (C4) if and only if* $\mathcal{U}(\rho) \subseteq \overline{\text{conv}}(\text{supp}(\mathbb{P}))$.

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 $\rho(\boldsymbol{x}'\tilde{\boldsymbol{\omega}})$ is a continuous sub-linear function of $\boldsymbol{x}$. Then $\rho(\boldsymbol{x}'\tilde{\boldsymbol{\omega}}) = \sup_{\boldsymbol{\omega}\in\mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{\omega}$ for some closed convex set $\mathcal{U}(\rho)$ (Theorem 3.1.1 of Hiriart-Urruty and Lemaréchal (2001)). The implication then follows from the translation invariance property. The reverse implication is straightforward.

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

$$\sup_{\boldsymbol{\omega}\in\mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{\omega} = \rho(\boldsymbol{x}'\tilde{\boldsymbol{\omega}}) \leqslant \rho\left(\sup_{\boldsymbol{\omega}\in\text{supp}(\mathbb{P})} \boldsymbol{x}'\boldsymbol{\omega}\right) = \sup_{\boldsymbol{\omega}\in\text{supp}(\mathbb{P})} \boldsymbol{x}'\boldsymbol{\omega}$$

For the reverse implication note that if $\mathcal{U}(\rho) \subseteq \overline{\text{conv}}(\text{supp}(\mathbb{P}))$ and $\boldsymbol{x}'\tilde{\boldsymbol{\omega}} + x_0 \leqslant 0\, a.s.$ then $\rho(\boldsymbol{x}'\tilde{\boldsymbol{\omega}} + x_0) = x_0 + \sup_{\boldsymbol{\omega}\in\mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{\omega} \leqslant x_0 + \sup_{\boldsymbol{\omega}\in\text{supp}(\mathbb{P})} \boldsymbol{x}'\boldsymbol{\omega} \leqslant 0$. Together with sub-additivity of ρ this implies that if $\tilde{v}_{\boldsymbol{x},x_0} \leqslant \tilde{v}_{\boldsymbol{y},y_0}$ then $\rho(\tilde{v}_{\boldsymbol{x},x_0}) \leqslant \rho(\tilde{v}_{\boldsymbol{x}-\boldsymbol{y},x_0-y_0}) + \rho(\tilde{v}_{\boldsymbol{y},y_0}) \leqslant \rho(\tilde{v}_{\boldsymbol{y},y_0})$. ☐

Note that in the proof of Theorem 2.3 necessity of $\mathcal{U}(\rho) \subseteq \overline{\text{conv}}(\text{supp}(\mathbb{P}))$ was because of dominance between a constant ($\sup_{\boldsymbol{\omega}\in\text{supp}(\mathbb{P})} \boldsymbol{x}'\boldsymbol{\omega}$) and a linear ($\boldsymbol{x}'\tilde{\boldsymbol{\omega}}$) 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 $\mathcal{V}_0$).

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 family of absolutely continuous probability distribution $\mathcal{Q}$ induces a convex uncertainty set $\mathcal{U}(\mathcal{Q}) := \overline{\text{conv}}(\{\mathbb{E}_{\mathbb{Q}}[\boldsymbol{\omega}] : \mathbb{Q} \in \mathcal{Q}\}) \subseteq \overline{\text{conv}}(\text{supp}(\mathbb{P}))$, the converse does not always hold. For instance, if we let $\mathbb{P}$ be the uniform probability distribution over a compact convex set C and $\boldsymbol{\omega}^0$ be an extreme point of C, we have that $\mathcal{U} = \{\boldsymbol{\omega}^0\}$ is a convex uncertainty set that will induce a coherent risk measure over $\mathcal{V}$ through (3). However, there is no $\mathbb{Q} \ll \mathbb{P}$ such that $\mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{\omega}}) = \boldsymbol{\omega}^0$ and hence by Theorem 2.2 and (2) there cannot be a coherent risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ that coincides with this measure in $\mathcal{V}$. Therefore there are coherent risk measures over $\mathcal{V}$ that cannot be extended to coherent measures over $L_1(\Omega, \mathcal{F}, \mathbb{P})$. Fortunately, as we will see in Section 5.1, if $\mathbb{P}$ is a finite probability distrbution then any coherent risk measures over $\mathcal{V}$ can be extended to a coherent risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$. However, we will also show that this *coherent extension* can fail to preserve additional desirable properties such as the ones in the following definition.

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 properties.*

(D1) Comonotonicity*: $\rho(\tilde{u} + \tilde{v}) = \rho(\tilde{u}) + \rho(\tilde{v})$ for all $\tilde{u}, \tilde{v}$ such that $\left(\tilde{u}(\boldsymbol{\omega}^1) - \tilde{u}(\boldsymbol{\omega}^2)\right)\left(\tilde{v}(\boldsymbol{\omega}^1) - \tilde{v}(\boldsymbol{\omega}^2)\right) \geqslant 0$.*

(D2) Law Invariance*: $\rho(\tilde{u}) = \rho(\tilde{v})$ for all $\tilde{u}, \tilde{v}$ that have the same distribution under $\mathbb{P}$.*

Example 2.1. *One of the most well known distortion risk measure is the* conditional value at risk *which is given by*

$$\mathrm{CVaR}_{\delta}(\tilde{v}) := \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{\delta} \mathbb{E}[(\tilde{v} - t)^+] \right\} \tag{4}$$

In the finite support and non-atomic cases we can get a more precise description of $\mathcal{U}(\rho)$ when ρ is a distortion risk measure.

2.1. Finite Support Case

We now consider the case of a uniform probability distribution with finite support. That is $\mathrm{supp}(\mathbb{P}) = \{\boldsymbol{\omega}^1, \ldots, \boldsymbol{\omega}^N\} \subseteq \mathbb{R}^d$ and $\mathbb{P}(\boldsymbol{\omega} = \boldsymbol{\omega}^i) = \frac{1}{N}$ for all $i = 1, \ldots, N$. In this setting we may assume $\Omega = \{\boldsymbol{\omega}^1, \ldots, \boldsymbol{\omega}^N\}$ and that $\mathcal{F}$ is the σ-algebra of all subsets of Ω. Furthermore, under these assumptions every probability measure $\mathbb{Q}$ satisfies $\mathbb{Q} \ll \mathbb{P}$ and can be represented as a vector $\boldsymbol{q} \in \Delta^N := \{\boldsymbol{q} \in \mathbb{R}^N_+ : \sum_{i=1}^N q_i = 1\}$, where $\mathbb{Q}(\boldsymbol{\omega} = \boldsymbol{\omega}^i) = q_i$ for all $i = 1, \ldots, N$. Similarly every random variable $\tilde{v} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ is representable by means of a vector $\boldsymbol{v} \in \mathbb{R}^N$ where $v_i := \tilde{v}(\boldsymbol{\omega}^i)$ for all $i = 1, \ldots, N$. 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 $\mathcal{V}_0$. By letting $\boldsymbol{O} := \left[\boldsymbol{\omega}^1 | \ldots | \boldsymbol{\omega}^N\right]$, we have that any $\tilde{v}_{\boldsymbol{x}, x_0} \in \mathcal{V}$ is representable by a vector $\boldsymbol{v}_{\boldsymbol{x}, x_0} = \boldsymbol{O}'\boldsymbol{x} + x_0 \boldsymbol{e}$, where $(\boldsymbol{v}_{\boldsymbol{x}, x_0})_i = \tilde{v}_{\boldsymbol{x}, x_0}(\boldsymbol{\omega}^i) = (\boldsymbol{\omega}^i)'\boldsymbol{x} + x_0$ is the i-th scenario for $\tilde{v}_{\boldsymbol{x}, x_0}$.

In this setting we have the following version of Theorem 2 for distortion risk measures introduced in Bertsimas and Brown (2009).

Theorem 2.5. *ρ is a distortion risk measure on $L_1(\Omega, \mathcal{F}, \mathbb{P})$ if and only if there exists $\boldsymbol{q} \in \Delta^N$ such that*

$$\rho(\tilde{v}) = \max_{\sigma \in S_N} \sum_{i=1}^{N} q_{\sigma(i)} v_i \tag{5}$$

where S_N is the group of permutations of N elements. Furthermore, in this representation we can additionally choose $\boldsymbol{q} \in \widehat{\Delta}^N := \{\boldsymbol{q} \in \mathbb{R}^N_+ : \sum_{i=1}^N q_i = 1,\ q_1 \geqslant \ldots \geqslant q_N\}$, in which case

$$\rho(\tilde{v}) = \sum_{i=1}^{N} q_i v_{[i]} = \boldsymbol{q}'\boldsymbol{v}_{[\cdot]} \tag{6}$$

where $\boldsymbol{v}_{[\cdot]} := (v_{[1]}, \ldots, v_{[N]})'$, and $v_{[i]}$ are the decreasing order statistics of v_i, i.e. $v_{[1]} \geqslant \ldots \geqslant v_{[N]}$.

Bertsimas and Brown also developed a very interesting geometric interpretation of distortion risk measures by introducing a version of Theorem 2.3 as follows.

Definition 2.6. *For $\boldsymbol{q} \in \mathbb{R}^N$ and $\Omega = \{\boldsymbol{\omega}^1, \ldots, \boldsymbol{\omega}^N\} \subseteq \mathbb{R}^d$ let the $\boldsymbol{q}-$permutahull of Ω be given by*

$$\Pi_{\boldsymbol{q}}(\Omega) := \operatorname{conv}\left(\left\{\sum_{i=1}^N q_{\sigma(i)}\boldsymbol{\omega}^i : \sigma \in S_N\right\}\right) \tag{7}$$

where S_N is the group of permutations of N elements.

