Robust distortion risk measures

Carole Bernard1,2 | Silvana M. Pesenti3 | Steven Vanduffel2

1Department of Accounting, Law and Finance, Grenoble Ecole de Management, Grenoble, France
2Department of Economics and Political Science, Vrije Universiteit Brussel, Brussels, Belgium
3Department of Statistical Sciences, University of Toronto, Toronto, Canada

Abstract

The robustness of risk measures to changes in underlying loss distributions (distributional uncertainty) is of crucial importance in making well-informed decisions. In this paper, we quantify, for the class of distortion risk measures with an absolutely continuous distortion function, its robustness to distributional uncertainty by deriving its largest (smallest) value when the underlying loss distribution has a known mean and variance and, furthermore, lies within a ball—specified through the Wasserstein distance—around a reference distribution. We employ the technique of isotonic projections to provide for these distortion risk measures a complete characterization of sharp bounds on their value, and we obtain quasi-explicit bounds in the case of Value-at-Risk and Range-Value-at-Risk. We extend our results to account for uncertainty in the first two moments and provide applications to portfolio optimization and to model risk assessment.

KEYWORDS

distortion risk measures, distributional robustness, model uncertainty, Range Value-at-Risk, risk bounds, Wasserstein distance
1 | INTRODUCTION

Many decisions in financial risk management are based on evaluations of so-called risk or performance measures, for example, Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR) (BCBS, 2013; EIOPA, 2009), which are limit cases of Range Value-at-Risk (RVaR) (Cont et al., 2010). Risk managers rely on such measures, as they provide a classification of risk severities by associating to every loss distribution a real number (Artzner et al., 1999). Many risk measures are, however, highly sensitive to changes in the underlying loss distribution, and this vulnerability can contribute to adverse decisions; see Cont et al. (2010), Kou et al. (2013), Embrechts et al. (2015), and Pesenti et al. (2016). Here, we quantify distributional uncertainty via bounds on the values of risk measures. Specifically, we study worst-case and best-case values of risk measures—that is, the largest (resp. smallest) value a risk measure can attain when the loss distribution belongs to a set of plausible alternative distributions.

The literature on risk bounds presents two main approaches. The first studies risk bounds for the aggregate portfolio loss $S = X_1 + \cdots + X_n$ given that the components have known marginal distributions but unknown interdependence; see, among others, Denuit et al. (1999), Embrechts and Puccetti (2006), Wang and Wang (2011), and Embrechts et al. (2013). Although explicit results can be derived in certain cases of interest, the bounds obtained are typically too wide to be practically useful. Specifically, the upper bound on a distortion risk measure is always at least as large as its value under the assumption of a comonotonic dependence, a situation that is arguably very extreme. Papers that study bounds with additional dependence constraints or by imposing modeling structures include Bernard et al. (2017), Lux and Rüschendorf (2019), and Wang et al. (2019). Another stream of the literature is concerned with deriving risk bounds of the aggregate loss $S$ under partial knowledge of its moments. Since only information on the moments of $S$ is required, these risk bounds apply to nonlinear loss models, such as re-insurance portfolios, and to loss models for which the explicit aggregation is not known analytically, for example, when $S$ is obtained via simulation models. As (higher) moments of the aggregate loss $S$ depend both on the marginal distributions and their interdependence, one, thus, retains some aspects of the marginal distributions and dependence information. Relevant papers include Hürlimann (2002) for the case of VaR and Cornilly et al. (2018) and Zhu and Shao (2018) for the case of (concave) distortion risk measures. Risk bounds that use additional refined information, such as uni-modality or symmetry, can be found in, for example, Li et al. (2018).

This paper is situated in the second stream, in that we consider risk bounds of an aggregate loss $S$ under knowledge of its first two moments. However, we additionally impose a probability distance constraint on $S$, specified via the Wasserstein distance. Specifically, all alternative aggregate loss distributions, over which the worst- and best-cases are sought—the uncertainty set—lie within a tolerance distance from a reference loss distribution. Thus, the size of the tolerance distance determines the uncertainty around the reference distribution, which may also entail uncertainties on its components. Indeed, if the tolerance is small enough, the uncertainty set contains only the reference distribution, whereas if the tolerance distance increases to infinity, we recover bounds with moments constraints only. Therefore, the tolerance distance allows for a refined notion of model risk, which results in a continuous spectrum of risk bounds interpolating from the risk of the reference distribution to the bounds on the aggregate loss under the knowledge of the first two moments only.

Worst- (best-) case values under probability distance constraints have been considered by Glasserman and Xu (2014) and Lam (2016), who use the Kullback–Leibler divergence, as well as by Blanchet and Murthy (2019), who utilize distances stemming from mass transportation. These
papers, however, consider the expected value of a function of the loss random variable, whereas here we study the class of distortion risk measures. Moreover, we use a distance that allows for comparison of distributions on differing support, which is important in our context as it is not clear a priori whether the worst-case distribution shares the same support as the reference distribution. For example, a discrete reference distribution, together with a particular choice of distortion risk measure (e.g., inverse-S-shaped), will result in a continuous worst-case distribution. In this work, we quantify the distance between the alternative distribution function and the reference distribution via the Wasserstein distance of order 2. The Wasserstein distance has been widely applied to model distributional uncertainty in financial contexts (see, e.g., Bartletic al., 2020, Chen & Xie, 2021, and Pflug & Wozabal, 2007), and as a distance from reference model (see e.g., Pesenti, 2022 and Pesenti & Jaimungal, 2023).

A related stream of literature studies distributionally robust optimization, leading to an optimum in the worst-case. Relevant literature, in which the worst-case value of a distortion risk measure is considered, includes Cai et al. (2023) and Pesenti et al. (2020). Cai et al. (2023) consider distortion risk measures under the assumption that the uncertainty set of the components $X_1, \ldots, X_n$ is characterized by the knowledge of their moments, the components having compact support and additional convex constraints. Pesenti et al. (2020) consider signed Choquet Integrals (see Section 3.15 for a definition) and uncertainty sets characterized by so-called closedness-under-concentration. Neither of these works considers the Wasserstein distance, and the closedness-under-concentration condition is not compatible with uncertainty sets that are $\sqrt{\varepsilon}$-Wasserstein balls around an (arbitrary) reference distribution.

