Computational aspects of robust optimized certainty equivalents and option pricing

Daniel Bartl 1 | Samuel Drapeau 2 | Ludovic Tangpi 3

1 Department of Mathematics, University of Konstanz, Konstanz, Germany
2 School of Mathematical Sciences and Shanghai Advanced Institute of Finance (SAIF/CAFR), Shanghai Jiao Tong University, Shanghai, China
3 Department of Operations Research and Financial Engineering, Princeton University, Charleton Street Princeton, New Jersey

Abstract
Accounting for model uncertainty in risk management and option pricing leads to infinite-dimensional optimization problems that are both analytically and numerically intractable. In this article, we study when this hurdle can be overcome for the so-called optimized certainty equivalent (OCE) risk measure—including the average value-at-risk as a special case. First, we focus on the case where the uncertainty is modeled by a nonlinear expectation that penalizes distributions that are “far” in terms of optimal-transport distance (e.g., Wasserstein distance) from a given baseline distribution. It turns out that the computation of the robust OCE reduces to a finite-dimensional problem, which in some cases can even be solved explicitly. This principle also applies to the shortfall risk measure as well as for the pricing of European options. Further, we derive convex dual representations of the robust OCE for measurable claims without any assumptions on the set of distributions. Finally, we give conditions on the latter set under which the robust average value-at-risk is a tail risk measure.

Keywords
average value-at-risk, convex duality, distribution uncertainty, optimized certainty equivalent, optimal transport, penalization, robust option pricing, Wasserstein distance

1 INTRODUCTION

In this article, we study properties of the optimized certainty equivalent (OCE) of Ben-Tal and Teboulle (1986, 2007), and to a wider extent option pricing, under model uncertainty. In the context of risk assessment, the rationale behind the definition of the OCE is as follows. Assume that a financial agent...
faces a future uncertain loss profile with distribution $\mu$, and she wants to assess its risk. In her assessment, given a loss function $l : \mathbb{R} \to (-\infty, \infty]$, she computes the expectation $\int l(x)\mu(dx)$ representing the present average cost of her losses. She can, however, reduce her overall future losses by allocating some cash $m$, resulting in the present value $\int (l(x) - m)\mu(dx) + m$. Minimizing over all possible allocations defines the optimal cost or OCE of $\mu$ with respect to the loss function $l$:

$$\text{OCE}(l) := \inf_{m \in \mathbb{R}} \left( \int l(x - m)\mu(dx) + m \right).$$  \hspace{1cm} (1)

Now, if the distribution $\mu$ of future loss is not perfectly known, the risk-averse financial agent will consider the overall cost of an allocation $m$ to be given by $\mathcal{R}(l(\cdot - m)) + m$, where $\mathcal{R}$ is a nonlinear expectation. To wit, $\mathcal{R}$ models the degree of conservatism or risk aversion of the investor and can be interpolated between the linear case $\mathcal{R}(\cdot) = \int \cdot d\mu$ and the worst case $\mathcal{R}(\cdot) = \sup_{\mu \in \mathcal{M}_1(\mathbb{R})} \int \cdot d\mu$, with $\mathcal{M}_1(\mathbb{R})$ the set of probability measures on $\mathbb{R}$. Hence, the natural definition of the robust OCE is

$$\mathcal{OCE}(l) := \inf_{m \in \mathbb{R}} (\mathcal{R}(l(\cdot - m)) + m),$$  \hspace{1cm} (2)

that is, the minimal allocation cost when the future expected loss is written in terms of the nonlinear expectation $\mathcal{R}$.

The classical OCE satisfies sound economical properties discussed in Ben-Tal and Teboulle (2007). In particular, it is a convex monetary risk measure in the sense of Artzner, Delbaen, Eber, and Heath (1999) and Föllmer and Schied (2002), which is additionally law invariant, see Frittelli and Gianin (2002) for definition and consequences. Furthermore, depending on the specification of the loss function $l$, it includes classical risk measures such as the entropic risk measure, the average value-at-risk, see Rockafellar and Uryasev (2000), the monotone mean variance of Maccheroni, Marinacci, and Rustichini (2006), and as a scaling limit, the shortfall risk measure of Föllmer and Schied (2002). Stated as a classical unconstrained one-dimensional optimization problem, the OCE is a smooth quantification instrument, see Cheridito and Li (2009) and Cherny and Kupper (2007). The computation of the risk as well as the risk contributions can be explicitly stated in terms of first-order conditions and efficiently implemented using Fourier transform methods (see Drapeau, Kupper, & Papapantoleon, 2014) and stochastic root finding methods (see Hamm, Salfeld, & Weber, 2014, as well as Dunkel & Weber, 2010). However, when facing model uncertainty, these properties become a priori challenging due to the potential infinite-dimensional nature of the optimization problem (2). In addition, it is not clear by how much the resulting robust quantification of risk deviates from its nonrobust counterpart, a crucial question in practice.

The goal of this paper is to study the robust OCE and provide several ways to reduce the complexity stemming from the robustness to get explicit formulas allowing for a quantification of the risk under model misspecification. Our first main result focuses on the case where

$$\mathcal{R}(f) := \sup_{\mu \in \mathcal{M}_1(\mathbb{R})} \left( \int f d\mu - \varphi(d_c(\mu_0, \mu)) \right),$$

with $\varphi$ being a positive function, $d_c$ an optimal transport-like distance with cost function $c(x, y)$ such as the Wasserstein distance, and $\mu_0$ a fixed, given distribution. The underlying intuition is that, using past information for instance, one knows a priori that a distribution $\mu_0$ is likely to be the true distribution of
BARTL ET AL.

the financial loss whose risk is being assessed. Due to uncertainty about this estimation, one considers
every possible distribution, penalizing, however, those that are “far away” from the baseline distribution
\( \mu_0 \) in terms of the distance \( d_c \). See further concrete financial motivations for such an approach in
Remark 2.9. After proving the general duality formula,

\[
\sup_{\mu \in \mathcal{M}_1(\mathbb{R})} \left( \int f \, d\mu - \varphi(d_c(\mu_0, \mu)) \right) = \inf_{\lambda \geq 0} \left( \int f^\lambda \, d\mu_0 + \varphi^*(\lambda) \right)
\]

with \( \varphi^* \) being the convex conjugate of \( \varphi \) in Theorem 2.4, we show in Theorem 2.7 that

\[
\mathcal{OCE}(l) = \inf_{\lambda \geq 0} (\text{OCE}(l^\lambda) + \varphi^*(\lambda)), \quad \text{where} \quad l^\lambda(x) \coloneqq \sup_{y \in \mathbb{R}} (l(y) - \lambda c(x, y)).
\]

This formula for the robust OCE shows that the \textit{infinite-dimensional} optimization problem of comput-
ing \( \mathcal{OCE}(l) \) simplifies into a finite-dimensional problem of computing the OCE of the distribution \( \mu_0 \)
and with modified loss function \( l^\lambda \).

We stress that the duality formula (3) is interesting on its own, and is valid for measurable functions \( f \),
lower semicontinuous cost, and penalty functions. Adjusting the penalty function \( \varphi \) enables to set the
level of risk aversion. A particular case considered in the literature arises when choosing \( \varphi \coloneqq \infty 1_{(\delta, \infty]} \)
for some \( \delta > 0 \), as the nonlinear expectation \( \mathcal{R} \) becomes

\[
\mathcal{R}(f) = \sup_{\{\mu \in \mathcal{M}_1(\mathbb{R}) : d_c(\mu_0, \mu) \leq \delta\}} \int f \, d\mu,
\]

the worst-case expectation over a ball around the baseline distribution \( \mu_0 \). In this case, taking \( d_c \) to be
the first-order Wasserstein distance and \( l(x) = x^+ / \alpha \) so that the OCE becomes the average value-at-risk
at level \( \alpha \), one has

\[
\text{AVR}_{\alpha} = \text{AVR}_{\alpha}(\mu_0) + \delta / \alpha,
\]

see Example 2.10 for different penalizations \( \varphi \) and different distances. In other terms, the robust aver-
age value-at-risk is the same as the standard average value-at-risk plus an “uncertainty premium” \( \delta / \alpha \).
The above formula was obtained by Wozabal (2014) in the case where the existence of a dominating
probability measure is assumed. In the particular case where \( \varphi = \infty 1_{(\delta, \infty]} \), convex dual representations
of \( \mathcal{R} \) and applications thereof to stochastic programming have been recently studied. The first con-
tribution is due to Esfahani and Kuhn (2018) who derive the duality formula when \( d_c \) is the first-order
Wasserstein distance, \( \mu_0 \) the empirical measure, and the function \( f \) is convex with a special structure.
We refer to Zhao and Guan (2018) for a similar setting and another class of objective functions \( f \).
Blanchet and Murthy (2016) and Gao and Kleywegt (2016) obtain the result on a general Polish space
and for lower semicontinuous cost functions \( c \) and upper semicontinuous integrable functions \( f \). The
proof given in the present paper is essentially more direct. It relies on a version of Choquet’s capaci-
tability theorem and applies for measurable functions.

The application of this duality goes well beyond dimension reduction for robust OCEs. It also
enables us to derive a finite-dimensional representation of the robust shortfall risk measure of
Föllmer and Schied (2002). Interestingly, a “martingale version” of (3) provides a finite-dimensional
analytical formula for robust European option pricing, see Proposition 2.13. As an example, if
we consider the price CALL\((k)\) of a call option on a stock with time one price \( S \) and time
zero price \( s \) and risk-neutral distribution \( \mu_0 \), for the robust price of a linearly penalized call,
we obtain

$$\text{CALL}(k) := \sup_{\mu \in \mathcal{M}_1(\mathbb{R}) : \int_{\mathbb{R}_d} S \, d\mu = s} \left( \int (x - k)^+ \, d\mu - d_c(\mu_0, \mu) \right)$$

$$= \inf_{\alpha \in \mathbb{R}} \left( \text{CALL} \left( k - \frac{2\alpha + 1}{2} \right) + \frac{\alpha^2}{2} \right).$$

when $d_c$ is the second-order Wasserstein distance. In other terms, robust pricing of European options with respect to penalization of a Wasserstein distance simply boils down to optimizing the payoff—here the strike—of the option price with respect to the baseline risk-neutral distribution.