Now, note that (6) is equivalent to $\rho(\widetilde{v}) = \sup_{\sigma \in S_N} \sum_{i=1}^N q_{\sigma(i)}\widetilde{v}(\boldsymbol{\omega}^i)$. Hence, similarly to (2), if ρ is a distortion risk measure and $\widetilde{v}_{\boldsymbol{x},x_0} \in \mathcal{V}$ we have

$$\rho(\widetilde{v}_{\boldsymbol{x},x_0}) = \sup_{\sigma \in S_N} \sum_{i=1}^N q_{\sigma(i)} \left((\boldsymbol{\omega}^i)'\boldsymbol{x} + x_0\right) = x_0 + \sup_{\sigma \in S_N} \sum_{i=1}^N q_{\sigma(i)}\ (\boldsymbol{\omega}^i)'\boldsymbol{x} = x_0 + \sup_{\boldsymbol{\omega} \in \Pi_q(\Omega)} \boldsymbol{\omega}'\boldsymbol{x}.$$

This leads to the following version Theorem 2.3.

Theorem 2.7. [3]*If ρ is a distortion risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ then there exists $\boldsymbol{q} \in \widehat{\Delta}^N$ such that $\mathcal{U}(\rho) = \Pi_q(\Omega)$.*

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

$$\boldsymbol{h}_j^\delta := \begin{cases} \frac{1}{\delta N} & j \leqslant \delta N \\ 0 & \textit{otherwise} \end{cases}. \tag{8}$$

2.2. Non-Atomic Case

For non-atomic probability spaces a characterization similar to the one in Section 2.1 was given in Shapiro (2011). To describe it we assume $(\Omega, \mathcal{F}, \mathbb{P})$ is such that $\Omega = \mathbb{R}^d$, $\mathcal{F}$ is the σ-algebra of all Lebesque measurable sets of $\mathbb{R}^d$ and that $\mathbb{P}$ is non-atomic. In this case the appropriate version of permutations of Ω are the *measure preserving maps.*

Definition 2.8. *A one-to-one map $T : \Omega \to \Omega$ is* measure preserving *if and only if*

(i) $T(A) \in \mathcal{F}$ for every $A \in \mathcal{F}$.

(ii) $P(A) = P(T(A))$ for every $A \in \mathcal{F}$.

We let $\mathcal{T}$ be the set of all measure preserving mappings.

The following analog of Theorem 2.5 is shown in in Shapiro (2011).

Theorem 2.9. *ρ is a distortion risk measure on $L_1(\Omega, \mathcal{F}, \mathbb{P})$ if and only if there exists*

$$g_\rho \in \Delta := \left\{g \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}) \,:\, 0 \leqslant g \quad a.s., \quad \int_\Omega g(\boldsymbol{\omega}) d\mathbb{P}(\boldsymbol{\omega}) = 1\right\} \tag{9}$$

such that

$$\rho(\tilde{v}) = \sup_{T \in \mathcal{T}} \int_\Omega \tilde{v}(\boldsymbol{\omega}) g_\rho(T(\boldsymbol{\omega})) d\mathbb{P}(\boldsymbol{\omega}) \tag{10}$$

$$= \sup_{g \in \mathcal{G}_\rho} \int_\Omega \tilde{v}(\boldsymbol{\omega}) g(\boldsymbol{\omega}) d\mathbb{P}(\boldsymbol{\omega}) \tag{11}$$

were $\mathcal{G}_\rho := \overline{\operatorname{conv}}\left(\{g_\rho \circ T \,:\, T \in \mathcal{T}\}\right)$ is a weakly- compact set. Hence, for any $\tilde{v} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ the supremum in (11) is attained at some $g_{\tilde{v}} \in \mathcal{G}_\rho$ and we have*

$$\rho(\tilde{v}) = \int_\Omega \tilde{v}(\boldsymbol{\omega}) \boldsymbol{g}_{\tilde{v}}(\boldsymbol{\omega}) d\mathbb{P}(\boldsymbol{\omega}). \tag{12}$$

[3]This is essentially Corollary 4.3 of Bertsimas and Brown (2009).

Example 2.3. *For $\rho = \mathrm{CVaR}_\delta$ we have characterization* (10) *with $g_{\mathrm{CVaR}_\delta} = 1_A$ for any $A \in \mathcal{F}$ with $\mathbb{P}(A) = \delta$. Furthermore,*

$$\mathcal{G}_{\mathrm{CVaR}_\delta} = \left\{ g \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}) \,:\, 0 \leqslant g \leqslant 1/\delta \quad a.s., \quad \int_\Omega g(\boldsymbol{\omega}) d\mathbb{P}(\boldsymbol{\omega}) = 1 \right\},$$

and $\mathrm{ext}\left(\mathcal{G}_{\mathrm{CVaR}_\delta}\right) = \left\{ 1_{T(A)} \,:\, T \in \mathcal{T} \right\}$. *Hence, by the Bauer maximum principle, we have that for every $\tilde{v} \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ there exists $A_{\tilde{v}} \in \mathcal{F}$ with $\mathbb{P}(A_{\tilde{v}}) = \delta$ such that*

$$\mathrm{CVaR}_\delta(\tilde{v}) = \frac{1}{\delta} \int_\Omega \tilde{v}(\boldsymbol{\omega}) \mathbf{1}_{A_{\tilde{v}}}(\boldsymbol{\omega}) d\mathbb{P}(\boldsymbol{\omega}) \tag{13}$$

3. Distortion Uncertainty Sets for Non-Atomic Distributions

Using Theorem 2.9 we can easily extend the geometric characterization in Theorem 2.7 to the non-atomic case as follows.

Definition 3.1. *For a given $g \in \Delta$, let*

$$\Pi_g := \overline{\mathrm{conv}} \left(\left\{ \int_\Omega \boldsymbol{\omega} g(T(\boldsymbol{\omega})) d\mathbb{P}(\boldsymbol{\omega}) : T \in \mathcal{T} \right\} \right) \tag{14}$$

Corollary 3.2. *If ρ is a distortion risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ then there exists $g \in \Delta$ such that $\mathcal{U}(\rho) = \Pi_g$.*

Characterizing sets Π_g can be quite involved. However, if $\mathbb{P}$ is the uniform distribution over a convex set C and $g = \mathbf{1}_A$ for $A \in \mathcal{F}$ with $\mathbb{P}(A) = \delta$ we can use the following lemma to simplify the construction of Π_g.

Lemma 3.3. *Let $\mathbb{P}$ be the uniform distribution over a compact convex set C and let $A \in \mathcal{F}$ with $\mathbb{P}(A) = \delta$. For $\boldsymbol{x} \in S^{n-1} := \{\boldsymbol{x} \in \mathbb{R}^n \,:\, \|\boldsymbol{x}\|_2 = 1\}$ and $\delta \in (0, 1/2]$ let $H_{\boldsymbol{x},\delta} := \{\boldsymbol{\omega} \in \mathbb{R}^n \,:\, \boldsymbol{x}'\boldsymbol{\omega} \leqslant l(\boldsymbol{x}, \delta)\}$ where $l(\boldsymbol{x}, \delta)$ is such that $\mathbb{P}\left(C \cap H_{\boldsymbol{x},\delta}\right) = \delta$. Then*

$$\Pi_{\mathbf{1}_A} = \overline{\mathrm{conv}} \left(\left\{ \frac{1}{\delta} \int_{C \cap H_{\boldsymbol{x},\delta}} \boldsymbol{\omega} d\mathbb{P}(\boldsymbol{\omega}) \,:\, \boldsymbol{x} \in S^{n-1} \right\} \right). \tag{15}$$

Proof. Let $E = \left\{ \frac{1}{\delta} \int_{C \cap H_{\boldsymbol{x},\delta}} \boldsymbol{\omega} d\mathbb{P}(\boldsymbol{\omega}) \,:\, \boldsymbol{x} \in S^{n-1} \right\} \subseteq C$.

$\mathrm{CVaR}_\delta(\tilde{v}_{\boldsymbol{x}}) = \sup_{\boldsymbol{\omega} \in \mathcal{U}(\mathrm{CVaR}_\delta)} \boldsymbol{x}'\boldsymbol{\omega}$ is the support function of $\mathcal{U}(\mathrm{CVaR}_\delta) = \Pi_{\mathbf{1}_A}$. Hence to show (15) it suffices to prove that CVaR_δ is also the support function of $\overline{\mathrm{conv}}(E)$ (e.g. Corollary 3.1.2. Hiriart-Urruty and Lemaréchal (2001)).

Now let $\mathbb{F}_{\boldsymbol{x}}$ be the probability distribution of random variable $\widetilde{v}_{\boldsymbol{x}}$ and $\mathrm{VaR}_\delta(\widetilde{v}_{\boldsymbol{x}}) := \inf \left\{ t \in \mathbb{R} \,:\, \mathbb{F}\left((-\infty, t]\right) \geqslant 1 - \delta \right\}$ be the *Value-at-Risk* of $\widetilde{v}_{\boldsymbol{x}}$. We then have

$$\begin{aligned} \mathrm{CVaR}_\delta(\tilde{v}_{\boldsymbol{x}}) &= \frac{1}{\delta} \int_{\mathrm{VaR}_\delta(\tilde{v}_{\boldsymbol{x}})}^{+\infty} r d\mathbb{F}(r) \\ &= \frac{1}{\delta} \int_{C \cap H_{\boldsymbol{x},\delta}} \boldsymbol{x}'\boldsymbol{\omega} d\mathbb{P}(\boldsymbol{\omega}) \\ &= \boldsymbol{x}' \left(\frac{1}{\delta} \int_{C \cap H_{\boldsymbol{x},\delta}} \boldsymbol{\omega} d\mathbb{P}(\boldsymbol{\omega}) \right) \end{aligned}$$

Then $\mathrm{CVaR}_\delta(\tilde{v}_{\boldsymbol{x}}) = \max_{\boldsymbol{\omega} \in E} \boldsymbol{x}'\boldsymbol{\omega}$. The result follows by noting that by Minkowski's Theorem we have $\mathrm{ext}(\overline{\mathrm{conv}}(E)) \subseteq E$ and hence the supremum of $\boldsymbol{x}'\boldsymbol{\omega}$ over $\boldsymbol{\omega} \in \overline{\mathrm{conv}}(E)$ is attained at $\boldsymbol{\omega} \in E$. $\square$

Lemma 3.3 can be used to construct $\mathcal{U}(\mathrm{CVaR}_\delta)$ for some simple distributions.