In this paper, we derive for any given distortion risk measure (with absolutely continuous distortion function) its worst- and best-case values when the loss distribution has known first two moments and lies within a tolerance distance from a reference distribution, specified through the Wasserstein distance. To derive these bounds, we use the technique of isotonic projections (Németh, 2003). As a result, we obtain quasi analytical best- and worst-case bounds that significantly improve existing moment bounds. We find that for small Wasserstein tolerance distances, the worst-case distribution function (the distribution function attaining the worst-case value) is close (in a probabilistic sense) to the reference distribution, whereas for large tolerance distances, the worst-case distribution function is no longer affected by the reference distribution. Thus, when the Wasserstein tolerance distance goes to infinity, the reference distribution becomes irrelevant, and we recover, as corollaries to our results, known moment bounds on distortion risk measures; see Li (2018) and Zhu and Shao (2018). Indeed, Li (2018) considers worst-case concave distortion risk measures with given mean and standard deviation, and Zhu and Shao (2018) consider the best and worst-cases of the entire class of distortion risk measure under the knowledge of the first two moments. Neither considers a Wasserstein constraint.

We further apply our results to obtain quasi-explicit bounds on RVaR and (via a direct proof) VaR. We find that for small Wasserstein tolerance distances, the worst-case distribution functions for RVaR and VaR are no longer two-point distributions, thus making the bounds attractive for financial risk management applications.

This paper is structured as follows: Section 2 introduces the necessary notation and formulates the problem. The main results—the quasi-analytical best- and worst-case values of distortion risk measures under a Wasserstein distance constraint—are presented in Section 3. Section 4 contains two extensions. In Section 4.1, we calculate bounds of VaR and RVaR. In Section 4.2, we extend our results to complete moment uncertainty—that is, when the uncertainty set is a $\sqrt{\varepsilon}$-Wasserstein ball around a reference distribution without any moment constraints. We provide applications of the risk bounds to portfolio optimization and to model risk assessment of an insurance portfolio,
including rationales for choosing the Wasserstein distance, in Section 5. All proofs are relegated to the appendix.

2 PROBLEM FORMULATION

We consider an atomless probability space \((\Omega, \mathcal{A}, \mathbb{P})\) and let \(L^2 = L^2(\Omega, \mathcal{A}, \mathbb{P})\) be the set of all square integrable random variables on that space. We denote by \(\mathcal{M}^2 = \{ G(x) = \mathbb{P}(X \leq x) \mid X \in L^2 \}\) the corresponding space of distribution functions with finite second moment. The (left-continuous) inverse, also called the quantile function, of a distribution function \(G \in \mathcal{M}^2\) is \(G^{-1}(u) = \inf\{y \in \mathbb{R} \mid G(y) \geq u\}\), and we denote its right-continuous inverse by \(G^{-1,+}(u) = \inf\{y \in \mathbb{R} \mid G(y) > u\}\), \(0 < u < 1\). Throughout the exposition, we write \(U \sim \mathbb{U}(0,1)\) for a standard uniform random variable on \((0,1)\).

2.1 Distortion risk measures

In this section, we recall the class of distortion risk measures that contains risk measures commonly encountered in financial applications, such as VaR and TVaR. A distortion risk measure, evaluated at a distribution function \(G \in \mathcal{M}^2\), is defined via the Choquet integral

\[
H_g(G) = - \int_{-\infty}^{0} 1 - g(1 - G(x)) \, dx + \int_{0}^{+\infty} g(1 - G(x)) \, dx,
\]

whenever at least one of the two integrals is finite. The function \(g : [0, 1] \rightarrow [0, 1]\) refers to a distortion function—that is, a non-decreasing function satisfying \(g(0) = 0\) and \(g(1) = 1\). If \(g\) is absolutely continuous, then the distortion risk measure \(H_g\) has the following representation (Dhaene et al., 2012):

\[
H_g(G) = \int_{0}^{1} \gamma(u) G^{-1}(u) \, du,
\]

with weight function \(\gamma(u) = \partial_- g(x)|_{x=1-u}, 0 < u < 1\), which satisfies \(\int_{0}^{1} \gamma(u) \, du = 1\) and where \(\partial_-\) denotes the derivative from the left. In what follows, we may sometimes write \(H_g(G^{-1})\) instead of \(H_g(G)\).

Assumption 2.1. We assume that representation (1) holds and that \(\int_{0}^{1} |\gamma(u)|^2 \, du < +\infty\).

Throughout the paper, we use the following three concave distortion risk measures to illustrate our results: the dual power distortion with parameter \(\beta > 0\),

\[
g(x) = 1 - (1 - x)^\beta, \quad \text{and} \quad \gamma(u) = \beta u^{\beta - 1};
\]

the Wang transform (Wang, 1996) with parameter \(0 < q_0 < 1\),

\[
g(x) = \Phi(\Phi^{-1}(x) + \Phi^{-1}(q_0)), \quad \text{and} \quad \gamma(u) = \frac{\phi(\Phi^{-1}(1-u) + \Phi^{-1}(q_0))}{\phi(\Phi^{-1}(1-u))},
\]

where \(\Phi^{-1}\) denotes the inverse of the cumulative distribution function of the standard normal distribution, and \(\phi\) denotes the probability density function of the standard normal distribution.
where $\Phi$ and $\phi$ denote the standard normal distribution and its density, respectively; and the Tail Value-at-Risk (TVaR$_{\alpha}$), also called Expected Shortfall, at level $0 < \alpha < 1$ (Acerbi, 2002) with

$$g(x) = \min \{ \frac{x}{1-\alpha}, 1 \} \quad \text{and} \quad \gamma(u) = \frac{1}{1-\alpha} \mathbb{1}_{(\alpha,1)}(u).$$

(4)

In Section 4.1, we also consider the case of Value-at-Risk (VaR$_{\alpha}$), $0 < \alpha < 1$, a distortion risk measure with $g(x) = \mathbb{1}_{(1-\alpha,1]}(x)$ that does not satisfy Assumption 2.1, that is, it does not admit a representation as in Equation (1) as a Lebesgue integral.

### 2.2 Modeling distributional uncertainty

In financial risk management, distributional uncertainty is prevalent and bounds on the value of a risk measure, so-called worst- and best-case values or upper and lower bounds, are valuable tools for decision making. A worst- (best-) case value of a risk measure is defined as the largest (smallest) value a risk measure can attain when the underlying distribution belongs to a set of alternative distributions. Here, we consider subsets of $\mathcal{M}^2$ that are characterized via a tolerance distance to a reference distribution, specified through the Wasserstein distance of order 2. Distribution functions belonging to such an uncertainty set may be viewed as alternatives to the reference distribution, representing model misspecification or alternative model assumptions. To formalize this, we recall the definition of the Wasserstein distance. The Wasserstein distance exhibits desirable properties, such as no restriction on the support of distribution functions, which is in contrast to the Kullback–Leibler divergence; see Villani (2008) for a discussion of the Wasserstein distance.