Coming back to the OCE, we further investigate alternative representations of the robust OCE when the nonlinear expectation $\mathcal{R}$ is the worst-case expectation over an arbitrary set $D \subseteq \mathcal{M}_1(\mathbb{R})$. In particular, we derive the representation of the robust $\text{AV} \@ \mathcal{R}$ as a tail risk measure. This representation, first proved by Rockafellar and Uryasev (2000) in the nonrobust case, has important applications in optimization problems and does not carry over to the robust case unless stronger structural and topological assumptions are put on the set $D$, see Proposition 3.1. Under such assumptions, we derive a Kusuoka-type representation (see Kusuoka, 2001; Jouini, Schachermayer & Touzi, 2006) for “robustifications” of law-invariant risk measures, see Corollary 3.3. Finally, when defined on random variables, convex dual representations of OCE are particularly relevant, see, for instance, Cherny and Kupper (2007) for applications to optimization problems and Backhoff and Tangpi (2016) for applications to dynamic representations. In the present article, we derive a dual representation of the robust OCE on the set of bounded measurable random variables, without any topological assumption on the sample space.

Financial modeling under model ambiguity, known as robust finance, is currently the subject of intensive research. Earliest papers dealing with robust risk measures with the ambiguity set given by a Wasserstein ball include Pflug, Pichler, and Wozabal (2012) and Wozabal (2014) in the context of portfolio optimization problems. Among other more recent papers, we refer to Hobson (1998), Cheridito, Kupper, and Tangpi (2017), Beiglböck, Henry-Labordère, and Penkner (2011), and Dolinsky and Soner (2014) for superhedging problems and Bartl (2019), Blanchard and Carassus (2018), Matoussi, Possamaï, and Zhou (2015), Nutz (2016), and Neufeld and Sikic (2018) for robust utility maximization. Distributionally robust problems are also studied in statistics, economics, and operations research (see, e.g., Hansen & Sargent 2001; Huber 1981).

The paper is organized as follows: The next section summarizes our main findings. Namely, a duality result for the nonlinear operator $\mathcal{R}$ for general penalty functions and measurable cost functions. This result is then used to provide finite-dimensional representations of the OCE, shortfall risk measure, and price of European options in the presence of model ambiguity. Furthermore, we give conditions under which the robust $\text{AV} \@ \mathcal{R}$ is a tail-risk measure, and a convex dual representation when the OCE is defined on random variables. We also give several edifying examples. In the final section, we provide detailed proofs.

2 | MAIN RESULTS FOR UNCERTAINTY GIVEN BY WASSERSTEIN DISTANCES

We start this section with briefly defining our notation, a short review on the distances emerging from optimal transport, and the duality theorem hinted in (3). Throughout, let $d \in \mathbb{N}$ and $X \subseteq \mathbb{R}^d$ be a closed set. We denote by $\mathcal{M}_1(X)$ the set of all probabilities on the Borel $\sigma$-field of $X$. For
a measurable function $f : X \to (-\infty, \infty]$ bounded from below and $\mu \in \mathcal{M}_1(X)$, we denote by $\int f \, d\mu = \int_X f(x) \mu(dx) \in (-\infty, \infty]$ the integral of $f$ against $\mu$. Now fix some measurable function $c : X \times X \to [0, \infty]$ and define the cost of transportation by

$$d_c(\mu, \nu) := \inf \left\{ \int_{X \times X} c(x, y) \pi(dx, dy) : \pi \in \mathcal{M}_1(X \times X) \text{ such that } \pi(\cdot \times X) = \mu \text{ and } \pi(X \times \cdot) = \nu \right\}$$

for $\mu, \nu \in \mathcal{M}_1(X)$. If $c$ is assumed to be lower semicontinuous, this problem has a dual formulation

$$d_c(\mu, \nu) = \sup \left\{ \int_X f \, d\mu + \int_X g \, d\nu : f, g : X \to \mathbb{R} \text{ bounded continuous such that } f(x) + g(y) \leq c(x, y) \text{ for all } x, y \in X \right\},$$

see, for instance, Villani (2008, Theorem 5.9) for a proof.

Remark 2.1. For the specific choice $c(x, y) = |y - x|^{p}$ with $p \geq 1$, the function $d_c^{1/p}$ defines the celebrated Wasserstein distance of order $p$. In case of compact $X$, this is, in fact, a metric on $\mathcal{M}_1(X)$ compatible with the weak topology, that is the coarsest topology making all mappings $\mu \mapsto \int f \, d\mu$ continuous for continuous and bounded $f$. However, for general $X \subset \mathbb{R}^d$, one has $d_c(\mu_n, \mu) \to 0$ if and only if $\mu_n$ converges weakly to $\mu$ and $\int_X |x|^p \mu_n(dx) \to \int_X |x|^p \mu(dx)$, see Villani (2008, Theorem 6.8).

For a function $f : X \to (-\infty, \infty]$ and $\lambda \geq 0$, its $\lambda c$-transform is defined and denoted by

$$f^{\lambda c}(x) := \sup \{ f(y) - \lambda c(x, y) : y \in X \text{ such that } f(y) < \infty \}.$$

Remark 2.2. If $f$ and $c$ are continuous, then $f^{\lambda c}$ is lower semicontinuous. In general, if $f$ is only assumed to be measurable, then $f^{\lambda c}$ is not necessarily. However, if $f$ is measurable, it follows, for instance, from Bertsekas and Shreve (1978, Section 7) that $f^{\lambda c}$ is universally measurable and in particular $\mu$-measurable for every $\mu \in \mathcal{M}_1(X)$. This implies that the integral $\int f^{\lambda c} \, d\mu$ is well defined whenever $f$ is measurable and bounded from below. Note also that $f^{\lambda c}$ is a well-known modification of the classical Fenchel–Legendre transform studied in the context of optimal transport under the name “$c$-transform,” see, for instance, Villani (2008, Section 5).

From now on, we call

- a **cost function** any lower semicontinuous function $c : X \times X \to [0, \infty]$ with $\inf_{x \in X} c(x, y) = 0$ for all $x \in X$, and for every $r \geq 0$, there is $k \geq 0$ such that $c(x, y) \geq r$ if $|x - y| \geq k$;

- a **penalization function** any convex, increasing lower semicontinuous function $\varphi : [0, \infty] \to [0, \infty]$ with $\varphi(0) = 0$ and neither $\varphi$ nor $\varphi^*$ being constant $0$.

Remark 2.3. For the typical case when the cost function depends only on the difference—with slight abuse of notations $c(y - x) = c(x, y)$—the assumptions of a cost function are satisfied whenever $\inf_{x \in X} c(x) = 0$ and $\liminf_{|x| \to \infty} c(x) = \infty$. In particular, $c(x, y) = |y - x|^p$ with $p > 0$ corresponding to the Wasserstein distances are all cost functions.

**Theorem 2.4.** Given a closed set $X \subset \mathbb{R}^d$, a cost function $c$ and a penalization function $\varphi$, then it holds that

$$\sup_{\mu \in \mathcal{M}_1(X)} \left( \int_X f \, d\mu - \varphi(d_c(\mu_0, \mu)) \right) = \inf_{\lambda \geq 0} \left( \int_X f^{\lambda c} \, d\mu_0 + \varphi^*(\lambda) \right)$$

(6)
for every measurable function \( f : X \to (-\infty, \infty] \) bounded from below. Moreover, the infimum over \( \lambda \) is attained.

**Example 2.5.** Typical examples of penalization function \( \varphi \) we have in mind:

- In case of \( \varphi = \infty 1_{(\delta, \infty]} \) for some fixed \( \delta > 0 \), one computes \( \varphi^*(\lambda) = \delta \lambda \) so that

  \[
  \sup_{\mu \in \mathcal{M}_1(X) : d_c(\mu_0, \mu) \leq \delta} \int f \, d\mu = \inf_{\lambda \geq 0} \left( \int f^{\lambda c} \, d\mu_0 + \delta \lambda \right). \tag{7}
  \]

- For \( \varphi(x) = x \), one gets \( \varphi^* = \infty 1_{(1, \infty)} \) and as \( f^{\lambda c} \leq f^c \) for all \( \lambda \leq 1 \), the formula simplifies to

  \[
  \sup_{\mu \in \mathcal{M}_1(X)} \left( \int f \, d\mu - d_c(\mu_0, \mu) \right) = \int f^c \, d\mu_0.
  \]

- Other examples might include \( \varphi(x) = x^p/p \) for some \( p > 1 \) for which \( \varphi^*(\lambda) = \lambda^q/q \) where \( 1/p + 1/q = 1 \), or \( \varphi(x) = \exp(x) - 1 \) for which \( \varphi^*(\lambda) = \lambda \log \lambda - \lambda + 1 \).

**Remark 2.6.** Note that the formula (7) has recently been proven by several authors, under different assumptions. In Esfahani and Kuhn (2018) and Zhao and Guan (2018), the authors focus on the case \( c(x, y) = |x - y| \) and \( \mu_0 \) an empirical measure. The closest set of assumptions to ours is in Blanchet and Murthy (2016). Therein, the authors work on a general Polish space \( X \), assume \( c \) to be lower semicontinuous and real-valued, and prove duality for \((\mu_0, \text{integrable})\) upper semicontinuous functions \( f \). See also Gao and Kleywegt (2016). Further, note that the techniques differ among all proofs. Our proof builds on convex and Choquet’s regularity result for functional defined on measurable functions and is significantly shorter.

### 2.1 Optimized certainty equivalents and shortfall risk measures

Throughout this section, let \( X = \mathbb{R} \) and \( \mu_0 \in \mathcal{M}_1(\mathbb{R}) \) be some fixed baseline distribution. For any “loss” function \( l : \mathbb{R} \to (-\infty, \infty] \) measurable and bounded from below, recall that the OCE and the shortfall risk measure with respect to \( \mu_0 \) are defined by

\[
\text{OCE}(l) := \inf_{m \in \mathbb{R}} \left( \int l(\cdot - m) \, d\mu_0 + m \right) \quad \text{and} \quad \text{ES}(l) := \inf \left\{ m \in \mathbb{R} : \int l(\cdot - m) \, d\mu_0 \leq 0 \right\}.
\]

For the remainder, we fix a cost function \( c \), which for notational simplicity depends only on the difference, and a penalization function \( \varphi \). Given \( f : \mathbb{R} \to (\infty, \infty] \) measurable and bounded from below, define

\[
\mathcal{R}(f) := \sup_{\mu \in \mathcal{M}_1(\mathbb{R})} \left( \int f \, d\mu - \varphi(d_c(\mu_0, \mu)) \right),
\]

and the robust OCE/shortfall risk measure

\[
\mathcal{OCE}(l) := \inf_{m \in \mathbb{R}} \left( \mathcal{R}(l(\cdot - m)) + m \right) \quad \text{and} \quad \mathcal{ES}(l) := \inf \left\{ m \in \mathbb{R} : \mathcal{R}(l(\cdot - m)) \leq 0 \right\}
\]

for \( l : \mathbb{R} \to (-\infty, \infty] \) measurable and bounded from below.