Example 3.1. *Let $\mathbb{P}$ be the uniform distribution over the unit square. Using Lemma 3.3 and Example 2.3 we can see that the uncertainty set associated to* CVaR_δ *is the convex hull of the centroids of a series of triangles and quadrilaterals. More specifically* $\mathcal{U}(\mathrm{CVaR}_\delta)$ *is symmetric over each orthant and its boundary on the non-negative orthant is given by the parametric curve*

$$\left\{\left(1-\frac{\cot^2(t)}{12\delta}-\delta,-\frac{\cot(t)}{6\delta}\right) \,:\, t\in[0,s(\delta)]\right\}\cup\left\{\left(1-\frac{\sqrt{8\delta\tan(t)}}{3},1-\frac{\sqrt{8\delta\cot(t)}}{3}\right),:\, t\in[s(\delta),e(\delta)]\right\}$$
$$\cup\left\{\left(\frac{\cot(t)}{6\delta},1-\frac{\cot^2(t)}{12\delta}-\delta\right) \,:\, t\in[e(\delta),\pi/2]\right\}$$

where $s(\delta) = 2\cot^{-1}\left(\frac{2\delta}{-1+\delta\sqrt{\frac{1}{\delta^2}+4}}\right)$ *and* $e(\delta) = -2\tan^{-1}\left(2\delta-\sqrt{4\delta^2+1}\right)$. *Figure 1 shows* $\mathcal{U}(\mathrm{CVaR}_\delta)$ *for* $\delta\in\{0.01,0.1,0.4\}$

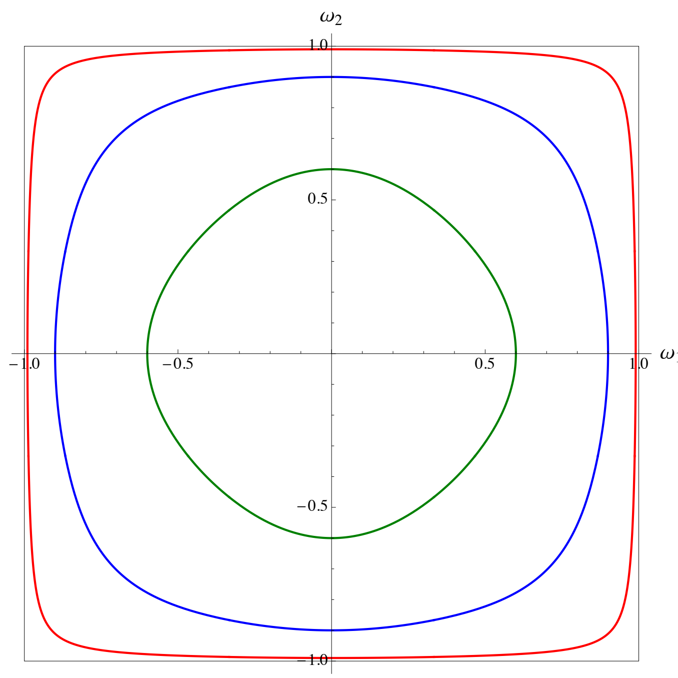

Figure 1: $\mathcal{U}(\mathrm{CVaR}_\delta)$ for uniform distribution on the unit square.

It is unlikely that we can get closed form expressions similar to Example 3.1 for the n-dimensional cube or other general high dimensional convex sets as this would require multidimensional integration. However, we can give closed form expressions for the uncertainty sets of distortion risk measures for *elliptical* distributions.

Lemma 3.4. *Let* $\mu\in\mathbb{R}^d$, $B\in\mathbb{R}^{d\times d}$ *be a non-singular matrix and* $\mathbb{P}$ *be any probability distribution such that* $\tilde{u}_{\boldsymbol{x}} := \boldsymbol{x}'B^{-1}(\tilde{\boldsymbol{\omega}}-\mu)$ *has the same continuous probability distribution for every* $\boldsymbol{x}\in S^{n-1} := \{\boldsymbol{x}\in\mathbb{R}^n \,:\, \|\boldsymbol{x}\|_2 = 1\}$ *(e.g.* $\mathbb{P}$ *is the uniform distribution over the ellipsoid* $\left\{\boldsymbol{\omega}\in\mathbb{R}^d \,:\, \|B(\boldsymbol{\omega}-\mu)\|_2\leqslant 1\right\}$ *or* $\mathbb{P}=\mathcal{N}(\mu,BB')$*). Then, for any distortion risk measure* ρ *we have*

$$\mathcal{U}(\rho) = \left\{\boldsymbol{\omega}\in\mathbb{R}^n \,:\, \left\|B^{-1}(\boldsymbol{\omega}-\mu)\right\|_2\leqslant\rho(\tilde{u}_{\boldsymbol{x}_0})\right\} \tag{16}$$

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

Proof. Because $\rho(\widetilde{v}_{\boldsymbol{x}}) = \sup_{\boldsymbol{\omega}\in\mathcal{U}(\rho)} \boldsymbol{x}'\boldsymbol{\omega}$ we have that $\mathcal{U}(\rho) = \left\{\boldsymbol{\omega} \in \mathbb{R}^d : \boldsymbol{x}'\boldsymbol{\omega} \leqslant \rho(\widetilde{v}_{\boldsymbol{x}}) \text{ for all } x \in S^{n-1}\right\}$ (e.g. Corollary 3.1.2. Hiriart-Urruty and Lemaréchal (2001)). We then have that

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

The first equality comes from translation invariance 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 $\widetilde{u}_{\boldsymbol{x}}$. □

4. Restricted Distortion Risk Measures and Epsilon Scalings

We consider the following well known modification of a risk measure (e.g. see Lagos et al. (2011) and equation (6.68) in Shapiro et al. (2009)).

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

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 $\epsilon \in [0,1]$, $\widehat{\rho}_\epsilon$ is also a distortion risk measure over $\mathcal{S}$. It is also easy to see that for $\epsilon \geqslant 0$ the only property that $\widehat{\rho}_\epsilon$ may fail to inherit is monotonicity. If $\mathcal{S} = \mathcal{V}$, we can further see that the uncertainty set associated to $\widehat{\rho}_\epsilon$ is $\mathcal{U}(\widehat{\rho}_\epsilon) = \overline{\boldsymbol{\omega}} + \epsilon(\mathcal{U}(\rho) - \overline{\boldsymbol{\omega}})$ where $\overline{\boldsymbol{\omega}} := \mathbb{E}[\widetilde{\boldsymbol{\omega}}]$. If $\overline{\boldsymbol{\omega}} \in \mathcal{U}(\rho)$ then $\mathcal{U}(\widehat{\rho}_\epsilon)$ is a scaling of $\mathcal{U}(\rho)$ around $\overline{\boldsymbol{\omega}}$, which allows us to give simple conditions for the monotonicity of $\widehat{\rho}_\epsilon$. However, not all uncertainty sets contain the mean of $\widetilde{\boldsymbol{\omega}}$. Fortunately we can show that the mean is always contained in the uncertainty sets associated to distortion risk measures by using the following proposition whose proof for discrete and non-atomic distributions can be found in Bertsimas and Brown (2009) and Shapiro (2011) respectively.

Proposition 4.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be such that $\mathbb{P}$ is finite or non-atomic and ρ be a distortion risk measure. Then there exists a probability measure $\mathbb{F}$ on $[0,1]$ with zero mass at 0 such that*

$$\rho(\widetilde{v}) = \int_0^1 \mathrm{CVaR}_\delta(\widetilde{v})\, d\mathbb{F}(\delta) \tag{17}$$

Corollary 4.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be such that $\mathbb{P}$ is finite or non-atomic and ρ be a distortion risk measure. Then $\overline{\boldsymbol{\omega}} \in \mathcal{U}(\rho)$.*

Proof. For any $\boldsymbol{x} \in \mathbb{R}^n$ we have $\mathbb{E}(\widetilde{v}_{\boldsymbol{x}}) \leqslant \mathrm{CVaR}_\delta(\widetilde{v}_{\boldsymbol{x}})$ for all $\delta \in (0,1]$. Hence by Proposition 4.2 we have $\overline{\boldsymbol{\omega}}'\boldsymbol{x} = \mathbb{E}(\widetilde{v}_{\boldsymbol{x}}) \leqslant \rho(\widetilde{v}_{\boldsymbol{x}})$. The result follows by noting that $\mathcal{U}(\rho) = \left\{\boldsymbol{\omega} \in \mathbb{R}^d : \boldsymbol{x}'\boldsymbol{\omega} \leqslant \rho(\widetilde{v}_{\boldsymbol{x}}) \text{ for all } x \in S^{n-1}\right\}$ □

Because for distortion risk measures $\mathcal{U}(\widehat{\rho}_\epsilon)$ is a scaling of $\mathcal{U}(\rho)$ around $\overline{\boldsymbol{\omega}}$ we refer to $\widehat{\rho}_\epsilon$ as the *epsilon scaling* of ρ. Now, for $\epsilon \in [0,1]$ we have that $\mathcal{U}(\widehat{\rho}_\epsilon)$ is contraction of $\mathcal{U}(\rho)$ that is always contained in $\mathcal{U}(\rho)$ and hence in $\overline{\mathrm{conv}}(\mathrm{supp}(\mathbb{P}))$. Similarly, for $\epsilon > 1$ we have that $\mathcal{U}(\widehat{\rho}_\epsilon)$ is an expansion of $\mathcal{U}(\rho)$ and hence might not be contained $\overline{\mathrm{conv}}(\mathrm{supp}(\mathbb{P}))$. However, if the containment does hold, Theorem 2.3 implies $\widehat{\rho}_\epsilon$ will be a distortion risk measure on $\mathcal{V}$.