**Definition 2.2 (Wasserstein distance of order 2).** The Wasserstein distance of order 2 between $G_1, G_2 \in \mathcal{M}^2$ is (Villani, 2008)

$$d_W(G_1, G_2) = \inf \left\{ \left[ E \left( (X_1 - X_2)^2 \right) \right]^{\frac{1}{2}} \left| X_1 \sim G_1, X_2 \sim G_2 \right. \right\},$$

where the infimum is taken over all bivariate distributions with marginals $G_1$ and $G_2$.

For distributions on the real line, the Wasserstein distance admits the following well-known representation (Dall’Aglio, 1956)

$$d_W(G_1, G_2) = \left( \int_0^1 \left( G_1^{-1}(u) - G_2^{-1}(u) \right)^2 du \right)^{\frac{1}{2}}.$$

Hence, the Wasserstein distance between $G_1$ and $G_2$ is uniquely determined by their corresponding quantile functions, and we may write $d_W(G_1^{-1}, G_2^{-1})$ instead of $d_W(G_1, G_2)$.

We denote by $F \in \mathcal{M}^2$ the reference distribution and its first two moments by $\int x \, dF(x) = \mu_F \in \mathbb{R}$ and $\int x^2 \, dF(x) = \mu_F^2 + \sigma_F^2$, $\sigma_F > 0$, respectively. Throughout, we fix the reference distribution $F$ and consider distributional uncertainty sets of the type

$$\mathcal{M}_\varepsilon(\mu, \sigma) = \left\{ G \in \mathcal{M}^2 \left| \int x \, dG(x) = \mu, \int x^2 \, dG(x) = \mu^2 + \sigma^2, d_W(F, G) \leq \sqrt{\varepsilon} \right. \right\},$$
where \( \mu \in \mathbb{R}, \sigma > 0, \) and \( 0 \leq \varepsilon \leq +\infty. \) The set \( \mathcal{M}_\varepsilon(\mu, \sigma), \) thus, contains all distribution functions whose first two moments are \( \mu \) and \( \mu^2 + \sigma^2, \) respectively, and that lie within a \( \sqrt{\varepsilon} \)-Wasserstein ball around the reference distribution \( F. \) The \( \varepsilon \)-constraint in the definition of \( \mathcal{M}_\varepsilon(\mu, \sigma) \) is redundant when \( \varepsilon = +\infty, \) and in this special case, we denote the uncertainty set \( \mathcal{M}_\infty(\mu, \sigma) \) by \( \mathcal{M}(\mu, \sigma). \) Note that in applications, one may often encounter that \( (\mu, \sigma) = (\mu_F, \sigma_F), \) but our set-up allows them to be different, which could for example be useful when their values are estimated using different data sources.

The best- and worst-case values of a distortion risk measure \( H_g \) over the distributional uncertainty set \( \mathcal{M}_\varepsilon(\mu, \sigma) \) are, respectively, defined by

\[
\begin{align*}
\inf_{G \in \mathcal{M}_\varepsilon(\mu, \sigma)} H_g(G) & \quad (5a) \\
\sup_{G \in \mathcal{M}_\varepsilon(\mu, \sigma)} H_g(G). & \quad (5b)
\end{align*}
\]

In addition to the best- and worst-case values, we also study best- and worst-case distribution functions if they exist—that is, the distribution functions attaining Equations (5a) and (5b), respectively,

\[
\begin{align*}
\arg\min_{G \in \mathcal{M}_\varepsilon(\mu, \sigma)} H_g(G) & \quad (6a) \\
\arg\max_{G \in \mathcal{M}_\varepsilon(\mu, \sigma)} H_g(G). & \quad (6b)
\end{align*}
\]

Note that the square-integrability of \( g \) in Assumption 2.1 is not restrictive. Indeed if \( \int_0^1 |g(u)|^2 \, du = +\infty, \) then problems (5a) and (5b) are both equal to \( +\infty. \) We provide a proof in Appendix B, Lemma B.1. We impose two additional conditions that are necessary in order to ensure that the considered optimization problems are relevant and well-posed.

**Assumption 2.3.** We assume that \( g(\cdot) \neq 1; \) otherwise \( H_g(G) = \mu \) for all \( G \in \mathcal{M}_\varepsilon(\mu, \sigma) \) and problems (5) and (6) are obsolete.

**Assumption 2.4.** We assume that \( \mathcal{M}_\varepsilon(\mu, \sigma) \) contains at least two elements, and hence infinitely many; otherwise problems (5) and (6) are either ill-posed or trivial.

In this regard, Assumption 2.4 is equivalent to assuming that \( \varepsilon > (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2. \) To see this, note that for any \( G \in \mathcal{M}_\varepsilon(\mu, \sigma) \)

\[
d_W(F, G)^2 = \int_0^1 (F^{-1}(u) - G^{-1}(u))^2 \, du \\
= \mu_F^2 + \sigma_F^2 + \mu^2 + \sigma^2 - 2 \text{cov}(F^{-1}(U), G^{-1}(U)) - 2\mu\mu_F \\
= (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma\sigma_F(1 - \text{corr}(F^{-1}(U), G^{-1}(U))) \\
\geq (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2.
\]
Hence, if \( \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 \), then \( M_\varepsilon(\mu, \sigma) = \emptyset \), and if \( \varepsilon = (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 \), then \( M_\varepsilon(\mu, \sigma) \) is a singleton, containing only one distribution with quantile function

\[
G^{-1}(u) = \mu + \sigma \left( \frac{F^{-1}(u) - \mu_F}{\sigma_F} \right), \quad 0 < u < 1.
\]

Moreover, in this special case, \( G \) coincides with the reference distribution \( F \) if and only if \( \mu = \mu_F \) and \( \sigma = \sigma_F \). Thus, from now on, we assume that \( \varepsilon > (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 \).

## 3 BOUNDS ON DISTORTION RISK MEASURES

For ease of exposition, we first study worst-case values (5b) and worst-case distribution functions (6b) for the class of concave distortion risk measures that are distortion risk measures with concave distortion function \( g \). The corresponding results for the case of general distortion risk measures are presented in Section 3.2 and those for best-case values (5a) and best-case distributions (6a) in Section 3.3.

### 3.1 Upper bounds on concave distortion risk measures

Concave distortion risk measures are a subset of the class of distortion risk measures that are coherent and include the widely used TVaR. Furthermore, concave distortion risk measures span the class of law-invariant, coherent, and comonotone additive risk measures (Kusuoka, 2001). For the special case of concave distortion risk measures, we obtain analytic solutions to problems (5b) and (6b), which are presented in the following theorem.