The following is the first main result of this section that states that the infinite-dimensional problem of computing the quantities \( \mathcal{OCE}(l) \) and \( \mathcal{ES}(l) \) is, in fact, a finite-dimensional problem with different loss function \( l \).
Theorem 2.7. Given a loss function \( l : \mathbb{R} \to \mathbb{R} \) measurable and bounded from below, it holds that

\[
\text{OCE}(l) = \inf_{\lambda \geq 0} \left( \text{OCE}(l^\lambda c) + \varphi^*(\lambda) \right).
\]

If further \( \inf_x l(x) < 0 \), then it holds that

\[
\text{ES}(l) = \inf_{\lambda \geq 0} \text{ES}(l^\lambda c + \varphi^*(\lambda)).
\]

Remark 2.8. In the special case of \( \varphi(x) = x \), computations simplify and

\[
\text{OCE}(l) = \text{OCE}(l^c) \quad \text{and} \quad \text{ES}(l) = \text{ES}(l^c).
\]

Similar, if \( \varphi(x) = \infty 1_{(\delta, \infty]} \) for some \( \delta > 0 \), one gets

\[
\text{OCE}(l) = \inf_{\lambda \geq 0} \left( \text{OCE}(l^\lambda c) + \lambda \delta \right) \quad \text{and} \quad \text{ES}(l) = \inf_{\lambda \geq 0} \text{ES}(l^\lambda c + \lambda \delta).
\]

Remark 2.9. The main reason why the Wasserstein distance is a popular choice of distance to model ambiguity is that the empirical measure converges with respect to this distance, and it is not too strong as opposed to, for instance, the total variation distance, see, for instance, Dereich, Scheutzow, and Schottstedt (2008). For example, the distance between the true measure and the empirical measures converges in expectation, with nonasymptotic rates, roughly in the order \( n^{-1/2} \) if enough moments exist. This justifies the use of optimal-transport distances to model the ambiguity set in finance or any other field where the true distribution is approximated by the empirical measure built on available data. The convergence implies that the approximation becomes more accurate as the data sample size increases. Moreover, concentration inequalities suggest how to choose \( \delta \) in terms of \( N \). More precisely, by Fournier and Guillin (2015), if \( c(x, y) = |x - y|^p \) and the true distribution \( \mu \) is assumed to satisfy \( \int \exp(|x|^2p) \mu(dx) < \infty \), then it holds that

\[
P(\text{d}_e(\mu, \mu_N) \geq \delta) \leq C \exp(-cN\delta^2) \quad \text{for all} \quad \delta \in (0, \infty),
\]

where \( \mu_N := \frac{1}{N} \sum_{n=1}^N \delta_{x_n} \) is the empirical measure on \( N \) i.i.d. observations \( x_1, \ldots, x_N \) and \( C, c > 0 \) are constants (the strong exponential integrability assumption can be weakened to existence of moments, but the formula gets uglier). Also, refer to Bolley, Guillin, and Villani (2007) for similar bounds.

There are many other distances on the space of probability measures one could take into account when defining \( R \), and consequently \( \text{OCE} \) or \( \text{ES} \) (see, e.g., Gibbs & Francis, 2002, for a survey). We already noted before that the Wasserstein distance is stronger than weak convergence, which actually turns out to be necessary in the present setting: If one replaces \( d_e \) by a distance compatible with weak convergence or if \( \liminf_{x \to \infty} c(x) < \infty \) and \( \mu_0 = \delta_0 \), then always \( \text{OCE}(l) = \text{ES}(l) = \infty \) as soon as \( l \) is a convex and not constant loss function. A proof of these facts is given in Section 4.

While Theorem 2.7 reduces the infinite-dimensional problem to a finite-dimensional one, regardless of the computation of the classical OCE, the challenge of computing \( l^\lambda c \), usually for several different \( \lambda \)'s, remains. However, for many relevant examples, closed-form formulas exist.

Example 2.10 (Average value-at-risk). Let \( l(x) = x^\alpha / \alpha \) for some \( \alpha \in (0, 1) \) so that \( \text{OCE}(l) \) becomes the robust average value-at-risk \( \text{AV}@R \) at level \( \alpha \). Table 1 summarizes the relation between the nonrobust average value-at-risk \( \text{AV}@R \) and its robust counterpart \( \text{AV}@R \) for different choices of \( c \) and \( \varphi \).
This example gives a mathematical justification to an intuitively natural fact known as postvaluation adjustment. When computing the risk of a loss $\mu_0$, it is advisable to add a margin to hedge a possible model misspecification or a computational error, see, for instance, Damodaran (2008, Chapter 5).

**Example 2.11 (Monotone mean variance).** Let $l(x) = (((1 + x)^2 - 1)/2$ so that $\text{OCE}(l)$ becomes the robust monotone mean-variance risk measure, see Maccheroni et al. (2006). For the cost function $c(x, y) = (x - y)^2$, one has

$$\text{OCE}(l) = \inf_{\lambda > 1/2} \left( \text{OCE}(\frac{2\lambda}{2\lambda - 1}) + \frac{1}{4\lambda - 2} - \varphi^*(\lambda) \right).$$

For example, if $\varphi(x) = x$, this formula simplifies to $\text{OCE}(l) = \text{OCE}(2l) + 1/2$.

**Example 2.12 (Value-at-risk).** Let $c(x, y) = |x - y|^p$ for some $p > 0$. Then, for the robust value-at-risk at level $\alpha \in (0, 1)$, one has

$$\text{VAR} \varphi_R^\alpha := \inf \left\{ m \in \mathbb{R} : R(1_{(m, \infty)}) \leq \alpha \right\} = \inf \inf_{\lambda \geq 0} \left\{ m \in \mathbb{R} : \mu_0((m, \infty)) + e(m, \lambda) + \varphi^*(\lambda) \leq \alpha \right\},$$

where $e(m, \lambda) := \int_{|x - m|^1/\lambda} 1 - \lambda|x - m|^p \mu_0(dx)$. For example, if $\varphi(x) = x$ and $c(x, y) = |x - y|$, this formula simplifies and $\text{VAR} \varphi_R^\alpha = \inf \{ m \in \mathbb{R} : \mu_0((m, \infty)) + \int_{|m - 1, m|} (m - x) \mu_0(dx) \leq \alpha \}$.

### 2.2 Robust pricing of European options

The principle behind Theorem 2.7 is not limited to the OCE or ES, or other risk measures of similar form. It can be applied, for example, to option pricing, see also the recent paper Blanchet, Chen, and Zhou (2018) for applications to mean variance hedging.

Throughout, let $X = \mathbb{R}^d$ be the canonical space of a $d$-dimensional finance asset $S = (S^1, \ldots, S^d)$, that is, $S : \mathbb{R}^d \to \mathbb{R}^d$ is the identity. Here again, we fix a cost function $c$, which for notational simplicity depends only on the difference, and a penalization function $\varphi$. We further assume that $\liminf_{|x| \to \infty} c(x)/|x|^{1+\epsilon} = \infty$ for some $\epsilon > 0$. Let $\mu_0 \in \mathcal{M}_1(\mathbb{R}^d)$ be an integrable risk-neutral pricing measure for $S$, that is, $s = \int S d\mu_0 \in \mathbb{R}^d$ is the price of these assets at time 0 and assume that $\int c(x - y) \mu_0(dx) < \infty$ for every $y \in \mathbb{R}^d$. We further assume that the interest rate is 0.

Given a European type of option payoff $H := h(S)$ where $h : \mathbb{R}^d \to \mathbb{R}$ is measurable and bounded from below, we denote by

$$\text{PRICE}(H) := \int_{\mathbb{R}^d} h(x) \mu_0(dx) = \int h(S) d\mu_0.$$
its risk-neutral price. Taking uncertainty in the pricing measure into account, an analog of $\mathcal{R}$ in the previous section, consists of considering all probabilities consistent with the asset prices, that is,

$$\text{PRICE}(H) := \sup_{\mu \in \mathcal{M}_1(\mathbb{R}^d) : \int_{\mathbb{R}^d} S \, d\mu = s} \left( \int h \, d\mu - \varphi(d_c(\mu_0, \mu)) \right).$$

As the additional constraint $\int S \, d\mu = s$ is satisfied if and only if $\inf_{\alpha \in \mathbb{R}^d} \int \alpha \cdot (S - s) \, d\mu > -\infty$, formally applying a minimax theorem, one obtains

**Proposition 2.13.** For every measurable payoff $h : \mathbb{R}^d \to \mathbb{R}$ such that $\sup_{x \in \mathbb{R}^d} |h(x)|/(1 + |x|) < \infty$, one has

$$\text{PRICE}(H) = \inf_{\alpha \in \mathbb{R}^d} \inf_{\lambda \geq 0} \left( \int h^{\lambda, \alpha} \, d\mu_0 + \varphi^*(\lambda) \right),$$

where

$$h^{\lambda, \alpha}(x) := \sup_{y \in \mathbb{R}^d} (h(y) + \alpha \cdot (y - s) - \lambda c(y - x)) \quad \text{for } x \in \mathbb{R}^d, \alpha \in \mathbb{R}^d, \lambda \geq 0.$$

The latter function $h^{\lambda, \alpha}$ can be interpreted as a modified payoff priced against the original risk-neutral pricing distribution $\mu_0$. In particular, if $\varphi(x) = x$ the formula again simplifies to

$$\text{PRICE}(H) = \inf_{\alpha \in \mathbb{R}^d} \int h^{\alpha} \, d\mu_0.$$

**Example 2.14 (Robust Call).** Let us consider the case of a call option with maturity 1 and strike $k$ on a single asset, that is, $h(x) = (x - k)^+$ and $d = 1$. For ease of notations, we denote by $\text{CALL}(k)$ the corresponding price. For the cost $c(x, y) = (x - y)^2/2$, the robust call at strike $k$ satisfies

$$\text{CALL}(k) = \inf_{\alpha \in \mathbb{R}} \inf_{\lambda > 0} \left( \text{CALL} \left( k - \frac{2\alpha + 1}{2\lambda} \right) + \frac{\alpha^2}{2\lambda} + \varphi^*(\lambda) \right).$$

Again, if $\varphi(x) = x$, the robust price of a call simplifies to

$$\text{CALL}(k) = \inf_{\alpha \in \mathbb{R}} \left( \text{CALL} \left( k - \frac{2\alpha + 1}{2} \right) + \frac{\alpha^2}{2} \right).$$

Figure 1 represents the standard Black and Scholes price versus robust price as a function of the strike. As expected, the largest spread between the Black and Scholes price and the robust one is at the money, while the cost of robustness vanishes for in or out of the money options.