Proposition 4.4. *Let $\epsilon \geqslant 0$, ρ be a distortion risk measure over $\mathcal{V}$. If $\overline{\boldsymbol{\omega}} + \epsilon(\mathcal{U}(\rho) - \overline{\boldsymbol{\omega}}) \subseteq \overline{\text{conv}}(\text{supp}(\mathbb{P}))$ Then $\widehat{\rho}_\epsilon$ is a distortion risk measure over $\mathcal{V}$.*

Proof. The result follows from the fact that any linear combination of measures satisfying (D1) and (D2) also satisfies them and from Theorem 2.3. □

If $\mathcal{S} = \mathcal{V}_0$, $\widehat{\rho}_\epsilon$ can be distortion even if the condition of Proposition 4.4 does not hold. Indeed, as the following lemma shows, we sometimes have that $\widetilde{u} \leqslant \widetilde{v}\, a.s.$ implies $\widetilde{u} = \widetilde{v}\, a.s$ for any $\widetilde{u}, \widetilde{v} \in \mathcal{V}_0$ and hence monotonicity is trivially satisfied. Together with Theorem 2.3 this gives another case in which the distortion risk measure properties are inherited by $\widehat{\rho}_\epsilon$

Lemma 4.5. *If $\mathbf{0} \in \text{ri}\,(\text{conv}(\text{supp}(\mathbb{P}))$ then for any $\widetilde{v} \in \mathcal{V}_0$, $\widetilde{v} \geqslant 0\, a.s.$ implies $\widetilde{v} = 0\, a.s.$*

Proof. Consider $\widetilde{v} = \boldsymbol{x}'\boldsymbol{\omega} \geqslant 0, a.s.$. If $\boldsymbol{x} = \mathbf{0}$ the result is direct. For the case $\boldsymbol{x} \neq \mathbf{0}$ we show that $\boldsymbol{x} \in \text{lin}(\text{supp}(\mathbb{P}))^\perp$ as this implies $\widetilde{v} = 0\ a.s.$

Assume for a contradiction that $\boldsymbol{x} \notin \text{lin}(\text{supp}(\mathbb{P}))^\perp$ and let $H := \{\boldsymbol{\omega} \in \mathbb{R}^d : \boldsymbol{x}'\boldsymbol{\omega} < 0\}$. Because of assumption $\boldsymbol{x} \notin \text{lin}(\text{supp}(\mathbb{P}))^\perp$ there exists $\boldsymbol{\omega}^1 \in H \cap \text{supp}(\mathbb{P})$. Because H is an open neighborhood of $\boldsymbol{\omega}^1$ we have $\mathbb{P}(H) > 0$ an hence $\mathbb{P}(H_0) > 0$ where $H_0 = H \cap \text{lin}(\text{supp}(\mathbb{P}))$. Then $\int_{H_0} \boldsymbol{\omega} d\mathbb{P}(\boldsymbol{\omega}) \in \mathbb{P}(H_0)H_0$ and because H_0 is a convex cone we have $\int_{H_0} \boldsymbol{\omega} d\mathbb{P}(\boldsymbol{\omega}) \in H_0$, which implies $\boldsymbol{x}' \int_{H_0} \boldsymbol{\omega} d\mathbb{P}(\boldsymbol{\omega}) < 0$. However, $\widetilde{v} \geqslant 0\, a.s.$ implies $0 \leqslant \int_{H_0} \boldsymbol{x}'\boldsymbol{\omega} d\mathbb{P}(\boldsymbol{\omega}) = \boldsymbol{x}' \int_{H_0} \boldsymbol{\omega} d\mathbb{P}(\boldsymbol{\omega})$, a contradiction.

□

Proposition 4.6. *Let $\mathbf{0} \in \text{ri}\,(\text{conv}(\text{supp}(\mathbb{P}))$ and ρ be a distortion risk measure over $\mathcal{V}_0$. Then $\widehat{\rho}_\epsilon$ is a distortion risk measure over $\mathcal{V}_0$ for any $\epsilon \geqslant 0$.*

Proof. Direct from Lemma 4.5, Theorem 2.3 and the preservation of (D1) and (D2) under linear combinations. □

Note that the condition in Lemma 4.5 holds if $\mathbb{P}$ has positive definite covariance matrix and $\mathbb{E}_{\mathbb{P}}(\widetilde{\boldsymbol{\omega}}) = \mathbf{0}$, but can also hold if $\mathbb{E}_{\mathbb{P}}(\widetilde{\boldsymbol{\omega}}) \neq \mathbf{0}$ or if $\mathbb{P}$ is concentrated in a linear subspace. In the latter case, $\widetilde{v} \geqslant 0\, a.s.$ implies $\widetilde{v} = 0\, a.s.$, but it does not imply $\widetilde{v}$ is identically zero in all $\mathbb{R}^d$.[4] Finally, note that by eliminating the monotonicity condition we obtain the following corollary of Theorem 6.4 of Shapiro et al. (2009)

Corollary 4.7. *A function $\rho : L_1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ satisfies (C1)–(C3) if and only of there exists a family of functions $\mathcal{J} \subseteq \mathcal{A} := \{f \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}) : \int_\Omega f(\boldsymbol{\omega}) d\mathbb{P}(\boldsymbol{\omega}) = 1\}$[5] such that $\rho(\tilde{v}) = \sup_{f \in \mathcal{J}} \int_\Omega \tilde{v}(\boldsymbol{\omega}) f(\boldsymbol{\omega}) d\mathbb{P}(\boldsymbol{\omega})$.*

Then by restricting ρ to $\mathcal{V}_0$ we obtain the analog of (2) given by $\rho(\widetilde{v}_{\boldsymbol{x}}) = \sup_{f \in \mathcal{J}} \int_\Omega \boldsymbol{x}'\boldsymbol{\omega} f(\boldsymbol{\omega}) d\mathbb{P}(\boldsymbol{\omega}) = \sup_{\boldsymbol{\omega} \in \widehat{\mathcal{U}}(\rho)} \boldsymbol{x}'\boldsymbol{\omega}$ where $\widehat{\mathcal{U}}(\rho) := \{\int_\Omega \boldsymbol{\omega} f(\boldsymbol{\omega}) d\mathbb{P}(\boldsymbol{\omega}) : f \in \mathcal{J}\} \subseteq \text{aff}(\text{supp}(\mathbb{P}))$. Hence the risk measures that do not satisfy monotonicity over all $L_1(\Omega, \mathcal{F}, \mathbb{P})$, but are coherent over $\mathcal{V}_0$ through (4.6) might have associated uncertainty sets that are not in $\overline{\text{conv}}(\text{supp}(\mathbb{P}))$, but they will be in $\text{aff}(\text{supp}(\mathbb{P}))$.

A similar adaptation can be done for Theorem 2.9 restricted to linear random variables. However, we now concentrate on finite probability distributions and in particular show that there are risk measures that are distortion risk measures over linear random variables that cannot be extended to distortion risk measures over all random variables.

5. Finite Support Case

We now study conditions for distortion risk measures over linear and affine random variables when $(\Omega, \mathcal{F}, \mathbb{P})$ is a uniform finite probability space such that $\Omega = \{\boldsymbol{\omega}^1, \ldots, \boldsymbol{\omega}^N\}$, $\mathcal{F}$ is the σ-algebra of all subsets of Ω and $\mathbb{P}(\boldsymbol{\omega} = \boldsymbol{\omega}^i) = \frac{1}{N}$.

[4] Consider for example $d = 2$, with $\mathbb{P}((-1, 1)) = \mathbb{P}((1, -1)) = 1/2$ and $\widetilde{v} = \omega_1 + \omega_2$.

[5] It is interesting notice that the proof of necessity of $f \geqslant 0$ for monotonicity in Shapiro et al. (2009) uses indicator random variables which are clearly not linear

Let ρ be a distortion risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$. Then, by Theorem 2.5, $\rho(\tilde{v}) = \sum_{i=1}^{N} q_i v_{[i]} = \boldsymbol{q}'\boldsymbol{v}_{[\cdot]}$ for some $\boldsymbol{q} \in \hat{\Delta}^N := \{\boldsymbol{q} \in \mathbb{R}_+^N : \sum_{i=1}^{N} q_i = 1,\ q_1 \geqslant \ldots \geqslant q_N\}$. Then, by defining $\boldsymbol{q}^\epsilon := \epsilon\ \boldsymbol{q} + (1-\epsilon)\ \boldsymbol{e}_N$, we have $\hat{\rho}_\epsilon(\tilde{v}) \ = \ \epsilon\ \boldsymbol{q}'\boldsymbol{v}_{[\cdot]} \ + \ (1-\epsilon)\ \boldsymbol{e}_N'\boldsymbol{v} = (\boldsymbol{q}^\epsilon)'\ \boldsymbol{v}_{[\cdot]}$. If $\epsilon \geqslant 0$ we have $\sum_{i=1}^{N} q_i^\epsilon = 1$ and $q_1^\epsilon \geqslant \ldots \geqslant q_N^\epsilon$, but not necessarily $q_i^\epsilon \geqslant 0$ for all $i = 1, \ldots, N$. Hence, while we do not have $\boldsymbol{q}^\epsilon \in \hat{\Delta}^N$ we do have $\boldsymbol{q}^\epsilon \in \check{\Delta}^N := \left\{\boldsymbol{q} \in \mathbb{R}^N : \sum_{i=1}^{N} q_i = 1,\ q_1 \geqslant \ldots \geqslant q_N\right\}$. If $q_N \geqslant 0$ then $\boldsymbol{q} \in \hat{\Delta}^N$ and by Theorem 2.5 $\hat{\rho}_\epsilon$ is also a distortion risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ so in particular $\boldsymbol{q}^\epsilon$ is a distortion risk measure for any $\epsilon \in [0,1]$. We will now see that it is possible for $\hat{\rho}_\epsilon$ to be a distortion risk measure over $\mathcal{V}$ even with $q_N < 0$. But before this, we present the following simple Lemma that shows that every risk measures associated to an element in $\check{\Delta}^N$ corresponds to $\hat{\rho}_\epsilon$ for some distortion risk measure ρ.