**Theorem 3.1** (Worst-case values and quantiles of concave distortions). Let Assumptions 2.1, 2.3, and 2.4 be fulfilled. Further, let \( g \) be a concave distortion function and denote by \( c_0 = \text{corr}(F^{-1}(U), \gamma(U)) \). Then, the following statements hold:

1. If \( (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma\sigma_F(1 - c_0) \), then the solution to Equation (6b) is unique and has a quantile function given by

\[
h_\lambda(u) = \mu + \sigma \left( \frac{\gamma(u) + \lambda F^{-1}(u) - a_\lambda}{b_\lambda} \right), \quad 0 < u < 1,
\]

where \( \lambda > 0 \) denotes the unique positive solution to \( d_W(F^{-1}, h_\lambda) = \sqrt{\varepsilon} \), which is explicitly given by

\[
\lambda = K \frac{C_{\gamma,F}^2 - V\sigma_F^2}{\sigma_F^2} \cdot \frac{C_{\gamma,F}}{\sigma_F^2},
\]

where \( V = \text{var}(\gamma(U)) \), \( C_{\gamma,F} = \text{cov}(F^{-1}(U), \gamma(U)) \), and \( K = \frac{\mu_F^2 + \sigma_F^2 + \mu^2 + \sigma^2 - 2\mu\mu_F - \varepsilon}{2} \geq 0 \) and where \( a_\lambda = \mathbb{E}(\gamma(U) + \lambda F^{-1}(U)) \) and \( b_\lambda = \text{std}(\gamma(U) + \lambda F^{-1}(U)) \). The corresponding worst-case value,

\[
\frac{C_{\gamma,F}^2 - V\sigma_F^2}{\sigma_F^2} \cdot \frac{C_{\gamma,F}}{\sigma_F^2},
\]

1 If \( c_0 = 1 \), then only case 2 applies.
that is, the solution to Equation (5b), is

$$H_g(h_\lambda) = \mu + \frac{\sigma}{b_\lambda}(V + \lambda C_{\gamma,F}) = \mu + \sigma \text{std}(\gamma(U)) \text{corr}(\gamma(U), \gamma(U) + \lambda F^{-1}(U)),$$

and $H_g(h_\lambda)$ is continuous in $\varepsilon$.

2. If $(\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma \sigma_F(1 - c_0) \leq \varepsilon$, then case 1 applies with $\lambda = 0$.

The worst-case value provided in Theorem 3.1 is sharp for all $(\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon \leq +\infty$ and attained by the worst-case quantile function $h_\lambda$. We observe that $h_\lambda$ is a weighted average of the reference quantile function $F^{-1}$ and the (non-decreasing) weight function $\gamma$ from the distortion risk measure. As the worst-case value $H_g$ is non-decreasing in the tolerance distance $\varepsilon$, we obtain that $\text{corr}(\gamma(U), \gamma(U) + \lambda F^{-1}(U))$ is non-decreasing in $\varepsilon$, which in turn implies that $\lambda$ is non-increasing in $\varepsilon$. Hence, we obtain that the influence of the reference distribution on the worst-case quantile function $h_\lambda$ diminishes with increasing tolerance distance $\varepsilon$. Furthermore, for $\varepsilon$ sufficiently large, that is, under case 2 of Theorem 3.1, $\lambda$ is zero and the worst-case quantile function is independent of the reference distribution. Note that when $\varepsilon = (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma \sigma_F(1 - c_0)$, the expression of $\lambda$ in Equation (7) is still valid and simplifies to $\lambda = 0$.

As a direct corollary, we obtain the worst-case values of TVaR. The proof is omitted as it follows by direct calculations from Theorem 3.1.

**Corollary 3.2** (Worst-case value of TVaR). Let Assumption 2.4 be fulfilled. Then, the worst-case value of the TVaR is given by

$$\sup_{G \in \mathcal{M}(\mu, \sigma)} TVaR_\alpha(G) = \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha} + 2\lambda(TVaR_\alpha(F) - \mu_F)} + \lambda^2 \sigma_F^2,$$

Here, if $\varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma \sigma_F(1 - c_0)$ then $\lambda$ is given by Equation (7) in Theorem 3.1, and if $\varepsilon \geq (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma \sigma_F(1 - c_0)$, then $\lambda = 0$.

Theorem 3.1 generalizes results that first appeared in Li (2018). This author derived sharp upper bounds on $H_g$ under the knowledge of the first two moments, that is, without considering an $\sqrt{\varepsilon}$-Wasserstein distance constraint or a reference distribution. Applying Theorem 3.1 with $\varepsilon = +\infty$, we obtain his result, which is stated here for completeness; see also Cornilly et al. (2018) and Zhu and Shao (2018).

**Corollary 3.3** (Worst-case values for $\varepsilon = +\infty$; Li (2018)). Let Assumptions 2.1 and 2.3 be fulfilled. Further, let $g$ be a concave distortion function; then

$$\sup_{G \in \mathcal{M}(\mu, \sigma)} H_g(G) = \mu + \sigma \text{std}(\gamma(U)) = \mu + \sqrt{\int_0^1 (\gamma(p) - 1)^2 \, dp}.$$

The bound is sharp and obtained by a unique maximizing distribution function with quantile function $h_0$ defined in Theorem 3.1.

**Example 3.4.** We illustrate the worst-case values of Theorem 3.1 for three concave distortion risk measures: the dual power distortion, the Wang transform, and TVaR defined in
Equations (2)–(4), respectively. The reference distribution $F$ is chosen to be standard normal, and we further set $\mu = \mu_F = 0$ and $\sigma = \sigma_F = 1$.

In the top panels of Figure 1, we observe that the upper bounds are indeed non-decreasing and continuous functions of $\epsilon$. The turquoise-colored vertical lines display $\epsilon^* = (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma\sigma_F(1 - c_0) = 2(1 - c_0)$. The parameter value $\epsilon^*$ indicates the transition from case 1 to case 2 in Theorem 3.1; indeed, for $\epsilon \geq \epsilon^*$ case 2 of Theorem 3.1 applies and the worst-case value is independent of $\epsilon$. For TVaR$_\alpha$ with $\alpha = 0.7$ (top right panel), the worst-case value is equal to $\mu + \sigma \sqrt{\alpha/1 - \alpha} = \sqrt{\alpha/1 - \alpha} = 1.53$, and we recover the well-known Cantelli upper bound.

The bottom panels of Figure 1 display the worst-case quantile functions for selected values of $\epsilon$. We observe that for $\epsilon = 0$, the worst-case quantile functions are equal to the reference quantile functions (blue solid line). When the Wasserstein distance $\epsilon$ increases, the influence of the reference distribution diminishes and, if $\epsilon > 2(1 - c_0)$, the worst-case quantile function (violet dash-dotted line) is independent of the reference distribution. This can be seen clearly for TVaR, where the violet dash-dotted quantile function corresponds to a two-point distribution.