**Remark 2.15.** The robust price $\text{PRICE}(H)$ can also be interpreted as the minimal superhedging price of $H$ when the shortfall risk is controlled by the nonlinear expectation $\mathcal{R}$. More precisely, Cheridito et al. (2017) give conditions under which the robust price can be represented as

$$\inf \{ m \in \mathbb{R} : \mathcal{R}(-m - \alpha \cdot (S - s) + H) \leq 0 \quad \text{for some } \alpha \in \mathbb{R}^d \}.$$
Remark 2.16 (Robust utility maximization). Another problem in robust mathematical finance where Theorem 2.4 directly applies is utility maximization. Indeed, given a utility function $U : \mathbb{R} \to \mathbb{R}$ bounded from above and a measurable claim $f : \mathbb{R}^d \to \mathbb{R}$, Theorem 2.4 implies that

$$
\sup_{\alpha \in \mathbb{R}^d} \inf_{\mu \in \mathcal{M}_1(\mathbb{R}^d)} \left( \int U(f(s) + \alpha \cdot (s - s)) \, d\mu + \varphi(d_{\mathbb{C}}(\mu_0, \mu)) \right) = \sup_{\alpha \in \mathbb{R}^d} \sup_{\lambda \geq 0} \left( \int U^{\lambda c, \alpha} \, d\mu_0 - \varphi^*(\lambda) \right),
$$

where $U^{\lambda c, \alpha}(x) := \inf_y (U(f(y) + \alpha \cdot (y - s)) + \lambda c(x, y))$. Note that one can also treat the multiperiod case with dynamic programming, see Bartl (2019, Section 2.3) for a discussion of the Wasserstein distance in this framework.

3 | RESULTS FOR GENERAL SETS

3.1 | Tail risk measures and Kusuoka-type representation

For this section, let

$$D \subseteq \mathcal{M}_1(\mathbb{R}) \quad \text{and} \quad R(f) := \sup_{\mu \in D} \int f \, d\mu$$

for $f : \mathbb{R} \to (-\infty, \infty]$ measurable and bounded from below. In the nonrobust setting—that is, $D$ being a singleton—it is well known that the average value-at-risk, see Example 2.10, is a risk measure capturing the “tail risk” by satisfying the representation

$$\text{AV@R}_\alpha = \frac{1}{\alpha} \int_0^\alpha \text{V@R}_\mu \, du,$$

(8)

where V@R is the value-at-risk, see Example 2.12. That is, AV@R is roughly speaking the average over the V@R below the $\alpha$-quantile; an important property, for instance, in optimal portfolio problems, see Rockafellar and Uryasev (2000). However, we will see in Section 3.2 that $\text{AV@R}_\mu$ equals the supremum over $\mu \in D$ of the average value at risk with respect to $\mu$, from which it easily follows that (8), in general, no longer holds true when $D$ consists of more than one element. In a similar manner,
one cannot expect any form of Kusuoka-type representation—abstract version of (8) where any law-invariant risk measure can be represented over the average value at risk—in general.

To prove a robust version of formula (8), stronger assumptions on the set $D$ are needed. If $D \neq \emptyset$ satisfies

$$D \text{ is tight and for all } \mu, \tilde{\mu} \in D \text{ there is } v \in D \text{ such that } \mu(-\infty, t], \tilde{\mu}(-\infty, t] \geq v(-\infty, t] \text{ for all } t \} \quad \text{(DIR)},$$

it also follows that the robust OCE has the same properties as the nonrobust one, see Corollary 4.2. Here, tight means that for every $\varepsilon > 0$, there is some compact set $K \subset \mathbb{R}$ satisfying $\sup_{\mu \in D} \mu(K) \leq \varepsilon$.

**Proposition 3.1.** Assume that (DIR) holds true. Then,

$$A\nu@R_\alpha = \frac{1}{\alpha} \int_0^\alpha A\nu@R_u \, du$$

for every $\alpha \in (0, 1)$.

**Example 3.2.** Let $(\Omega, \mathcal{F}, P)$ be a probability space carrying a Brownian motion $(W_t)_{t \in [0, T]}$, where $T \in (0, \infty)$, equipped with the completion of the natural filtration of $W$. Let $\sigma$, $\tilde{b}$, and $\tilde{b}$ be three real numbers such that $\sigma > 0$. Then, for every strictly increasing function $f : \mathbb{R} \to \mathbb{R}$ and $S_0 > 0$, the set

$$D := \{ P \circ f(S_T)^{-1} \text{ with } dS_t = S_t(b dt + \sigma dW_t) \text{ and } b \in [\tilde{b}, \tilde{b}] \}$$

satisfies (DIR).

Proposition 3.1 suggests a Kusuoka-type representation for robustifications of law-invariant risk measures. In fact, let $\rho$ be the risk measure defined as

$$\rho(\mu) := \sup_{v \in \mathcal{M}_1((0, 1])} \left( \int_{(0, 1]} A\nu@R_u(\mu) \, v(du) - \beta(v) \right)$$

for some penalty function $\beta : \mathcal{M}_1((0, 1]) \to (-\infty, \infty]$, and consider its robust counterpart given by $\rho(D) := \sup_{\mu \in D} \rho(\mu)$. The following corollary gives a representation of $\rho(D)$ in terms of $A\nu@R$. In particular, it shows that when $D$ satisfies (DIR), $A\nu@R$ constitute basic building blocks of robustifications of law-invariant risk measures.

**Corollary 3.3.** Assume that $D$ satisfies (DIR) and is closed in the weak topology. Then, it holds that

$$\rho(D) := \sup_{\mu \in D} \rho(\mu) = \sup_{v \in \mathcal{M}_1((0, 1])} \left( \int_{(0, 1]} A\nu@R_u(D) \, v(du) - \beta(v) \right).$$

### 3.2 Duality

Let $(\Omega, \mathcal{F})$ be a given measurable space endowed with a nonlinear expectation

$$\mathcal{E}(\cdot) := \sup_{P \in \mathcal{M}_1(\Omega)} \left( E_P[\cdot] - \beta(P) \right),$$

for some function $\beta : \mathcal{M}_1(\Omega) \to [0, \infty]$, where $\mathcal{M}_1(\Omega)$ is the set of probability measures on $\mathcal{F}$. In analogy to the first part of the paper, we define the robust OCE as

$$\mathcal{OCE}(X) = \inf_{m \in \mathbb{R}} (\mathcal{E}(l(X - m)) + m)$$
for every measurable function $X : \Omega \to \mathbb{R}$. Here, $l : \mathbb{R} \to \mathbb{R}$ is assumed to satisfy the usual assumptions

\[
\begin{align*}
l & \text{ is convex, increasing, bounded from below, and} \\
l(0) = 0, & \quad l^*(1) = 0, \\
& \text{ and } l(x) > x \text{ for } |x| \text{ large enough,}\end{align*}
\]

(CIB)

and $l^*(y) = \sup_{x \in \mathbb{R}} (xy - l(x))$ denotes the convex conjugate of $l$ for $y \in \mathbb{R}$ and $l^*(\infty) := \infty$. Note that $l^*(y) \geq 0$ and that if $l$ is continuously differentiable, then $l^*(1) = 0$ just says that $l'(0) = 1$. For the remainder of this section, $dQ/dP$ denotes the Radon–Nikodym derivative if $Q$ is absolutely continuous with respect to $P$ and $dQ/dP \equiv \infty$ otherwise.

**Theorem 3.4.** Assume that $l$ satisfies (CIB) and that $\beta$ is convex with $\inf_{P \in \mathcal{M}_1(\Omega)} \beta(P) = 0$. Then, one has

\[
\mathcal{OCE}(X) = \sup_{Q \in \mathcal{M}_1(\Omega)} \left( E_Q[X] - \inf_{P \in \mathcal{M}_1(\Omega)} \left( E_P \left[ l^* \left( \frac{dQ}{dP} \right) \right] + \beta(P) \right) \right)
\]

for every bounded measurable function $X : \Omega \to \mathbb{R}$.

When $\beta$ is the convex indicator of a nonempty convex subset $\mathcal{P}$ of $\mathcal{M}_1(\Omega)$, that is, $\beta = \infty 1_{\mathcal{P}}$, then $\mathcal{E}(\cdot) = \sup_{P \in \mathcal{P}} E_P[\cdot]$ and

\[
\mathcal{OCE}(X) = \sup_{Q \in \mathcal{M}_1(\Omega)} \left( E_Q[X] - \inf_{P \in \mathcal{P}} E_P \left[ l^* \left( \frac{dQ}{dP} \right) \right] \right).
\]

In particular, this implies that the robust OCE is equal to the supremum over $P \in \mathcal{P}$ of the OCE with respect to $P$.

**Example 3.5.** Throughout this example, let $\mathcal{E}(\cdot) = \sup_{P \in \mathcal{P}} E_P[\cdot]$ for some convex set $\mathcal{P}$.

- **Relative entropy:** Let $l(x) = (\exp(ax) - 1)/a$ for some $a > 0$. Then,

\[
\mathcal{OCE}(X) = \sup_{P \in \mathcal{P}} \frac{1}{a} \log E_P[\exp(aX)] = \sup_{Q \in \mathcal{M}_1(\Omega)} \left( E_Q[X] - \inf_{P \in \mathcal{P}} \frac{1}{a} E_Q \left[ \log \frac{dQ}{dP} \right] \right).
\]

This is a generalization of the well-known Gibbs variational principle, and $\inf_{P \in \mathcal{P}} \log dQ/dP$ can be seen as the Kullback–Leibler divergence between the probability measure $Q$ and the set $\mathcal{P}$.

- **Monotone mean variance:** Let $l(x) = (((1 + x)^+)^2 - 1)/2$. Then,

\[
\mathcal{OCE}(X) = \sup_{Q \in \mathcal{M}_1(\Omega)} \left( E_Q[X] - \inf_{P \in \mathcal{P}} E_P \left[ \frac{1}{2} \left( \frac{dQ}{dP} \right)^2 - 1 \right] \right).
\]

The function $\inf_{P \in \mathcal{P}} E_P[(dQ/dP)^2/2 - 1]$ can be seen as the Rényi divergence of order 2 between the probability measure $Q$ and the set $\mathcal{P}$.