Lemma 5.1. *Let $\mu(\tilde{\boldsymbol{v}}) \ = \boldsymbol{q}'\ \boldsymbol{v}_{[\cdot]}$ for some $\boldsymbol{q} \in \check{\Delta}^N$. Then $\mu = \hat{\rho}_\epsilon(\tilde{v})$ for some $\varepsilon \geqslant 0$ and distortion risk measure ρ.*

Proof. If $q_N \geqslant 0$ the result holds with $\varepsilon = 0$. For $q_N < 0$, let $\epsilon := 1 - N q_N$ and $\boldsymbol{q}^{DRM} := \frac{1}{\epsilon}\left(\boldsymbol{q} + (\epsilon - 1)\ \boldsymbol{e}_N\right)$. Then $\boldsymbol{q} = \epsilon\ \boldsymbol{q}^{DRM} + (1-\epsilon)\ \boldsymbol{e}_N$, $\epsilon > 1$, $\boldsymbol{q}^{DRM} \in \hat{\Delta}^N$ and the result follows with $\rho(\tilde{v}) = (\boldsymbol{q}^{DRM})'\ \boldsymbol{v}_{[\cdot]}$. □

5.1. Restriction to Linear and Affine-Linear Functions

We begin by studying the restriction of $\hat{\rho}_\epsilon$ to the space of affine-linear random variables $\mathcal{V}$. By noting that for $\boldsymbol{q}^\epsilon := \epsilon\ \boldsymbol{q} + (1-\epsilon)\ \boldsymbol{e}_N$ we have $\Pi_{\boldsymbol{q}^\epsilon} = \overline{\boldsymbol{\omega}} + \epsilon\,(\Pi_{\boldsymbol{q}} - \overline{\boldsymbol{\omega}})$ the permutahull characterization (8) for $\hat{\rho}_\epsilon$ becomes

$$\hat{\rho}_\epsilon\left(\tilde{v}_{\boldsymbol{x},x_0}\right) \quad = \quad x_0 + \sup_{\boldsymbol{\omega} \in \Pi_{\boldsymbol{q}^\epsilon}(\Omega)} \boldsymbol{\omega}'\boldsymbol{x} \tag{18}$$

We then have the following corollary of Proposition 4.4.

Corollary 5.2. *Let $\epsilon \geqslant 0$, $\rho(\tilde{\boldsymbol{v}}) = \boldsymbol{q}'\boldsymbol{v}_{[\cdot]}$ for $\boldsymbol{q} \in \hat{\Delta}^N$ and $\boldsymbol{q}^\epsilon = \epsilon\boldsymbol{q} + (1-\epsilon)\boldsymbol{e}_N \in \check{\Delta}^N$. If $\Pi_{\boldsymbol{q}^\epsilon} \subseteq conv(\Omega)$ then $\hat{\rho}_\epsilon$ is a distortion risk measure over $\mathcal{V}$.*

Even if the conditions of Corollary 5.2 do not hold, we still have that for any $\varepsilon \geqslant 0$ and distortion risk measure ρ, $\hat{\rho}_\epsilon$ will satisfy (C1)–(C3), (D1) and (D2) on all $L_1(\Omega, \mathcal{F}, \mathbb{P})$. If $\Pi_{\boldsymbol{q}^\epsilon} \subseteq \mathrm{conv}(\Omega)$ does indeed hold we have that $\hat{\rho}_\epsilon$ will additionally satisfy (C4), but only over $\mathcal{V}$. However, if $\Pi_{\boldsymbol{q}^\epsilon} \subseteq \mathrm{conv}(\Omega)$ we can also construct a risk measure that satisfies (C1)–(C4) on all $L_1(\Omega, \mathcal{F}, \mathbb{P})$ and coincides with $\hat{\rho}_\epsilon$ in $\mathcal{V}$.

Lemma 5.3. *Let $\epsilon \geqslant 0$, $\rho(\tilde{\boldsymbol{v}}) = \boldsymbol{q}'\boldsymbol{v}_{[\cdot]}$ for $\boldsymbol{q} \in \hat{\Delta}^N$ such that $\Pi_{\boldsymbol{q}^\epsilon} \subseteq conv(\Omega)$ for $\boldsymbol{q}^\epsilon$ as defined earlier. Then, for any $\mathcal{Q} \subseteq \Delta^N$ such that $\boldsymbol{O}\mathcal{Q} \subseteq \Pi_{\boldsymbol{q}^\epsilon}$ and $\mathrm{ext}(\Pi_{\boldsymbol{q}^\epsilon}) \subseteq \boldsymbol{O}\mathcal{Q}$ we have that*

$$\mu_{\mathcal{Q}}(\tilde{v}) := \sup_{\boldsymbol{\lambda} \in \mathcal{Q}} \mathbb{E}_{\boldsymbol{\lambda}}[\tilde{v}] = \sup_{\boldsymbol{\lambda} \in \mathcal{Q}} \boldsymbol{\lambda}'\boldsymbol{v}$$

is a coherent risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mu_{\mathcal{Q}}|_{\mathcal{V}} = \hat{\rho}_\epsilon|_{\mathcal{V}}$

Proof. Coherence is direct from Theorem 2.2 and equivalence in $\mathcal{V}$ follows because restricting $\mu_{\mathcal{Q}}$ to $\mathcal{V}$ we have $\mu_{\mathcal{Q}}(\tilde{v}_{\boldsymbol{x},x_0}) = x_0 + \sup_{\boldsymbol{\lambda} \in \mathcal{Q}} \boldsymbol{\lambda}'\boldsymbol{O}'\boldsymbol{x} = x_0 + \sup_{\boldsymbol{\omega} \in \Pi_{\boldsymbol{q}^\epsilon}} \boldsymbol{\omega}'\boldsymbol{x} = \hat{\rho}_\epsilon(\tilde{v}_{\boldsymbol{x},x_0})$. □

Any *coherent extension* $\mu_{\mathcal{Q}}$ of $\hat{\rho}_\epsilon$ repairs the possible non-coherence of $\hat{\rho}_\epsilon$ outside $\mathcal{V}$. However, the modification of $\hat{\rho}_\epsilon$ outside $\mathcal{V}$ might prevent the extension from satisfying (D1) and (D2) outside of $\mathcal{V}$. Hence, it is not clear if any coherent extension of $\hat{\rho}_\epsilon$ is also a distortion risk measure over all $L_1(\Omega, \mathcal{F}, \mathbb{P})$. We next show that there in fact exist cases where none of the coherent extensions of $\hat{\rho}_\epsilon$ are distortion risk measures over $L_1(\Omega, \mathcal{F}, \mathbb{P})$.

Proposition 5.4. *Let $\epsilon \geqslant 0$ and ρ be the distortion risk measure associated to a probability vector $\boldsymbol{q} \in \hat{\Delta}^N$ through (6) such that $\Pi_{\boldsymbol{q}^\epsilon} \subseteq conv(\Omega)$. Also let $\mathrm{ext}(\Pi_{\boldsymbol{q}^\epsilon}) = \{\boldsymbol{v}^1, \ldots, \boldsymbol{v}^m\}$. Then $\hat{\rho}_\epsilon$ has a coherent extension* *that is a distortion risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ if and only if the following Mixed Integer Programming (MIP) problem has optimal value equal to 0:*

$$
\begin{array}{lll}
\min & \sum_{k=1}^{m} \boldsymbol{e}' \, (\boldsymbol{x}^k + \boldsymbol{y}^k) & \\
s.t. & \boldsymbol{v}^k = \boldsymbol{O}\, \boldsymbol{z}^k + (\boldsymbol{x}^k - \boldsymbol{y}^k) & \forall k \in \{1, \ldots, m\} \\
& \boldsymbol{z}^k = \boldsymbol{G}^k \boldsymbol{e} & \forall k \in \{1, \ldots, m\} \\
& \boldsymbol{G}^k \leqslant \boldsymbol{P}^k & \forall k \in \{1, \ldots, m\} \\
& G^k_{i,j} \leqslant r_j & \forall i, j \in \{1, \ldots, N\},\, k \in \{1, \ldots, m\} \\
& P^k_{i,j} + r_j \leqslant G^k_{i,j} + 1 & \forall i, j \in \{1, \ldots, N\}, k \in \{1, \ldots, m\} \\
& \boldsymbol{G}^k \in [0,1]^{N \times N} & \forall k \in \{1, \ldots, m\} \\
& \boldsymbol{x}^k,\ \boldsymbol{y}^k \in \mathbb{R}^N_+ & \forall k \in \{1, \ldots, m\} \\
& \boldsymbol{P}^k \in \{0,1\}^{N \times N} & \forall k \in \{1, \ldots, m\} \\
& \boldsymbol{P}^k \boldsymbol{e} = \boldsymbol{e} & \forall k \in \{1, \ldots, m\} \\
& \boldsymbol{e}' \boldsymbol{P}^k = \boldsymbol{e}' & \forall k \in \{1, \ldots, m\} \\
& \sum_{i=1}^{N} r_i = 1 & \\
& \boldsymbol{r} \in \mathbb{R}^N_+ &
\end{array}
\tag{19}
$$