### 3.2 Upper bounds on general distortion risk measures

To characterize the worst-case values and quantile functions of general distortion risk measures, we introduce the concept of isotonic projections; see for example, Németh (2003). For this, we denote the space of square-integrable, non-decreasing, and left-continuous functions on $(0,1)$ by

$$\mathcal{K} = \left\{ k : (0, 1) \to \mathbb{R} \left| \int_0^1 k(u)^2 du < +\infty, \ k \text{ non-decreasing and left-continuous} \right. \right\}. \quad (8)$$

**Definition 3.5.** For $\lambda \geq 0$, denote by $k^\uparrow_\lambda$ the isotonic projection of $\gamma + \lambda F^{-1}$ onto the space of square-integrable non-decreasing and left-continuous functions on $(0,1)$ — that is, the unique solution to

$$k^\uparrow_\lambda = \arg \min_{k \in \mathcal{K}} ||\gamma + \lambda F^{-1} - k||^2, \quad (9)$$

where $|| \cdot ||$ denotes the $L^2$ norm. Whenever $\lambda = 0$, we write $\gamma^\uparrow$ instead of $k^\uparrow_0$, as $k^\uparrow_0$ is indeed the isotonic projection of $\gamma$ alone.

The isotonic projection differs from the isotonic regression (Barlow et al., 1972), which is a metric projection onto the set of finite dimensional and component-wise increasing vectors. We refer to Appendix A for further details and properties of isotonic projections.

We further define

$$\lambda^\dagger = \inf \left\{ \lambda \geq 0 \left| k^\uparrow_\lambda \text{ is not constant} \right. \right\}$$

and set $\inf \emptyset = +\infty$. Finally, we define for $\lambda > \lambda^\dagger$ the quantile function

$$k^\uparrow_\lambda(u) = \mu + \sigma \left( \frac{k^\uparrow_\lambda(u) - a_\lambda}{b_\lambda} \right), \quad 0 < u < 1, \quad (10)$$
FIGURE 1 Worst-case values and quantile functions of three concave distortion risk measures: the dual power distortion (left), the Wang transform (middle), and the TVaR (right). The top panels display the worst-case values as a function of $\varepsilon$; the bottom panels display the worst-case quantile functions for selected values of $\varepsilon$. [Color figure can be viewed at wileyonlinelibrary.com]
with \( a_\lambda = \mathbb{E}(k_\lambda^1(U)) \) and \( b_\lambda = \text{std}(k_\lambda^1(U)) \). Note that \( h_\lambda^1 \) is well-defined whenever \( \lambda > \lambda^\dagger \), as the function \( k_\lambda^1 \) is nonconstant for all \( \lambda > \lambda^\dagger \); see also Proposition A.1 in Appendix A.

Before stating the worst-case value of general distortion risk measures, we need a further assumption.

**Assumption 3.6.** We assume that \( \lambda^\dagger = 0 \).

Note that Assumption 3.6 is a requirement on the weight function and the reference distribution. Assumption 3.6 is satisfied, for instance, when \( \gamma \) is bounded and the reference distribution has a quantile function that is unbounded to the right, that is, \( F^{-1}(u) \to +\infty \) for \( u \to 1 \).

**Theorem 3.7** (Worst-case values and quantiles of general distortions). Let Assumptions 2.1, 2.3, 2.4, and 3.6 be fulfilled. Further, let \( g \) be a distortion function and denote \( c_0 = \lim_{\lambda \searrow 0} \text{corr}(F^{-1}(U), k_\lambda^1(U)) \). Then, the following statements hold:

1. If \( (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma\sigma_F(1 - c_0) \), then a solution to Equation (6b) is unique and is given by the quantile function \( h_\lambda^1 \) defined in Equation (10), where \( \lambda > 0 \) is the unique positive solution to \( d_W(F^{-1}, h_\lambda^1) = \sqrt{\varepsilon} \). Furthermore, the worst-case value, that is, the solution to Equation (5b), is

\[
H_g(h_\lambda^1) = \mu + \sigma \text{std}(\gamma(U))\text{corr}(\gamma(U), k_\lambda^1(U))
\]

and \( H_g(h_\lambda^1) \) is continuous in \( \varepsilon \).

2. Let \( (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma\sigma_F(1 - c_0) \leq \varepsilon \). If \( \gamma^\dagger \) is not constant, then case (i) applies with \( \lambda = 0 \). If \( \gamma^\dagger \) is constant, then the solution to Equation (5b) is equal to \( \mu \) but cannot be attained in general.

**Remark 3.8.** If \( \gamma \) is non-increasing, equivalently the distortion function \( g \) is convex, then its isotonic projection \( \gamma^\dagger \) is constant and thus equal to 1. Hence, for \( \varepsilon \) sufficiently large, case 2 of Theorem 3.7 implies that the upper bound is equal to \( \mu \). However, the well-known Hoeffding bounds for the expectation of a product of two random variables with given marginal distributions imply that for any admissible quantile function \( G^{-1} \), it holds that \( \int_0^1 \gamma(u)G^{-1}(u)du < \int_0^1 G^{-1}(u)du = \mu \) (note that \( \int_0^1 \gamma(u)du = 1 \) and \( \gamma \) is assumed to be nonconstant). Hence, the upper bound \( \mu \) is not attained. However, if, for example, there exists \( 0 < p < \frac{1}{2} \) such that \( \int_0^p \gamma(u)du = \int_0^1 \gamma(u)du \) and \( \gamma^\dagger = 1 \), then the upper bound \( \mu \) is attained. To see this, consider the quantile function \( G^{-1}(u) = \mu + \sigma F^{-1}(u) \) in which \( F^{-1}(u) \) takes value \( \frac{-1}{\sqrt{2p}} \) on \((0, p)\), value \( \frac{1}{\sqrt{2p}} \) on \((1-p, p)\), and value zero otherwise. It verifies that the mean of \( G^{-1} \) is \( \mu \) whereas its variance is equal to \( \sigma^2 \). Moreover, \( \int_0^1 \gamma(u)G^{-1}(u)du = \mu \), thus \( G^{-1} \) attains the worst-case bound. An example of \( \gamma \) satisfying these conditions is if \( \gamma(u) = 1 - 2u \) on \((0, 1/2)\) and \( \gamma(u) = 2u - 1 \) on \([1/2, 1)\).