- **Average value-at-risk:** Let $l(x) = x^+ / \alpha$ for some $\alpha \in (0, 1)$. Then,

\[
\mathcal{OCE}(X) = \sup \left\{ E_Q[X] : Q \in \mathcal{M}_1(\Omega) \text{ such that } dQ/dP \leq 1/\alpha \text{ for some } P \in \mathcal{P} \right\}.
\]
4 | PROOFS

4.1 | Proof for Section 2

Proof of Theorem 2.4. For every measurable function \( f : X \to (-\infty, \infty] \), which is bounded from below, define

\[
\Phi(f) := \inf_{\lambda \geq 0} \left( \int f^{\lambda c} \, d\mu_0 + \varphi^*(\lambda) \right).
\]

The goal is to apply Choquet’s theorem (in the form of Theorem A.1 in the Appendix) to the functional \( \Phi \), which requires to check the following four steps.

**Step 1:** Monotonicity and convexity. If \( f \leq g \), then \( f^{\lambda c} \leq g^{\lambda c} \) for any \( \lambda \) so that \( \Phi(f) \leq \Phi(g) \). Moreover, for \( t \in [0, 1] \) and \( \lambda', \lambda'' \geq 0 \), it holds that

\[
(f + g)^{\lambda c} \leq t f^{\lambda' c} + (1 - t) g^{\lambda'' c}
\]

for \( \lambda := t\lambda' + (1 - t)\lambda'' \), which implies (also using convexity of \( \varphi^* \)) that \( \Phi(tf + (1 - t)g) \leq t\Phi(f) + (1 - t)\Phi(g) \). Finally, for every \( m \in \mathbb{R} \), \( \lambda \geq 0 \), and \( x \in X \), it holds that

\[
m^{\lambda c}(x) = m - \inf_{y \in X} \lambda c(x, y) = m
\]

so that \( \Phi(m) = \inf_{\lambda \geq 0}(\varphi^*(\lambda) - m) = m \). As \( \Phi \) is monotone, it follows, in particular, that \( \Phi(f) \in \mathbb{R} \) whenever \( f \) is bounded.

**Step 2:** Continuity from above. Denote by \( C_b \) and \( U_b \) the set of bounded continuous and upper semi-continuous functions from \( X \) to \( \mathbb{R} \), respectively. We show that \( \Phi \) is continuous from above on \( C_b \). Let \( \varepsilon > 0 \) and let \((f_n)\) be a sequence in \( C_b \) that decreases pointwise to 0. Fix some \( m \) such that \( f_1 \leq m \) and \( \lambda > 0 \) such that \( \varphi^*(\lambda) < \varepsilon \). This is possible because \( \varphi \) is not constant by assumption; hence, \( \varphi^* \) is real-valued (and therefore continuous by convexity) on some neighborhood of 0. Further, fix \( k \) such that \( \mu_0([-k, k]^c) \leq \varepsilon \). By assumption, there is \( r > 0 \) such that \( \lambda c(x, y) \geq m \) whenever \( |x - y| \geq r \), hence

\[
f_n^{\lambda c}(x) = \sup_{y \in [x-r,x+r]} (f_n(y) - \lambda c(x, y)) \leq \sup_{y \in [x-r,x+r]} f_n(y).
\]

It follows from Dini’s lemma that \( f_n 1_{[-k-r,k+r]} \leq \varepsilon \) for \( n \) large, thus

\[
f_n^{\lambda c} \leq \varepsilon 1_{[-k,k]} + m 1_{[-k,k]^c}
\]

for \( n \) large. Therefore,

\[
\Phi(f_n) \leq \varphi^*(\lambda) + \int f_n^{\lambda c}(x) \mu_0(dx) \leq \varepsilon + \varepsilon \mu_0([-k,k]) + m \mu_0([-k,k]^c) \leq (2 + m)\varepsilon
\]

for \( n \) large and as \( \varepsilon > 0 \) was arbitrary, \( \Phi(f_n) \downarrow 0 = \Phi(0) \).

**Step 3:** Continuity from below. We show that \( \Phi(f_n) \uparrow \Phi(f) \) whenever \((f_n)\) is a sequence of measurable functions \( f_n : X \to (-\infty, \infty] \) bounded from below, which increases to \( f \). Because \( \Phi \) is
increasing, it suffices to show that $\Phi(f) \leq \sup_n \Phi(f_n)$. Assume that $\sup_n \Phi(f_n) < \infty$, because otherwise there is nothing to prove. For every $n$ fix $\lambda_n \geq 0$ such that

$$\varphi^*(\lambda_n) + \int f_n^{\lambda_n} \, d\mu_0 \leq \Phi(f_n) + \frac{1}{n},$$

and $m \in \mathbb{R}$ with $m \leq f_1 \leq f_n$ so that

$$f_n^{\lambda_n}(x) \geq \sup_{y \in X} (m - \lambda_n c(x, y)) = m.$$ 

Note that as $\varphi^*$ is convex and not constant by assumption, $\varphi^*(r_n) \to \infty$ for every sequence $(r_n)$ that converges to $\infty$. Therefore, $(\lambda_n)$ is bounded and, possibly after passing to a subsequence, $(\lambda_n)$ converges to some $\lambda \in [0, \infty)$. Note that

$$f^{\lambda c}(x) = \sup_{y \in X} \lim_n \inf_n (f_n(y) - \lambda_n c(x, y)) \leq \lim_n \inf_n f_n^{\lambda_n c}(x)$$

for every $x$ and by the same argument $\varphi^*(\lambda) \leq \lim_n \inf_n \varphi^*(\lambda_n)$. An application of Fatou’s lemma now implies

$$\Phi(f) \leq \varphi^*(\lambda) + \int f^{\lambda c} \, d\mu_0 \leq \lim_n \inf_n \left( \varphi^*(\lambda_n) + \int f_n^{\lambda_n c} \, d\mu_0 \right) = \sup_n \Phi(f_n) \leq \Phi(f),$$

where the last inequality holds because $\Phi$ is increasing and $f_n \leq f$ for every $n$. Thus, the claim follows.

**Step 4:** Computation of the convex conjugate. We claim that

$$\Phi^*_{C_b}(\mu) := \sup_{f \in C_b} \left( \int f \, d\mu - \Phi(f) \right) = \Phi^*_{U_b}(\mu) := \sup_{f \in U_b} \left( \int f \, d\mu - \Phi(f) \right) = \varphi(d_c(\mu_0, \mu))$$

with the convention that $d_c(\mu_0, \mu) = \infty$ if $\mu$ is not a probability. First, notice that $0 \leq \Phi^*_{C_b} \leq \Phi^*_{U_b}$ because $\Phi(0) = 0$ and $C_b$ is a subset of $U_b$. To show that $\Phi^*_{U_b}(\mu) \leq \varphi(d_c(\mu_0, \mu))$, one may assume that $d_c(\mu_0, \mu) < \infty$; otherwise, this is trivially satisfied (as by assumption $\varphi(\infty) = \infty$). Then, there is a probability $\pi$ on $X \times X$ with marginals $\pi(\cdot \times X) = \mu_0$ and $\pi(X \times \cdot) = \mu$ such that $\int c \, d\pi = d_c(\mu_0, \mu)$, see, for instance, Villani (2008, Theorem 5.9). For any $f \in U_b$ and $\lambda \geq 0$, the pointwise inequality

$$f(y) \leq \lambda c(x, y) + f^{\lambda c}(x) \quad \text{for all } x, y$$

integrated with respect to $\pi$ yields

$$\int_X f(y) \, d\mu(dy) = \int_{X \times X} f(y) \, \pi(dx, dy) \leq \lambda d_c(\mu_0, \mu) + \int_X f^{\lambda c}(x) \, \mu_0(dx).$$

By definition, it holds that $\lambda d_c(\mu_0, \mu) - \varphi^*(\lambda) \leq \varphi(d_c(\mu_0, \mu))$ for all $\lambda \geq 0$ so that $\int f \, d\mu - \Phi(f) \leq \varphi(d_c(\mu_0, \mu))$; hence, $\Phi^*_{C_b}(\mu) \leq \Phi^*_{U_b}(\mu) \leq \varphi(d_c(\mu_0, \mu))$.

To show the reverse inequality, fix some $\mu$ and $\epsilon > 0$. By the Fenchel–Moreau theorem, $\varphi(x) = \sup_{x \geq 0}(\lambda x - \varphi^*(x))$ for every $x \geq 0$; hence, there is $\lambda \geq 0$ such that
\[
\lambda d_c(\mu_0, \mu) - \varphi^*(\lambda) \geq \varphi(d_c(\mu_0, \mu)) - \varepsilon, \quad \text{if } \varphi(d_c(\mu_0, \mu)) < \infty,
\]
\[
\lambda d_c(\mu_0, \mu) - \varphi^*(\lambda) \geq 1/\varepsilon, \quad \text{if } d_c(\mu_0, \mu) < \infty \text{ but } \varphi(d_c(\mu_0, \mu)) = \infty,
\]
\[
\varphi^*(\lambda) \leq \varepsilon \quad \text{and} \quad \lambda > 0 \quad \text{if } d_c(\mu_0, \mu) = \infty.
\]

As \(d_{c^i}(\mu, \mu_0) = \lambda d_c(\mu_0, \mu)\), by the dual formula (5) for \(d_c\), there are bounded and continuous functions \(f, g\) such that
\[
f(x) + g(y) \leq \lambda c(x, y) \quad \text{for all } x, y \text{ and } \int f \, d\mu + \int g \, d\mu_0 \geq \begin{cases} 
\lambda d_c(\mu_0, \mu) - \varepsilon, & \text{if } d_c(\mu_0, \mu) < \infty, \\
1/\varepsilon, & \text{otherwise}.
\end{cases}
\]

Note that \(f(x) - \lambda c(x, y) \leq -g(y)\) from which it follows that \(f^{\lambda c} \leq -g\). As \(\varphi(d_c(\mu_0, \mu)) < \infty\) implies \(d_c(\mu_0, \mu)) < \infty\), we therefore get
\[
\int f \, d\mu - \Phi(f) \geq \int f \, d\mu + \int g \, d\mu_0 - \varphi^*(\lambda) \geq \begin{cases} 
\varphi(d_c(\mu_0, \mu)) - 2\varepsilon, & \text{if } d_c(\mu_0, \mu) < \infty, \\
1/\varepsilon - \varepsilon, & \text{if } d_c(\mu_0, \mu) = \infty.
\end{cases}
\]

As \(\varepsilon > 0\) was arbitrary, it follows that \(\Phi_{c_i}^*(\mu) \geq \varphi(d_c(\mu_0, \mu))\), which by the previous part shows \(\Phi_{C_i}^*(\mu) = \Phi_{U_i}^*(\mu) = \varphi(d_c(\mu_0, \mu))\). The representation (6) now follows from an application of Theorem A.1.