Proof. From Theorem 2.7 and (18) we have that there exists a distortion risk measure μ over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ that is equal to $\widehat{\rho}_\epsilon$ in $\mathcal{V}$ if and only if there exists $\boldsymbol{r} \in \Delta^N$ such that $\Pi_{\boldsymbol{r}} = \Pi_{\boldsymbol{q}^\epsilon}$. For this to hold it is necessary $S := \{\boldsymbol{r} \in \mathbb{R}^N \ : \boldsymbol{v}^k = \boldsymbol{O}\boldsymbol{P}^k\boldsymbol{r},\, \boldsymbol{P}^k \in \mathcal{P}^N \forall k \in \{1, \ldots, m\}, \sum_{i=1}^N r_i = 1, \boldsymbol{r} \in \mathbb{R}^N_+\} \neq \varnothing$ where $\mathcal{P}^N := \{P \in \{0,1\}^{N \times N} : \boldsymbol{P}\boldsymbol{e} = \boldsymbol{e},\, \boldsymbol{e}'\boldsymbol{P} = \boldsymbol{e}'\}$ is the set of permutation matrices of size N. We conclude by noting that the feasible region of (19) is obtained from S by applying a standard linearization of the bilinear terms $P^k_{i,j} r_j$ (e.g. Adams and Sherali (1986)) and by adding slack variables for the first set of equalities in the definition of S. Then S is non-empty if and only if the optimal value of (19) is 0. $\square$

Using Proposition 5.4 it is not difficult to find examples where a $\widehat{\rho}_\epsilon$ is a distortion risk measure over $\mathcal{V}$, but cannot be extended to a distortion risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$. For instance, consider the following example.

Example 5.1. *On $\mathbb{R}^2$ let $\Omega = \{(8600, 5000), (5700, 8100), (1300, 9900), (-9600, 3000), (8500, -5200)\}$. We take the distortion risk measure associated to $\boldsymbol{q} := (1/4,\, 1/4,\, 1/4,\, 1/4,\, 0)'$ and consider $\epsilon = 1/5$, for which the associated $\boldsymbol{q}^\epsilon \in \check{\Delta}^5$ is $\boldsymbol{q}^\epsilon = (27/100,\, 27/100,\, 27/100,\, 27/100,\, -2/25)$. Figure 2 shows $\Pi_{\boldsymbol{q}^\epsilon}$ in blue, $\Pi_{\boldsymbol{q}}$ in red, conv(Ω) in green and $\overline{\boldsymbol{\varpi}}$ as a green "+". Clearly $\Pi_{\boldsymbol{q}^\epsilon} \subseteq$ conv(Ω) holds and hence $\widehat{\rho}_\epsilon$ is a distortion risk measure over $\mathcal{V}$. Furthermore, with a symbolic computation software it is easy to find out that $\operatorname{ext}(\Pi_{\boldsymbol{q}^\epsilon}) = \{(905, 3866), (1920, 2781), (3460, 2151), (7275, 4566), (940, 7436)\}$. Then, by Proposition 5.4, to show that $\widehat{\rho}_\epsilon$ cannot be extended to a distortion risk measure over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ it suffices to show that the optimal value of (19) is greater than 0. By using the exact MIP solvers developed in Cook et al. (2011); Espinoza (2006) we were able to computationally prove that the optimal solution of (19) is greater or equal to 1000. We note that the optimum of the LP relaxation of (19) was in fact equal to 0 and hence using an exact LP solver such as the one developed in Applegate et al. (2007) would not have been enough. Furthermore, obtaining the lower bound of 1000 was computationally extremely challenging and required specialized tweaks to the MIP solver from Espinoza (2006) and all the advanced features in the solver from Cook et al. (2011).*

Example 5.1 $\rho(\widetilde{v}) = \boldsymbol{q}'\boldsymbol{v}_{[\cdot]}$ shows that $\{\Pi_{\boldsymbol{q}}\}_{\boldsymbol{q} \in \check{\Delta}^N : \Pi_{\boldsymbol{q}} \subseteq \mathrm{conv}(\Omega)} \subseteq \{\Pi_{\boldsymbol{q}}\}_{\boldsymbol{q} \in \widehat{\Delta}^N}$ does not hold in general and hence we obtain the following refinement of Corollary 4.3 of Bertsimas and Brown (2009).

Corollary 5.5. *Let $\widehat{\mathcal{R}}$ be the set of distortion risk measures over $L_1(\Omega, \mathcal{F}, \mathbb{P})$ and $\check{\mathcal{R}}$ be the set of distortion risk measures over $\mathcal{V}$. Then*

1. $\{\mathcal{U}(\rho)\}_{\rho \in \widehat{\mathcal{R}}} = \{\Pi_{\boldsymbol{q}}\}_{\boldsymbol{q} \in \widehat{\Delta}^N}$.
2. $\{\mathcal{U}(\rho)\}_{\rho \in \check{\mathcal{R}}} \supseteq \{\Pi_{\boldsymbol{q}}\}_{\boldsymbol{q} \in \check{\Delta}^N : \Pi_{\boldsymbol{q}} \subseteq conv(\Omega)}$.

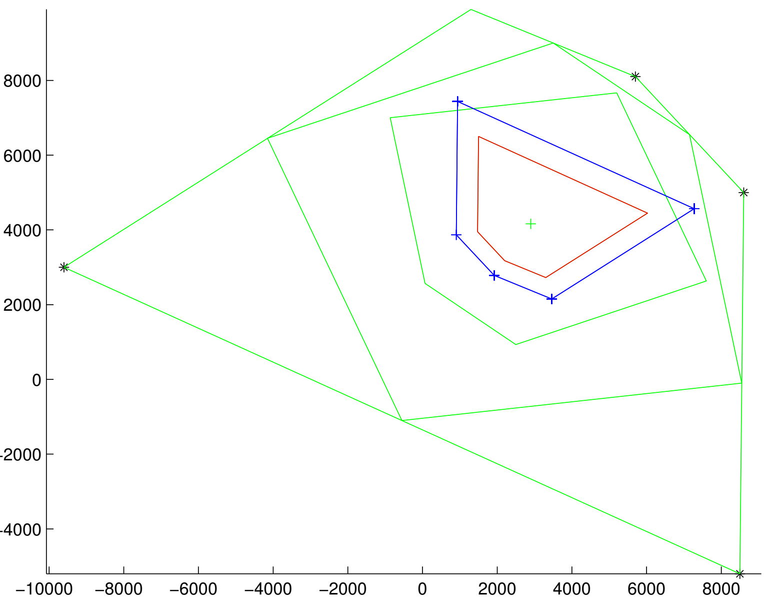

Figure 2: An Expansion of a $\boldsymbol{q}-permutahull$ that is not a $\boldsymbol{q}-permutahull$.

3. $\{\mathcal{U}(\rho)\}_{\rho\in\widehat{\mathcal{R}}} \subseteq \{\mathcal{U}(\rho)\}_{\rho\in\breve{\mathcal{R}}}$ *and the containment can be strict.*

Proof. 1. is Corollary 4.3 of Bertsimas and Brown (2009). 2. comes from Proposition 5.2 and 3. from Example 5.1. □

One problem with measures associated to $\boldsymbol{q} \in \breve{\Delta}^N$ is that it might be hard to check $\Pi_{\boldsymbol{q}} \subseteq \text{conv}(\Omega)$. Indeed, by letting $\boldsymbol{\omega}_0^i := \boldsymbol{\omega}^i - \overline{\boldsymbol{\omega}}$ for all i, $\Omega_0 := \{\boldsymbol{\omega}_0^1, \ldots, \boldsymbol{\omega}_0^N\}$ and A^o be the 1-polar of set A we have that checking $\Pi_{\boldsymbol{q}} \subseteq \text{conv}(\Omega)$ is equivalent to checking that $\text{conv}(\Omega_0)^o := \{\boldsymbol{a} \in \mathbb{R}^d : \boldsymbol{a}'\boldsymbol{\omega}_0^i \leqslant 1\ \forall i\}$ is contained in $\Pi_{\boldsymbol{q}}^o := \{\boldsymbol{a} \in \mathbb{R}^d : \boldsymbol{a}'\boldsymbol{O}\boldsymbol{P}\boldsymbol{q} \leqslant 1\ \forall \text{ permutation matrix } \boldsymbol{P}\}$. Unfortunately, $\Pi_{\boldsymbol{q}}^o$ may have as many as $N!$ constraints so checking this last containment will not be practical using the standard technique of maximizing the left hand sides of the inequalities $\Pi_{\boldsymbol{q}}^o$ over $\text{conv}(\Omega_0)^o$. We could instead use the polynomial sized extended formulation of $\Pi_{\boldsymbol{q}}^o$ introduced in Bertsimas and Brown (2009). However, checking that the projection of a polytope given by a polynomial number of linear inequalities is contained in another polytope with a polynomial number of linear inequalities is in general coNP-complete[6] and hence this approach could fail as well.

Fortunately the following corollary of Theorem 4.6 shows that for random variables over $\mathcal{V}_0$ it is sometimes possible to bypass condition $\Pi_{\boldsymbol{q}} \subseteq \text{conv}(\Omega)$.

Corollary 5.6. *Let* $\mathbf{0} \in \text{ri}\,(conv(\Omega))$ *and let* ρ *be a distortion risk measure over* $\mathcal{V}_0$*. Then* $\widehat{\rho}_\epsilon$ *is a distortion risk measure over* $\mathcal{V}_0$ *for any* $\epsilon \geqslant 0$*.*

Proof. It follows from Proposition 4.6 by noting that in the discrete case $\mathbf{0} \in \text{ri}\,(\text{conv}(\text{supp}(\mathbb{P}))$ is equivalent to $\mathbf{0} \in \text{ri}\,(\text{conv}(\Omega))$. □

[6]This follows from the fact that checking that a polytope given by as the convex hull of a polynomial number of points (V-polytope) is contained in a polytope given by a polynomial number of inequalities (H-polytope) is coNP-complete (Freund and Orlin, 1985) and that any V-polytope with a polynomial number of vertices can be written as the projection of an H-polytope with a polynomial number of inequalities.