If \( \gamma \) is non-decreasing, equivalently \( g \) is concave, then \( \gamma + \lambda F^{-1} \) is non-decreasing and thus equal to its isotonic projection. Hence, Theorem 3.7 reduces to Theorem 3.1 in the case of concave distortion risk measures. Theorem 3.7 generalizes results by Zhu and Shao (2018), who analyze problem (5b) in the special case in which the Wasserstein constraint is redundant, which we state in the subsequent corollary.
Corollary 3.9 (Worst-case values for $\varepsilon = +\infty$). Let Assumptions 2.1, 2.3, and 3.6 be fulfilled. Further, let $g$ be a distortion function; then, the following statements hold:

1. If $\gamma^\uparrow$ is not constant, then
   \[
   \sup_{G \in \mathcal{M}(\mu, \sigma)} H_g(G) = \mu + \sigma \sqrt{\int_0^1 (\gamma^\uparrow(p) - 1)^2 \, dp}.
   \]
   (11)
   The bound is sharp and attained by a distribution with quantile function $h_0^\uparrow$ defined in Equation (10).

2. If $\gamma^\uparrow$ is constant, then
   \[
   \sup_{G \in \mathcal{M}(\mu, \sigma)} H_g(G) = \mu
   \]
   and the bound cannot be attained in general.

3.3 Lower bounds on general distortion risk measures

In a similar way as in Section 3.2, we first introduce the notation and assumptions needed to derive the best-case values and quantile functions of general distortion risk measures.

Definition 3.10. For $\lambda \geq 0$, denote by $k^\downarrow_\lambda$ the isotonic projection of $\gamma - \lambda F^{-1}$ onto the set of square-integrable non-increasing functions and left-continuous—that is, the unique solution to
   \[
   k^\downarrow_\lambda = \arg\min_{k \in \mathcal{K}} ||\gamma - \lambda F^{-1} - k||^2.
   \]
   Whenever $\lambda = 0$, we write $\gamma^\downarrow$ instead of $k^\downarrow_0$, as $k^\downarrow_0$ is indeed the isotonic projection of $\gamma$ onto the non-increasing functions.

Further, we define
   \[
   \lambda^\downarrow = \inf \left\{ \lambda \geq 0 \left| k^\downarrow_\lambda \text{ is not constant} \right. \right\}
   \]
   and for all $\lambda > \lambda^\downarrow$ the quantile function
   \[
   h^\downarrow_\lambda(u) = \mu + \sigma \left( \frac{a_\lambda - k^\downarrow_\lambda(u)}{b_\lambda} \right), \quad 0 < u < 1,
   \]
   (13)
   with $a_\lambda = \mathbb{E}(k^\downarrow_\lambda(U))$ and $b_\lambda = \text{std}(k^\downarrow_\lambda(U))$. Note that $h^\downarrow_\lambda$ is well-defined whenever $\lambda > \lambda^\downarrow$, as the function $k^\downarrow_\lambda$ is nonconstant for all $\lambda > \lambda^\downarrow$; see also Proposition A.1 in Appendix A.

Assumption 3.11. We assume that $\lambda^\downarrow = 0$.

The next theorem states the best-case values and quantile functions of general distortion risk measures.
Theorem 3.12 (Best-case values and quantiles of general distortions). Let Assumptions 2.1, 2.3, 2.4, and 3.11 be fulfilled. Further, let \( g \) be a distortion function and denote \( c_0 = \lim_{\lambda \searrow 0} \text{corr}(F^{-1}(U), -k_1^\lambda(U)) \). Then, the following statements hold:

1. If \((\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma\sigma_F(1 - c_0)\), then a solution to Equation (6a) is unique and is given by the quantile function \( h_\lambda^1 \) defined in Equation (13), where \( \lambda > 0 \) is the unique positive solution to \( d_W(F^{-1}, h_\lambda^1) = \sqrt{\varepsilon} \). The corresponding best-case value, that is, the solution to Equation (5a), is

\[
H_g(h_\lambda^1) = \mu - \sigma \text{std}(\gamma(U)) \text{corr}(\gamma(U), k_\lambda^1(U)),
\]

and \( H_g \) is continuous in \( \varepsilon \).

2. Let \((\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma\sigma_F(1 - c_0) \leq \varepsilon \). If \( \gamma \) is not constant, then case (i) applies with \( \lambda = 0 \). If \( \gamma \) is constant, then the infimum is equal to \( \mu \) but cannot be attained.

Theorem 3.12 implies that for \( \varepsilon \) sufficiently large, that is, in case 2, the lower bound of any concave distortion risk measure is equal to \( \mu \) and not attained. The next corollary collects the results for \( \varepsilon = +\infty \).

Corollary 3.13 (Best-case values for \( \varepsilon = +\infty \)). Let Assumptions 2.1, 2.3, and 3.11 be fulfilled. Further, let \( g \) be a distortion function; then, the following statements hold:

1. If \( \gamma \) is not constant, then

\[
\inf_{G \in \mathcal{M}(\mu, \sigma)} H_g(G) = \mu - \sigma \text{std}(\gamma(U)) \text{corr}(\gamma(U), \gamma(U)).
\]

The bound is sharp and is attained by a distribution with quantile function \( h_0^1 \) defined in Equation (13).

2. If \( \gamma \) is constant, then

\[
\sup_{G \in \mathcal{M}(\mu, \sigma)} H_g(G) = \mu
\]

and the bound cannot be attained.

Remark 3.14. Distortion risk measures have an alternative representation that makes it possible to write minimization problems in terms of maximization problems. Specifically,

\[
\inf_{G_X \in \mathcal{M}(\mu, \sigma)} H_g(G_X) = - \sup_{G_{-X} \in \mathcal{M}(-\mu, \sigma)} H_{\tilde{g}}(G_{-X}) \tag{14}
\]

where \( \tilde{g} \) is the dual distortion of \( g \), given by \( \tilde{g}(x) = 1 - g(1 - x) \), and \( G_X, G_{-X} \) denote the distribution functions of the random variables \( X \) and \( -X \), respectively. Equation (14) follows from the fact that \( H_g(G_X) = -H_{\tilde{g}}(G_{-X}) \); see for example, Dhaene et al. (2012). This observation makes it possible to obtain Corollary 3.13, the statements on lower bounds, from Corollary 3.9, the corresponding statements on upper bounds, in a more direct manner. Such reasoning, however, cannot be extended when dealing with the Wasserstein distance constraints considered in this paper.
Remark 3.15. Theorems 3.7 and 3.12 can be generalized to signed Choquet integrals. A signed Choquet integral is, for a distribution function $G \in M^2$, defined by $H_g(G) = -\int_{-\infty}^{0} g(1 - g(1 - G(x)))dx + \int_{0}^{+\infty} g(1 - G(x))dx$, where $g : [0,1] \rightarrow \mathbb{R}$ is of bounded variation with $g(0) = 0$. Signed Choquet integrals generalize the class of distortion risk measures in that the distortion function $g$ can be decreasing. For absolutely continuous $g$, the signed Choquet integral admits a representation as in Equation (1), see Wang et al. (2020), thus, Theorems 3.7 and 3.12 can be extended to signed Choquet integrals with absolutely continuous distortion function.