As for the existence of an optimal \(\lambda \geq 0\), apply Step 3 to the constant sequence \(f_n = f\).

**Proof of Theorem 2.7.** First, note that as \(\varepsilon\) depends only on the difference, one has \(l(\cdot - m)^{\lambda c}(x) = l^{\lambda c}(x - m)\) for all \(m \in \mathbb{R}, x \in \mathbb{R}\), and \(\lambda \geq 0\). Now, by Theorem 2.4 one has
\[
\ThetaCE(l) = \inf_{m \in \mathbb{R}} \left( \sup_{\mu \in M_1(\mathbb{R})} \left( \int l(x-m) \mu(dx) - \varphi(d_c(\mu_0, \mu)) \right) + m \right)
\]
\[
= \inf_{m \in \mathbb{R}} \left( \inf_{\lambda \geq 0} \left( \int l(\cdot - m)^{\lambda c}(x) \mu_0(dx) + \varphi^*(\lambda) \right) + m \right)
\]
\[
= \inf_{\lambda \geq 0} \inf_{m \in \mathbb{R}} \left( \int l^{\lambda c}(x-m) \mu_0(dx) + \varphi^*(\lambda) + m \right) = \inf_{\lambda \geq 0} (\text{OCE}(l^{\lambda}) + \varphi^*(\lambda)),
\]
which completes the proof. The same arguments show that
\[
\mathcal{E}S(l) = \inf \left\{ m \in \mathbb{R} : \min_{\lambda \geq 0} \left( \int l^{\lambda c}(x-m) \mu_0(dx) + \varphi^*(\lambda) \right) \leq 0 \right\}
\]
\[
= \inf_{\lambda \geq 0} \inf \left\{ m \in \mathbb{R} : \int l^{\lambda c}(x-m) \mu_0(dx) + \varphi^*(\lambda) \leq 0 \right\}
\]
\[
= \inf_{\lambda \geq 0} \text{ES}(l^{\lambda c} + \varphi^*(\lambda)).
\]

**Proof of Remark 2.9.** We first prove that \(\ThetaCE(l) = \mathcal{E}S(l) = \infty\) if \(\lim_{x \to \infty} c(x) < \infty\) and \(l\) is a loss function. Because \(l\) is increasing, convex, and not constant, there exist \(a, b > 0\) such that \(l(x) \geq ax - b\) for every \(x \in \mathbb{R}\). Because \(\varphi\) is continuous at 0, there is \(\delta > 0\) such that \(\varphi(\delta) < \infty\). Moreover,
by assumption, there is some \( r \in \mathbb{R} \) and a sequence \( (x_k) \) in \( \mathbb{R} \) such that \( x_k \geq k \) and \( c(x_k) \leq r \). For simplicity, let us assume that \( c(0) = 0, \mu_0 = \delta_0 \), and define \( \mu_k = \left(1 - \frac{\delta}{r}\right)\delta_0 + \frac{\delta}{r}\delta_{x_k} \). Then,

\[
d_c(\mu_k, \mu_0) = \left(1 - \frac{\delta}{r}\right)c(0 - 0) + \frac{\delta}{r}c(x_k - 0) \leq \delta
\]

so that \( \varphi(d_c(\mu_k, \mu_0)) \leq \varphi(\delta) < \infty \). However, as

\[
\sup_k \int l(x - m) \mu_k(dx) \geq \sup_k \left(\left(1 - \frac{\delta}{r}\right)l(0 - m) + \frac{\delta}{r}(a(x_k - m) - b)\right) = \infty
\]

for every \( m \in \mathbb{R} \), it follows that \( OCE(l) = ES(l) = \infty \).

To show that \( OCE(l) = ES(l) = \infty \) if \( d_c \) is replaced by a distance compatible with weak convergence, let \( \mu_k = (k - 1)/k\delta_0 + 1/k\delta_{x_k} \), which converges weakly to \( \delta_0 \). As \( \varphi \) is continuous at 0, \( \varphi(d(\mu_k, \delta_0)) \to 0 \). However, \( \sup_k \int l(x - m) \mu_k(dx) = \infty \) for every \( m \), which again implies \( OCE(l) = ES(l) = \infty \).

\[\square\]

**Proof of Example 2.10.** For every \( \lambda \geq 0 \), it holds that

\[
\sup_{{y \in \mathbb{R}}} \left(\frac{1}{\alpha} y^+ - \lambda (x - y)^2\right) = \frac{1}{\alpha} \left(x + \frac{1}{4\lambda \alpha}\right)^+
\]

so that

\[
OCE(l_{\lambda c}) = OCE(l) + \frac{1}{4\lambda \alpha}
\]

for every \( \lambda \geq 0 \). Thus, Theorem 2.7 yields

\[
OCE(l) = OCE(l) + \inf_{\lambda \geq 0} \left(\varphi^*(\lambda) + \frac{1}{4\lambda \alpha}\right).
\]

It remains to plug the different \( \varphi^* \)s and compute the infimum. This proves the claim for \( p = 2 \).

Similarly, for \( p = 1 \), it holds that \( l_{\lambda c}(x) = \infty \) if \( \lambda < 1/\alpha \) and \( l_{\lambda c}(x) = x^+/\alpha \) else. Thus, it follows by Theorem 2.7 that

\[
OCE(l) = OCE(l) + \inf_{\lambda \geq 1/\alpha} \varphi^*(\lambda) = OCE(l) + \varphi^*(\frac{1}{\alpha}),
\]

where the last equality holds because \( \varphi^* \) is increasing.

\[\square\]

**Proof of Example 2.11.** It holds

\[
l_{\lambda c}(x) = \begin{cases} 
\infty, & \text{if } \lambda < 1/2, \\
\frac{2\lambda}{2\lambda - 1}l(x) + \frac{1}{4\lambda - 2}, & \text{else}.
\end{cases}
\]

Thus, for the OCE, one has \( OCE(l_{\lambda c}) = OCE\left(\frac{2\lambda}{2\lambda - 1}l\right) + \frac{1}{4\lambda - 2} \) so that by Theorem 2.7, it holds that

\[
OCE(l) = \inf_{\lambda > 1/2} \left(OCE\left(\frac{2\lambda}{2\lambda - 1}l\right) + \frac{1}{4\lambda - 2} - \varphi^*(\lambda)\right).
\]

\[\square\]
**Proof of Example 2.12.** Note that the value at risk is a special case of the expected shortfall, corresponding to the loss function $l = 1_{(0, \infty)} - \alpha$. Further, with the convention $0^{-1/p} = \infty$, it holds that

$$l^\lambda(x) = l(x) + (1 - \lambda|x|^p)1_{(-\lambda^{-1/p}, 0)}(x) - \alpha$$

for every $x$. Therefore, $\int l^\lambda(x - m) \mu_0(dx) = \mu_0((m, \infty)) + e(m, \lambda) - \alpha$ so that

$$\mathcal{V} @ R_\alpha = \mathcal{E} S(l) = \inf_{\lambda \geq 0} ES(l^\lambda + \varphi^*(\lambda))$$

$$= \inf_{\lambda \geq 0} \inf \{ m \in \mathbb{R} : \mu_0((m, \infty)) + e(m, \lambda) + \varphi^*(\lambda) \leq \alpha \},$$

by Theorem 2.7. The special case $\varphi(x) = x$ follows from Remark 2.8. □

**Proof of Proposition 2.13.** The proof is similar to the one of Theorem 2.4 and we only give a sketch. Denote by $B_{\text{lin}}$ the set of measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ for which $\sup_x f(x)/(1 + |x|)$ is finite, and by $U_{\text{lin}}$ and $C_{\text{lin}}$, the subsets of upper semicontinuous and continuous functions, respectively. For every $f \in B_{\text{lin}}$, define

$$\Phi(f) := \inf_{\alpha \in \mathbb{R}^d, \lambda \geq 0} \left( \int f^{\lambda c, \alpha} d\mu_0 - \varphi(d_\epsilon(\mu_0, \mu)) \right),$$

which is well defined as $f^{\lambda c, \alpha}(x) \geq f(x) - \alpha \cdot (x - s) - c(0) \geq -k(|x| + 1)$ for some $k > 0$ and $\int |x| \mu_0(dx) < \infty$.

As in Theorem 2.4, one checks that $\Phi$ is a convex function on $B_{\text{lin}}$ that is continuous from above on $C_{\text{lin}}$ (here the growth condition $\liminf_{|x| \to \infty} c(x)/(1 + |x|^{1+\epsilon}) = \infty$ is used). Moreover, similar arguments as in Theorem 2.4 show that $\Phi$ is continuous from below on $B_{\text{lin}}$ (here the condition that $\int c(x - y) \mu_0(dx) < \infty$ for every $y$ is used). By a version of Theorem A.1 (see Bartl, Cheridito, & Kupper, 2019, Theorem 2.2), it follows that

$$\Phi(f) = \sup_{\mu \in \mathcal{M}_1(\mathbb{R}^d)} \left( \int f d\mu - \Phi^*_{C_{\text{lin}}}(\mu) \right)$$

provided that

$$\Phi^*_{C_{\text{lin}}}(\mu) := \sup_{f \in C_{\text{lin}}} \left( \int f d\mu - \Phi(f) \right) = \sup_{f \in U_{\text{lin}}} \left( \int f d\mu - \Phi(f) \right) = : \Phi^*_{U_{\text{lin}}}(\mu).$$

As $\Phi(\alpha \cdot (S - s)) \leq 0$ and $\alpha \cdot (S - s) \in C_{\text{lin}}$ for every $\alpha \in \mathbb{R}^d$, similar computations as in Theorem 2.4 yield

$$\Phi^*_{C_{\text{lin}}}(\mu) = \Phi^*_{U_{\text{lin}}}(\mu) = \begin{cases} \varphi(d_\epsilon(\mu_0, \mu)), & \text{if } \int S - s d\mu = 0 \\ \infty, & \text{otherwise.} \end{cases}$$

This ends the proof. □

**Proof of Example 2.14.** We compute

$$h^{\lambda c, \alpha}(x) = \sup_{y \in \mathbb{R}} \left( (y - k)^+ + \alpha(y - s) - \frac{\lambda}{2} (y - x)^2 \right).$$
The first-order conditions yield
\[ 1_{(0,\infty)}(y - k) + \alpha - \lambda(y - x) \geq 0 \geq 1_{(0,\infty)}(y - k) + \alpha - \lambda(y - x). \]

For an optimizer, there are three cases:

- If \( y^* < k \), then \( y^* - x = \alpha / \lambda \), hence \( y^* = \alpha / \lambda + x \) with \( x \in (-\infty, k - \alpha / \lambda) \).
- If \( y^* > k \), then \( y^* - x = (\alpha + 1) / \lambda \), hence \( y^* = (\alpha + 1) / \lambda + x \) with \( x \in (k - (\alpha + 1) / \lambda, \infty) \).
- If \( y^* = k \) is impossible.