Note that in the case considered in Corollary 5.6, although we do not necessarilly have $\mathcal{U}(\widehat{\rho}_\epsilon) \nsubseteq \operatorname{conv}(\Omega)$, we do have $\mathcal{U}(\widehat{\rho}_\epsilon) \subseteq \operatorname{aff}(\Omega)$.

6. Computational Experiments

In the previous sections we have argued why, when considering linear and affine random variables, we should not discard risk measure that are only coherent or distortion risk measures over these types of random variables. However, we have not argued why it would be beneficial to use these restricted risk measures. In this section we present a computational example that illustrates how restricted risk measures such as the epsilon scalings could provide an advantage when risk measures are approximated using samples. For simplicity we will consider samples that are drawn from simple continuous distributions. However, in these cases safe tractable approximations of the associated uncertainty sets can be preferable to the sampled approximations (Nemirovski, 2012). In contrast, the experiments in this section are meant to illustrate the potential advantage of epsilon scalings in the case in which the samples are drawn from distributions that are not known explicitly such as those obtained from conditional simulations (e.g. Lagos et al. (2011); Vielma et al. (2009)). In this cases, it is not clear how to construct safe tractable approximations, so the only current option is to utilize the samples as potential scenarios.

It has been recently argued that the *Conditional Value-at-Risk* (CVaR) measure might be highly susceptible to estimation errors when being approximated by samples (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 $\max\{\boldsymbol{x}'\boldsymbol{r} \ : \ \boldsymbol{x}'\boldsymbol{e} = 1,\, \boldsymbol{x} \geqslant \boldsymbol{0}\}$. 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. Interpreting $-\boldsymbol{x}'\widehat{\boldsymbol{r}}$ as the losses of the portfolio, a classic and well studied approach to this problem is to minimize the Conditional Value at Risk of the losses:

$$\min_{\boldsymbol{x}} \left\{ \operatorname{CVaR}_\delta(-\boldsymbol{x}'\widetilde{\boldsymbol{r}}) : \boldsymbol{x}'\boldsymbol{e} = 1,\, \boldsymbol{x} \geqslant \boldsymbol{0} \right\}, \tag{20}$$

where $\operatorname{CVaR}_\delta(\widetilde{v}) := \min_t \left\{ t + \frac{1}{\delta}\mathbb{E}[(\widetilde{v} - t)^+] \right\}$. Evaluating CVaR requires multidimensional integration and hence is in general intractable. A data-driven approach for this issue is to rely on a finite sample of observations or scenarios of the returns, and 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, §5.1.1). Assume then that we have a finite 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 (20) is given by

$$\min_{\boldsymbol{x}} \left\{ \operatorname{CVaR}_\delta^N(-\boldsymbol{x}'\widetilde{\boldsymbol{r}}) : \boldsymbol{x}'\boldsymbol{e} = 1,\, \boldsymbol{x} \geqslant \boldsymbol{0} \right\}, \tag{21}$$

where $\operatorname{CVaR}_\delta^N(-\boldsymbol{x}'\widetilde{\boldsymbol{r}}) := \min_t \left\{ t + \frac{1}{\delta N}\sum_{i=1}^N [-\boldsymbol{x}'\boldsymbol{r}^i - t]^+ \right\}$. It is well known Rockafellar and Uryasev (2002, 2000) that this problem is equivalent to

$$\min_{\boldsymbol{x},t} \left\{ t + \frac{1}{\delta N}\sum_{i=1}^N [-\boldsymbol{x}'\boldsymbol{r}^i - t]^+ : t \in \mathbb{R},\ \boldsymbol{x}'\boldsymbol{e} = 1,\, \boldsymbol{x} \geqslant \boldsymbol{0} \right\}. \tag{22}$$

It is noted in Lim et al. (2011) that the optimal solutions of problems similar to (22) have a significant difference between their sampled $\operatorname{CVaR}_\delta^N$ and their real $\operatorname{CVaR}_\delta$. We aim to find an optimization problem that has solutions that

(i) Have little difference between their sampled $\operatorname{CVaR}_\delta^N$ and their real $\operatorname{CVaR}_\delta$.

(ii) Have values of their real CVaR_δ that are similar or better than the real CVaR_δ values of the solutions of (22).

To achieve this we propose replacing CVaR_δ by $\widehat{\mathrm{CVaR}}_{\delta\epsilon}$. That is to solve the sampled version of

$$\min\left\{\epsilon\ \mathrm{CVaR}_\delta(-\boldsymbol{x}'\tilde{\boldsymbol{r}}) + (1-\epsilon)\ \mathbb{E}[-\boldsymbol{x}'\tilde{\boldsymbol{r}}] : \boldsymbol{x}'\boldsymbol{e} = 1,\ \boldsymbol{x} \geqslant \boldsymbol{0}\right\}. \tag{23}$$

given by

$$\min\left\{\epsilon\ \left(t + \frac{1}{\delta N}\sum_{i=1}^{N}[-\boldsymbol{x}'\boldsymbol{r}^i - t]^+\right) - (1-\epsilon)\ \boldsymbol{x}'\overline{\boldsymbol{r}} : t \in \mathbb{R},\ \boldsymbol{x}'\boldsymbol{e} = 1,\ \boldsymbol{x} \geqslant \boldsymbol{0}\right\} \tag{24}$$

where $\overline{\boldsymbol{r}} := \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{r}^i$ and $\epsilon > 1$. The motivation behind this choice comes from the characterization of uncertainty sets for elliptical distributions given in Lemma 3.4. Assuming $B = I$ in this lemma, we have that using CVaR_δ^N to approximate CVaR_δ is equivalent to approximating a translation of the euclidean ball with the convex hull of the finite number of points given by $\widehat{\Omega}_{\boldsymbol{h}^\delta} := \left\{\sum_{i=1}^{N}\boldsymbol{h}^\delta_{\sigma(i)}\boldsymbol{r}^i : \sigma \in S_N\right\}$ for $\boldsymbol{h}^\delta$ defined in (8). To get a good approximation we need the number of extreme points of $\operatorname{conv}\left(\widehat{\Omega}_{\boldsymbol{h}^\delta}\right)$ to be quite large Ball (1997). While it is hard to predict which δ will yield the largest number of extreme points, it seems reasonable to pick $\delta = 0.5$ as it at least maximizes $\left|\widehat{\Omega}_{\boldsymbol{h}^\delta}\right|$. However, while this $\widehat{\Omega}_{\boldsymbol{h}^{0.5}}$ should approximate better the *shape* of the euclidean ball, it will actually approximate a ball of smaller radius that the one we are truly interested in (assuming the original δ is smaller than 0.5). Fortunately, we can use the following direct corollary of Lemma 3.4 to calculate the scaling that will give us the correct radius.

Corollary 6.1. *Let $\mu \in \mathbb{R}^d$, $B \in \mathbb{R}^{d\times d}$ be a non-singular matrix and $\mathbb{P}$ be any probability distribution such that $\tilde{u}_{\boldsymbol{x}} := \boldsymbol{x}'B^{-1}(\tilde{\boldsymbol{\omega}} - \mu)$ has the same continuous probability distribution for every $\boldsymbol{x} \in S^{n-1} := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_2 = 1\}$. Then, $\mathcal{U}(\mathrm{CVaR}_\delta) = \mathcal{U}\left(\widehat{\mathrm{CVaR}}_{0.5\epsilon}\right)$ for $\epsilon = \frac{\mathrm{CVaR}_\delta(\mu)}{\mathrm{CVaR}_{0.5}(\mu)}$.*

We begin our experiments with a Gaussian distribution as it satisfies the conditions of Corollary 6.1 and it also allows for the exact solution of (20).

6.1. Results for Gaussian Distribution

To generate the data for our experiments we utilize the same historical data for 200 stocks listed in SP-500 used in Vielma et al. (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 3.4, we can compute the exact optimum solving the following second-order conic problem

$$\min_{\boldsymbol{x},\boldsymbol{t}}\left\{\mathrm{CVaR}_\delta(\mu)\cdot t - \boldsymbol{x}'\overline{\boldsymbol{r}}) : \boldsymbol{x}'\boldsymbol{e} = 1,\ ||\Sigma\boldsymbol{x}||_2 \leqslant \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{t} \geqslant \boldsymbol{0}\right\} \tag{25}$$

where $\mu \sim N(0,1)$.

Our objective is to evaluate what sampled problems have optimal solutions that comply best with points (i)–(ii) above for $\delta \in \{0.1, 0.01\}$. For this we proceed as follows.

1. Generate N i.i.d. samples from our real distribution $\mathcal{N}(\mu, \Sigma)$.
2. Solve the sampled CVaR problem (22) for $\delta = 0.01$ and $\delta = 0.1$ and save the solutions $\boldsymbol{x}^*_{\mathrm{CVaR}_\delta}$ with optimal value $z^*(\boldsymbol{x}^*_{\mathrm{CVaR}_\delta})$.
3. Compute $\epsilon = \frac{\mathrm{CVaR}_\delta(\mu)}{\mathrm{CVaR}_{0.5}(\mu)}$ where $\mu \sim N(0,1)$.
4. Solve the sampled $\widehat{\mathrm{CVaR}}_{\delta\epsilon}$ problem (24) for $\delta = 0.01$ and $\delta = 0.1$ and save the solutions $\boldsymbol{x}^*_{\widehat{\mathrm{CVaR}}_{\delta,\epsilon}}$ with optimal value $z^*(\boldsymbol{x}^*_{\widehat{\mathrm{CVaR}}_{\delta,\epsilon}})$.