Example 3.16 (Continued). Figure 2 illustrates the best- and worst-case values (Theorems 3.7 and 3.12) of three concave distortion risk measures: the dual power distortion, the Wang transform, and TVaR, with a standard normal reference distribution, $\mu_F = \mu = 0$, and $\sigma_F = \sigma = 1$. In the upper panels of Figure 2, the solid blue line corresponds to the upper bound and the dashed blue line corresponds to the lower bound. The black line depicts the reference risk measure $H_g(F)$ and the vertical turquoise lines display the critical $\varepsilon$ value for the transition between case 1 and case 2 in Theorem 3.7 (upper bound, solid line) and Theorem 3.12 (lower bound, dashed line).

As all three risk measures are concave, their lower bounds—for $\varepsilon$ sufficiently large (case 2 in Theorem 3.12)—are all equal to $\mu = 0$ but not attained. That the lower bounds are not attained can be seen in the lower panels, where the corresponding best-case quantile functions are displayed. The best-case quantile functions become flatter for larger $\varepsilon$, and for $\varepsilon$ sufficiently large, the quantile functions are no longer defined.

4 | EXTENSIONS

In this section, we consider two extensions. First, we derive explicit bounds on RVaR and (via a direct proof) the bounds on VaR. Second, we extend the results to the uncertainty sets solely described by the Wasserstein distance.

4.1 | Bounds on Range Value-at-Risk, Value-at-Risk, and Tail Value-at-risk

In this section, we provide bounds on the risk measures RVaR, TVaR, and VaR. Specifically, we first calculate the best- and worst-case values of RVaR and then derive the TVaR bounds as limiting cases. The bounds on VaR are derived via a direct proof as VaR does not satisfy Assumption 2.1. The RVaR at levels $0 < \alpha < \beta \leq 1$ is defined by (Cont et al., 2010)

$$g(x) = \min \left\{ \max \left\{ \frac{x+\beta-1}{\beta-\alpha}, 0 \right\}, 1 \right\} \text{ and } \gamma(u) = \frac{1}{\beta - \alpha} 1_{[\alpha, \beta]}(u).$$

By letting $\beta \nearrow 1$, we obtain TVaR; that is, for any $G \in M^2$, it holds that

$$\text{TVaR}_\alpha(G) = \lim_{\beta \nearrow 1} \text{RVaR}_{\alpha,\beta}(G).$$

Next, recall that $\text{VaR}_\alpha(G) = G^{-1}(\alpha)$ and denote $\text{VaR}^+(\alpha) = G^{-1,+}(\alpha)$. Both VaR and $\text{VaR}^+$ are distortion risk measures (VaR with $g(x) = 1_{[1-\alpha, 1]}(x)$ and $\text{VaR}^+$ with $g(x) = 1_{[1-\alpha, 1]}(x)$, see Dhaene et al. (2012)); however, they do not belong to the class of distortion risk measures considered in
BERNARD et al.

**FIGURE 2** Best- and worst-case values and quantile functions of three concave risk measures; the dual power distortion (left), the Wang transform (middle), and TVaR (right). The top panels display the best- and worst-case values as a function of $\varepsilon$; the bottom panels display the best-case quantile functions for selected values of $\varepsilon$. The vertical turquoise lines in the upper panels correspond to the transition of case 1 to case 2, in the upper bound (Theorem 3.7, solid line) and lower bound (Theorem 3.12, dashed line). [Color figure can be viewed at wileyonlinelibrary.com]
this paper as they do not satisfy Assumption 2.1. Furthermore, it holds that for any $G \in \mathcal{M}^2$

\[
\text{VaR}_\alpha(G) = \lim_{\alpha' \searrow \alpha} \text{RVaR}_{\alpha',\alpha}(G) \quad \text{and} \quad \text{VaR}_\alpha^+(G) = \lim_{\beta \nearrow \alpha} \text{RVaR}_{\alpha,\beta}(G).
\]

The worst-case values of these risk measures under constraints on the first two moments but without a Wasserstein constraint have been extensively studied; indicatively see Kaas and Goovaerts (1986), Lo (1987), Grundy (1991), Hürlimann (2005), and De Schepper and Heijnen (2010) for the VaR, Li et al. (2018) for the RVaR, and Natarajan et al. (2010) for the TVaR. We collect the worst- and best-case values of these risk measures without a Wasserstein constraint in the subsequent corollaries. The bounds on TVaR and RVaR also follow from Corollaries 3.9 and 3.13.

**Corollary 4.1** (Worst-case values for $\varepsilon = +\infty$). For $0 < \alpha < \beta \leq 1$, the worst-case values of $\text{VaR}_\alpha$, $\text{VaR}_\alpha^+$, $\text{RVaR}_{\alpha,\beta}$, and $\text{TVaR}_{\alpha,\beta}$ coincide; that is,

\[
\sup_{G \in \mathcal{M}(\mu,\sigma)} \text{VaR}_\alpha(G) = \sup_{G \in \mathcal{M}(\mu,\sigma)} \text{VaR}_\alpha^+(G) = \sup_{G \in \mathcal{M}(\mu,\sigma)} \text{RVaR}_{\alpha,\beta}(G) = \sup_{G \in \mathcal{M}(\mu,\sigma)} \text{TVaR}_\alpha(G) = \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}}.
\]

While the worst-case value of $\text{VaR}_\alpha$ cannot be attained, the worst-case distribution of $\text{VaR}_\alpha^+$, $\text{RVaR}_{\alpha,\beta}$, and $\text{TVaR}_{\alpha}$ is a two-point distribution with quantile function

\[
h_0^+(u) = \begin{cases} 
\mu - \sigma \sqrt{\frac{1-\alpha}{\alpha}} & 0 < u \leq \alpha, \\
\mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}} & \alpha < u < 1.
\end{cases}
\]

**Corollary 4.2** (Best-case values for $\varepsilon = +\infty$). For $0 < \alpha < \beta \leq 1$, the best-case values of $\text{VaR}_\alpha$, $\text{VaR}_\alpha^+$, $\text{RVaR}_{\alpha,\beta}$, and $\text{TVaR}_{\alpha,\beta}$ are

\[
\inf_{G \in \mathcal{M}(\mu,\sigma)} \text{VaR}_\alpha(G) = \inf_{G \in \mathcal{M}(\mu,\sigma)} \text{VaR}_\alpha^+(G) = \mu - \sigma \sqrt{\frac{1-\alpha}{\alpha}},
\]

\[
\inf_{G \in \mathcal{M}(\mu,\sigma)} \text{RVaR}_{\alpha,\beta}(G) = \mu - \sigma \sqrt{\frac{1-\beta}{\beta}}, \quad \text{and} \quad \inf_{G \in \mathcal{M}(\mu,\sigma)} \text{TVaR}_\alpha(G) = \mu.
\]