Hence, we have three cases:

- For \( x \in (-\infty, k - (\alpha + 1) / \lambda) \), it follows that
  \[ h^{c,a}(x) = \left( \frac{\alpha + 1}{\lambda} + x - k \right) + \alpha\left( \frac{\alpha + 1}{\lambda} + x - s \right) - \frac{(\alpha + 1)^2}{2\lambda} = \left( x - \left( k - \frac{2\alpha + 1}{2\lambda} \right) \right) + \alpha(x - s) + \frac{\alpha^2}{2\lambda}. \]

- For \( x \in [k - \alpha / \lambda, \infty) \), it follows that
  \[ h^{c,a}(x) = \left( \frac{\alpha + 1}{\lambda} + x - k \right) + \alpha\left( \frac{\alpha + 1}{\lambda} + x - s \right) - \frac{(\alpha + 1)^2}{2\lambda} = \left( x - \left( k - \frac{2\alpha + 1}{2\lambda} \right) \right) + \alpha(x - s) + \frac{\alpha^2}{2\lambda}. \]

- For \( x \in (k - (\alpha + 1) / \lambda, k - \alpha / \lambda) \), it follows that
  \[ h^{c,a}(x) = \left( \alpha(x - s) + \frac{\alpha^2}{2\lambda} \right) \vee \left( \left( x - \left( k - \frac{2\alpha + 1}{2\lambda} \right) \right) + \alpha(x - s) + \frac{\alpha^2}{2\lambda} \right). \]

Hence,
\[ h^{c,a}(x) = \left( x - \left( k - \frac{2\alpha + 1}{2\lambda} \right) \right)^+ + \alpha(x - s) + \frac{\alpha^2}{2\lambda}. \]

Therefore, Proposition 2.13 and the fact that \( \mu_0 \) is a pricing measure yield
\[ \text{CALL}(k) = \inf_{\alpha \in \mathbb{R}} \inf_{\lambda > 0} \left( \text{CALL} \left( k - \frac{2\alpha + 1}{2\lambda} \right) + \frac{\alpha^2}{2\lambda} + \varphi^*(\lambda) \right). \]

Plugging the special cases of \( \varphi \) into this equation yields the claim.

\[ \square \]

### 4.2 | Proofs for Section 3.1

The main argument for the proof of Proposition 3.1 is given in the next lemma.

**Lemma 4.1.** Assume that \( D \) satisfies (DIR). Then, there exists \( \mu^* \in \mathcal{M}_1 \) such that
\[ \sup_{\mu \in D} \int f \, d\mu = \int f \, d\mu^* \]
for every increasing, continuous function \( f : \mathbb{R} \to \mathbb{R} \) that is bounded from below. If, in addition, \( D \) is closed in the weak topology induced by all continuous bounded functions, then \( \mu^* \in D \).
Proof. First, assume that $f$ is bounded. As $D$ is tight, it can be checked that $\overline{F}$ defined by $\overline{F}(t) := \inf_{\mu \in D} F_\mu(t)$ where $F_\mu(t) := \mu(-\infty, t]$ is a cumulative distribution function. Furthermore, $f$ being increasing, continuous, and bounded, it defines a finite Borel measure $df$ on the real line. Hence, $df$ is regular and $\tau$-additive, see, for instance, Bogachev (2007, Proposition 7.2.2). Let us first show that

$$\int \overline{F} df = \inf_{\mu \in D} \int F_\mu df. \tag{11}$$

Each cumulative distribution function $F_\mu$ is increasing and right-continuous, hence upper semicontinuous. Because $D$ satisfies (DIR), the net $(F_\mu)_{\mu \in D}$ is decreasing. Thus, $(1 - F_\mu)_{\mu \in D}$ is an increasing net of nonnegative lower semicontinuous functions such that $1 - \overline{F} = \lim_{\mu} (1 - F_\mu)$. It therefore follows from Bogachev (2007, Lemma 7.2.6) that

$$\sup_{\mu \in D} \int 1 - F_\mu df = \lim_{\mu} \int 1 - F_\mu df = \int 1 - \overline{F} df,$$

which shows (11). Moreover, as $f$ is continuous, one has $\int \overline{F}(x-) df(x) = \int \overline{F}(x) df(x)$. Hence, integration by parts yield

$$\int f \overline{F} = f(\infty) - \int \overline{F} df = f(\infty) - \int \inf_{\mu \in D} F_\mu df = f(\infty) - \inf_{\mu \in D} \int F_\mu df$$

$$= f(\infty) - \inf_{\mu \in D} \left( f(\infty) - \int f dF_\mu \right) = \sup_{\mu \in D} \int f dF_\mu,$$

showing (10) whenever $f$ is bounded, with $\mu^*$ being the distribution associated with $\overline{F}$. If $f$ is not bounded, we approximate $f$ from below by $f^n := f \wedge n$.

If $D$ is also closed, then it follows from Prokhorov’s theorem and tightness that $D$ is compact. Suppose for contradiction that $\mu^* \notin D$. Then, by the strong separation theorem and (10) that was already proven, there exists a continuous bounded and increasing function $f : \mathbb{R} \rightarrow \mathbb{R}$ such that $\int f d\mu^* > \sup_{\mu \in D} \int f d\mu$, which clearly contradicts (10). Thus, $\mu^* \in D$. $\square$

Corollary 4.2. Assume that (DIR) holds and that $l : \mathbb{R} \rightarrow \mathbb{R}$ is convex, increasing, bounded from below, and that $l(x) > x$ for $|x|$ large enough. Then, there exists $\mu^* \in \mathcal{M}_1$ and $m^* \in \mathbb{R}$ such that

$$\mathcal{OE}(l) = \inf_{m \in \mathbb{R}} \left( \int l(x - m) \mu^*(dx) + m \right) = \int l(x - m^*) \mu^*(dx) + m^*.$$

In particular, $m^*$ is characterized by

$$\int l_-(x - m^*) \mu^*(dx) \leq 1 \leq \int l_+(x - m^*) \mu^*(dx)$$

for the right- and left-hand derivatives $l_-$ and $l_+$ of $l$. If $l$ is continuously differentiable, then inequalities in the above formula are equalities.

Proof. The existence of a $\mu^* \in \mathcal{M}_1$ such that $\mathcal{OE}(l) = \inf_{m \in \mathbb{R}} (\int l(x - m) \mu^*(dx) + m)$ follows directly from Lemma 4.1. Therefore, the existence and characterization of an optimal allocation $m^*$ can be deduced from the nonrobust case, see, for instance, Ben-Tal and Teboulle (2007). $\square$
Proof of Proposition 3.1. It follows from Lemma 4.1 and Corollary 4.2 that one has $\mu^*(-\infty, t] = \inf_{\mu \in D} \mu((-\infty, t])$ for every $t$, which implies $\mathcal{V}@R_a = \inf\{m \in \mathbb{R} : \mu^*((m, \infty)) \leq a\}$ and $\mathcal{A} \mathcal{V}@R_a = \inf_{m \in \mathbb{R}} (\int (-m + m) / \alpha \mu^*(dx) + m)$. Thus, Föllmer and Schied (2011, Lemma 4.51) yields

$$\mathcal{A} \mathcal{V}@R_a = \frac{1}{\alpha} \int_0^a \inf\{m \in \mathbb{R} : \mu^*((m, \infty)) \leq u\} du = \frac{1}{\alpha} \int_0^a \mathcal{V}@R_u du.$$ 

Proof of Example 3.2. For every $b \leq b \leq b$, the process $S_t = S_0 \exp((b - \frac{1}{2} \sigma^2)t + \sigma W_t)$ is the solution of $dS_t = S_t (b dt + \sigma dW_t)$. Further, as $S_0 > 0$ and $f$ is strictly increasing, one has

$$P\left(f \left(S_T^b \right) \leq x \right) \leq P\left(f \left(S_T^a \right) \leq x \right) \leq P\left(f \left(S_T^b \right) \leq x \right)$$

for all $x$. Thus, a straightforward computation shows that $D$ satisfies (DIR).

Proof of Corollary 3.3. By Lemma 4.1, there is $\mu^*$ such that $\mathcal{A} \mathcal{V}@R_u = \mathcal{A} \mathcal{V}@R_u(\mu^*)$ for every $u \in (0, 1]$. Hence, it follows by definition that

$$\rho(D) \leq \sup_{\nu \in \mathcal{M}_1([0, 1])} \left( \int_{(0, 1]} \mathcal{A} \mathcal{V}@R_u \nu(du) - \beta(\nu) \right).$$

On the other hand, as $D$ is closed, it holds that $\mu^* \in D$ and thus $\sup_{\mu \in D} \int_{(0, 1]} \mathcal{A} \mathcal{V}@R_u(\mu) \nu(du) \geq \int_{(0, 1]} \mathcal{A} \mathcal{V}@R_u(\mu^*) \nu(du) = \int_{(0, 1]} \mathcal{A} \mathcal{V}@R_u(\nu(du))$, which proves the reverse inequality.