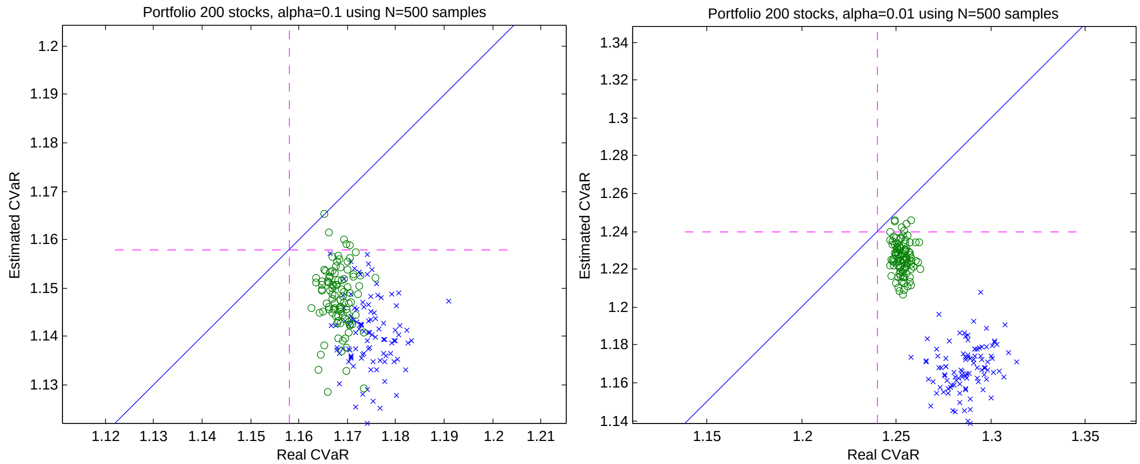

Figure 3: Mitigating Estimation Errors of CVaR.

5. Plot $\mathrm{CVaR}_\delta(-\boldsymbol{x}'\widetilde{\boldsymbol{r}})$ v/s $z^*(\boldsymbol{x})$ for $\boldsymbol{x} \in \left\{\boldsymbol{x}^*_{\mathrm{CVaR}_\delta}, \boldsymbol{x}^*_{\widehat{\mathrm{CVaR}}_{\delta,\epsilon}}\right\}$.
6. Repeat steps 1–4 100 times.

Figure 3 shows the results for this experiment. Blue x's correspond to $\boldsymbol{x}^*_{\mathrm{CVaR}_\delta}$ and green circles correspond to $\boldsymbol{x}^*_{\widehat{\mathrm{CVaR}}_{\delta,\epsilon}}$. A magenta line show the exact CVaR_δ as given by (25). As expected (e.g. (Shapiro et al., 2009, Proposition 5.6)), the sampled CVaR consistently underestimate the real CVaR and the epsilon scaling reduces this downward bias. In addition, the epsilon scaling reduces variability of both the sampled and real CVaR of the optimal solutions and tends to provide better solutions. This is confirmed in Table 1, where we present the 95% confidence intervals (cf. Shapiro et al. (2009)) for portfolio sized of 20 and 200 stocks, $\delta \in \{0.1, 0.01\}$ and sample sizes of $N = 100$, 200 and 500.

Port. Size	δ	CVaR_δ	Type	Confidence Interval (95%) $N = 100$	$N = 200$	$N = 500$
20	0.01	1.3799	CVaR_δ^N	[1.2414,1.4023]	[1.2801,1.3995]	[1.3324,1.3934]
			$\widehat{\mathrm{CVaR}}_\delta^N$	[1.3472,1.3926]	[1.3622,1.3906]	[1.3745,1.3857]
20	0.1	1.2501	CVaR_δ^N	[1.2079,1.2694]	[1.2267,1.2646]	[1.2431,1.2644]
			$\widehat{\mathrm{CVaR}}_\delta^N$	[1.2236,1.2657]	[1.2349,1.2661]	[1.2460,1.2630]
200	0.01	1.2398	CVaR_δ^N	[1.0834,1.2901]	[1.1252,1.2756]	[1.1677,1.2607]
			$\widehat{\mathrm{CVaR}}_\delta^N$	[1.1707,1.2710]	[1.2034,1.2603]	[1.2259,1.2493]
200	0.1	1.1579	CVaR_δ^N	[1.0846,1.1960]	[1.1177,1.1824]	[1.1408,1.1719]
			$\widehat{\mathrm{CVaR}}_\delta^N$	[1.1079,1.1880]	[1.1328,1.1710]	[1.1480,1.1678]

Table 1: 95% Confidence Interval for portfolio of 20 and 200 stocks for different values of δ and N

6.2. Results for Independent Uniform Distribution

To study a case in which the conditions of Corollary 6.1 do not hold we repeat the previous experiment assuming that each stock has a return $r_i = \mu_i + \varepsilon_i$, where ε_i as an independent random variable uniformly distributed in $[-1, 1]$. In this case $\mathcal{U}(\mathrm{CVaR}_\delta)$ for different δs are not scalings of one another (e.g. see

Figure 1) and even evaluating CVaR_δ requires multidimensional integration. For this reason we calculate ϵ in line 3 as

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

for $M = 100,000$.

The results for this experiment are shown in Figure 4. We can see that we obtain a behaviour similar

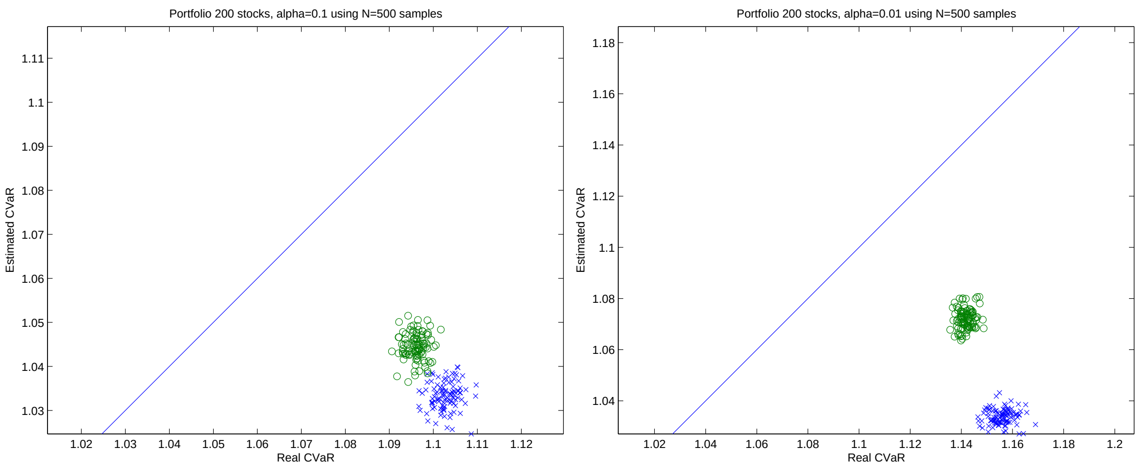

Figure 4: Mitigating Estimation Errors of CVaR.

to the Gaussian case even though in this case the shapes of $\mathcal{U}(\text{CVaR}_\delta)$ and $\mathcal{U}\left(\widehat{\text{CVaR}_{0.5\epsilon}}\right)$ could be quite different.

While these experiments are far from conclusive, they do illustrate how using epsilon scalings can produce results that are significantly different than those obtained from the original risk measure and in particular can have more stable estimators.

7. Acknowledgements

Daniel Espinoza was supported by FONDECYT Grant 1110024 and ICM Grant P10-024-F and Eduardo Moreno was supported by Anillo Grant ACT88 and CMM Basal Project.

References

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

Applegate, D., Cook, W.J., Dash, S., Espinoza, D.G., 2007. Exact solutions to linear programming problems. Oper. Res. Lett. 35, 693–699.

Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk. Mathematical finance 9, 203–228.

Ball, K.M., 1997. An elementary introduction to modern convex geometry, in: Levy, S. (Ed.), Flavors of Geometry. Cambridge University Press, Cambridge. volume 31 of *Mathematical Sciences Research Institute Publications*, pp. 1–58.

Ben-Tal, A., Ghaoui, L., Nemirovskiĭ, . Robust optimization.

Bertsimas, D., Brown, D., 2009. Constructing uncertainty sets for robust linear optimization. Operations research 57, 1483–1495.

Cook, W., Koch, T., Steffy, D.E., Wolter, K., 2011. An exact rational mixed-integer programming solver, in: Günlük, O., Woeginger, G.J. (Eds.), IPCO, Springer. pp. 104–116.

Espinoza, D.G., 2006. On Linear Programming, Integer Programming and Cutting Planes. Ph.D. thesis. Georgia Institute of Technology.

Freund, R.M., Orlin, J.B., 1985. On the complexity of four polyhedral set containment problems. Mathematical Programming 33, 139–145.

Hiriart-Urruty, J.B., Lemaréchal, C., 2001. Fundamentals of Convex Analysis. Springer Verlag, Heidelberg.

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, 15 – 20. LAGOS'11 - VI Latin-American Algorithms, Graphs and Optimization Symposium.

Lim, A.E., Shanthikumar, J.G., Vahn, G.Y., 2011. Conditional value-at-risk in portfolio optimization: Coherent but fragile. Operations Research Letters 39, 163 – 171.

Natarajan, K., Pachamanova, D., Sim, M., 2009. Constructing risk measures from uncertainty sets. Operations research 57, 1129–1141.

Nemirovski, A., 2012. On safe tractable approximations of chance constraints. European Journal of Operational Research 219, 707–718.

Rockafellar, R., Uryasev, S., 2000. Optimization of conditional value-at-risk. Journal of risk 2, 21–42.

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

Shapiro, A., 2011. On kusuoka representation of law invariant risk measures. Optimization Online `http://www.optimization-online.org/DB_HTML/2011/09/3164.html`.

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, 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.