Proof of Theorem 3.4. Denote by $\text{dom}(\beta) := \{P \in \mathcal{M}_1(\Omega) : \beta(P) < \infty\}$ the domain of $\beta$ that is a nonempty convex set by assumption. Therefore, the function $J$ defined by

$$J : \mathbb{R} \times \text{dom}(\beta) \to \mathbb{R}, \quad (m, P) \mapsto E_P[l(X - m)] + m - \beta(P)$$

is convex and continuous in $m$, and concave in $P$. Moreover, as $X$ is a bounded random variable and $l$ is bounded from below with $\lim_{x \to \infty} l(x)/x = \infty$ by assumption, there exists some $m_0 \in \mathbb{R}$ such that the first and last equalities in the following equation hold

$$\inf_{m \in \mathbb{R}} \sup_{P \in \text{dom}(\beta)} J(m, P) = \inf_{m \in [-m_0, m_0]} \sup_{P \in \text{dom}(\beta)} J(m, P) = \sup_{P \in \text{dom}(\beta)} \inf_{m \in [-m_0, m_0]} J(m, P) = \sup_{P \in \text{dom}(\beta)} \inf_{m \in \mathbb{R}} J(m, P).$$

The middle equality follows from a minimax theorem, see, for instance, Fan (1953, Theorem 2). Now notice that the left-hand side equals $\mathcal{OCE}(X)$ and the right-hand side the supremum over $P \in \text{dom}(\beta)$ of the OCE under $P$. In particular, it follows from the classical representation of the OCE that

$$\mathcal{OCE}(X) = \sup_{P \in \text{dom}(\beta)} \sup_{Q \in \mathcal{M}_1(\Omega)} \left( E_Q[X] - E_P \left[ l\left( \frac{dQ}{dP} \right) \right] - \beta(P) \right)$$

$$= \sup_{Q \in \mathcal{M}_1(\Omega)} \left( E_Q[X] - \inf_{P \in \mathcal{M}_1(\Omega)} \left( E_P \left[ l\left( \frac{dQ}{dP} \right) \right] + \beta(P) \right) \right).$$

\[ \square \]
ENDNOTES

1 $\varphi^*$ denotes the convex conjugate of $\varphi$, that is, $\varphi^*(y) = \sup_x (xy - \varphi(x))$.

2 $D$ is endowed with the ordering $\mu \preceq \nu$ if and only if $\mu(-\infty, t] \geq \nu(-\infty, t]$ for every $t$.

REFERENCES

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

Backhoff, J., & Tangpi, L. (2016). On the dynamic representation of some time-inconsistent risk measures in a Brownian filtration. Preprint arXiv:1608.07498.

Bartl, D. (2019). Exponential utility maximization under model uncertainty for unbounded endowments. The Annals of Applied Probability, 29(1), 577–612.

Bartl, D., Cheridito, P., & Kupper, M. (2019). Robust expected utility maximization with medial limits. Journal of Mathematical Analysis and Applications, 471(1-2), 752–775.

Beiglböck, M., Henry-Labordère, P., & Penkner, F. (2011). Model-independent bounds for option prices: A mass transport approach. Finance and Stochastics, 17(3), 477–501.

Ben-Tal, A. & Teboulle, M. (1986). Expected utility, penalty functions and duality in stochastic nonlinear programming. Management Science, 32, 1445–1466.

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

Bertsekas, D., & Shreve, S. (1978). Stochastic optimal control: The discrete-time case (Volume 23). New York: Academic Press.

Blanchard, R., & Carassus, L. (2018). Multiple-priors optimal investment in discrete time for unbounded utility function. The Annals of Applied Probability, 28(3), 1856–1892.

Blanchet, J., Chen, L., & Zhou, Y. (2018). Distributionally robust mean-variance portfolio selection with Wasserstein distances. Preprint arXiv:1802.04885.

Blanchet, J., & Murthy, K. (2016). Quantifying distributional model risk via optimal transport. Preprint arXiv:1604.01446.

Bogachev, V. I. (2007). Measure theory (Volume 2). New York: Springer.

Bolley, F., Guillin, A., & Villani, C. (2007). Quantitative concentration inequalities for empirical measures on non-compact spaces. Probability Theory Related Fields, 137, 541–593.

Cheridito, P., Kupper, M., & Tangpi, L. (2017). Duality formulas for robust pricing and hedging in discrete time. SIAM Journal on Financial Mathematics, 8(1), 738–765.

Cheridito, P., & Li, T. (2009). Risk measures on Orlicz hearts. Mathematical Finance, 19(2), 189–214.

Cherny, A. S., & Kupper, M. (2007). Divergence utilities. Available at SSRN: https://ssrn.com/abstract=1023525.

Damodaran, A. (2008). Strategic risk taking: A framework for risk management. Waterloo, NSW: Wharton School Publishing.

Dereich, S., Scheutzow, M., & Schottstedt, R. (2008). Constructive quantization: Approximation by empirical measures. Annales de l’Institut Henri Poincaré Probability and Statistics, 49(4), 1183–1203.

Dolinsky, Y., & Soner, H. (2014). Martingale optimal transport and robust hedging in continuous time. Probability Theory Related Fields, 160(1), 391–427.

Drapeau, S., Kupper, M., & Papapantoleon, A. (2014). A Fourier approach to the computation of CVaR and optimized certainty equivalents. Journal of Risk, 16(6), 3–29.

Dunkel, J., & Weber, S. (2010). Stochastic root finding and efficient estimation of convex risk measures. Operations Research, 58(5), 1505–1521.

Esfahani, P. M., & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. Mathematical Programming, 171, 115. https://doi.org/10.1007/s10107-017-1172-1.
Fan, K. (1953). Minimax theorems. *Proceedings of the National Academy of Sciences of the United States of America*, 39, 42–47.

Fournier, N., & Guillin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory Related Fields*, 162(3–4), 707–738.

Föllmer, H., & Schied, A. (2002). Convex measures of risk and trading constraint. *Finance and Stochastics*, 6(4), 429–447.

Föllmer, H., & Schied, A. (2011). *Stochastic finance: An introduction in discrete time* (3rd ed.). Berlin: Walter de Gruyter.

Frittelli, M., & Gianin, E. R. (2002). Putting order in risk measures. *Journal of Banking and Finance*, 26(7), 1473–1486.

Gao, R., & Kleywegt, A. J. (2016). *Distributionally robust stochastic optimization with Wasserstein distance*. Preprint arXiv:1604.02199.

Gibbs, A., & Su, F. E. (2002). On choosing and bounding probability metrics *International Statistical Review*, 70(3), 419–435.

Hamm, A.-M., Salfeld, T., & Weber, S. (2014). Stochastic root finding for optimized certainty equivalents. In *Proceedings of the 2013 Winter Simulation Conference*, Pasupathy.

Hansen, L. P., & Sargent, T. J. (2001). Robust control and model uncertainty. *American Economic Review*, 91(2), 60–66.

Hobson, D. (1998). Robust hedging of the lookback option. *Finance and Stochastics*, 2(4), 329–347.

Huber, P. J. (1981). *Robust statistics*. Hoboken, NJ: John Wiley & Sons.

Jouini, E., Schachermayer, W., & Touzi, N. (2006). Law invariant risk measures have the Fatou property. *Advanced Mathematical Economics*, 9, 49–71.

Kusuoka, S. (2001). On law-invariant coherent risk measures. *Advanced Mathematical Economics*, 3, 83–93.

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

Matoussi, A., Possamaï, D., & Zhou, C. (2015). Robust utility maximization in nondominated models with 2BSDE: The uncertain volatility model. *Mathematical Finance*, 25(2), 258–287.

Neufeld, A., & Sikic, M. (2018). Robust utility maximization in discrete-time markets with frictions. *SIAM Journal on Control and Optimization*, 56(3), 1912–1937.

Nutz, M. (2016). Utility maximization under model uncertainty in discrete time. *Mathematical Finance*, 26(2), 252–268.

Pflug, G., Pichler, A., & Wozabal, D. (2012). The $1/n$ investment strategy is optimal under high model ambiguity. *Journal of Banking and Finance*, 36, 410–417.

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

Villani, C. (2008). *Optimal transport: Old and new* (Volume 338). Berlin: Springer Science & Business Media.

Wozabal, D. (2014). Robustifying convex risk measures for linear portfolio: A nonparametric approach. *Operations Research Letters*, 42(6), 1302–1315.

Zhao, C., & Guan, Y. (2018). Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters*, 46(2), 262–267.

**How to cite this article:** Bartl D, Drapeau S, Tangpi L. Computational aspects of robust optimized certainty equivalents and option pricing. *Mathematical Finance*. 2020;30:287–309. https://doi.org/10.1111/mafi.12203

**APPENDIX A**

Let $X \subseteq \mathbb{R}^d$ be closed, denote by $B_b^{-}$ the set of all measurable functions from $X$ to $(-\infty, \infty]$ that are bounded from below, and by $U_b$ and $C_b$ the subsets of bounded upper semicontinuous (respectively,
bounded continuous) functions. Further, write \( \mathcal{M}(X) \) for the set of all countably additive, finite, positive Borel measures on \( X \), that is, \( \mathcal{M}(X) := \{t\mu : \mu \in \mathcal{M}_1(X) \text{ and } t \geq 0\} \). The following theorem, which builds on Choquet’s theory on the regularity of capacities, is a slight modification of Theorem 2.2 in Bartl et al. (2019).

**Theorem A.1.** Let \( \Phi : B_{b^-} \rightarrow (-\infty, \infty] \) be a monotone convex functional such that \( \Phi(f) < \infty \) whenever \( f \) is bounded. If

- \( \Phi(f_n) \downarrow \Phi(0) \) for every sequence \( (f_n) \) in \( C_b \) that decreases pointwise to 0,
- \( \Phi^*_c(\mu) := \sup_{f \in C_b} (\int f \, d\mu - \Phi(f)) = \sup_{f \in U_b} (\int f \, d\mu - \Phi(f)) =: \Phi^*_u(\mu) \) for every \( \mu \in \mathcal{M}(X) \),
- \( \Phi(f_n) \uparrow \Phi(f) \) for every sequence \( (f_n) \) in \( B_{b^-} \) that increases pointwise to \( f \in B_{b^-} \),

then

\[
\Phi(f) = \sup_{\mu \in \mathcal{M}(X)} \left( \int f \, d\mu - \Phi^*_c(\mu) \right) \quad \text{for every } f \in B_{b^-}. \tag{A.1}
\]

---

*Proof.* If \( B_{b^-} \) is replaced by the set of all bounded measurable functions, this is exactly the statement of Bartl et al. (2019, Theorem 2.2), so that (A.1) holds for all bounded measurable \( f \). For general \( f \in B_{b^-} \), notice that (A.1) holds for every \( f \wedge n \); hence, the claim follows from the third assumption \( \Phi(f) = \sup_n \Phi(f \wedge n) \), interchanging two suprema, and the monotone convergence theorem applied under each \( \mu \in \mathcal{M}(X) \). \( \square \)