While the best-case values of $\text{VaR}_\alpha^+$ and $\text{TVaR}_{\alpha,\beta}$ cannot be attained, the best-case distribution corresponding to the best-case $\text{VaR}_\alpha$ has a quantile function given by Equation (15) whereas for the best-case $\text{RVaR}_{\alpha,\beta}$, it is a two-point distribution with quantile function

\[
h_0^+(u) = \begin{cases} 
\mu - \sigma \sqrt{\frac{1-\beta}{\beta}} & 0 < u \leq \beta, \\
\mu + \sigma \sqrt{\frac{\beta}{1-\beta}} & \beta < u < 1.
\end{cases}
\]
Next, we study the lower and upper bound on RVaR when, in addition to the moment constraints, the distributions in the uncertainty set lie within an $\sqrt{\varepsilon}$-Wasserstein ball of the reference distribution $F$.

**Proposition 4.3** (Worst-case quantiles of RVaR). Under the assumptions of Theorem 3.7 case 1, a solution to Equation (6b) with $\text{RVaR}_{\alpha,\beta}$ is unique and is given by $h^\uparrow_\lambda$ defined in Equation (10), where $k^\uparrow_\lambda$ is

$$k^\uparrow_\lambda(u) = \begin{cases} 
\lambda F^{-1}(u) & 0 < u \leq \alpha, \\
\frac{1}{\beta - \alpha} + \lambda F^{-1}(u) & \alpha < u \leq w_0, \\
c & w_0 < u \leq w_1, \\
\lambda F^{-1}(u) & w_1 < u < 1, 
\end{cases}$$

and $w_0$, $w_1$, and $c$, with $\alpha \leq w_0 \leq \beta \leq w_1$, $c < +\infty$, satisfy

$$\lambda F^{-1}(w_0) = \begin{cases} 
c - \frac{1}{\beta - \alpha}, & \text{if } \frac{1}{\beta - \alpha} \leq c - \lambda F^{-1}(\alpha), \\
\lambda F^{-1}(\alpha), & \text{otherwise},
\end{cases}$$

$$\lambda F^{-1}(w_1) = \begin{cases} 
\lambda F^{-1}(1) & \text{if } c \geq \lambda F^{-1}(1), \\
c & \text{otherwise},
\end{cases}$$

$$c = \frac{1}{w_1 - w_0} \frac{\beta - w_0}{\beta - \alpha} + \frac{\lambda}{w_1 - w_0} \int_{w_0}^{w_1} F^{-1}(u) \, du.$$

**Proposition 4.4** (Best-case quantiles of RVaR). Under the assumptions of Theorem 3.12 case 1, a solution to Equation (6a) with $\text{RVaR}_{\alpha,\beta}$ is unique and is given by $h^\downarrow_\lambda$ defined in Theorem 3.12, where $k^\downarrow_\lambda$ is

$$k^\downarrow_\lambda(u) = \begin{cases} 
-\lambda F^{-1}(u) & 0 < u \leq z_0, \\
-c & z_0 < u \leq z_1, \\
-\lambda F^{-1}(u) + \frac{1}{\beta - \alpha}, & z_1 < u \leq \beta, \\
-\lambda F^{-1}(u) & \beta < u < 1,
\end{cases}$$

and $z_0$, $z_1$, and $c$, with $z_0 \leq \alpha \leq z_1 \leq \beta$, $c > -\infty$, solve

$$\lambda F^{-1}(z_1) = \begin{cases} 
\lambda F^{-1}(z_0) + \frac{1}{\beta - \alpha} & \text{if } \frac{1}{\beta - \alpha} < \lambda F^{-1}(\beta) - c, \\
\lambda F^{-1}(\beta) & \text{otherwise},
\end{cases}$$
TABLE 1 Best- and worst-case quantile functions of VaR_\alpha, VaR^+_\alpha, RVaR_{\alpha, \beta}, and TVaR_\alpha for case 1 of Theorems 3.7 and 3.12, respectively; that is, for \varepsilon sufficiently small.

| Risk measure   | Best-case quantile | Worst-case quantile |
|----------------|--------------------|---------------------|
| RVaR_{\alpha, \beta} | Proposition 4.4; | Proposition 4.3; |
| VaR_\alpha      | Proposition 4.4 with z_1 = \beta; | Not attained; |
| VaR^+_\alpha   | Not attained;      | Proposition 4.3 with w_0 = \alpha; |
| TVaR_\alpha     | Proposition 4.4 with \beta = 1; | Proposition 4.3 with \beta = 1 and w_0 = w_1 = 1. |

\[
\lambda F^{-1}(z_0) = \begin{cases} 
\lambda F^{-1}(0) & \text{if } c \leq \lambda F^{-1}(0), \\
 c & \text{otherwise}, 
\end{cases}
\]

\[
c = -\frac{z_1 - \alpha}{z_1 - z_0} \frac{1}{\beta - \alpha} + \frac{\lambda}{z_1 - z_0} \int_{z_0}^{z_1} F^{-1}(u) \, du.
\]

We verified with numerical experiments that the closed-form expressions in Equations (16) and (17) match well with numerically obtained isotonic and antitonic projections.\footnote{All numerically obtained isotonic projections are constructed using LSQISOTONIC, a built-in function in Matlab.}

The subsequent corollary and proposition provide the best- and worst-case quantile functions of TVaR and VaR, respectively. The results are also summarized in Table 1.

Corollary 4.5 (Best- and worst-case quantiles of TVaR). Under the assumptions of case 1 in Theorems 3.7 and 3.12, respectively, it holds that the worst-case quantile function of TVaR_\alpha is given by Proposition 4.3 with \beta = 1 and w_0 = w_1 = 1. The corresponding worst-case value of TVaR_\alpha is given in Corollary 3.2. The best-case quantile function of TVaR_\alpha is given by Proposition 4.4 with \beta = 1.

As VaR and VaR^+ do not satisfy Assumption 1, we provide a direct proof of the next proposition.

Proposition 4.6 (Best- and worst-case quantile of VaR). Let \(c_0 = \text{corr}(F^{-1}(U), 1_{U \in (\alpha, 1)})\) and \((\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F(1 - c_0)\). It holds that the worst-case quantile function of VaR^+_\alpha is given by Proposition 4.3 with w_0 = \alpha. The worst-case value of VaR_\alpha is equal to that of VaR^+_\alpha but not attained. The best-case quantile function of VaR_\alpha is given by Proposition 4.4 with \(z_1 = \beta\). The best-case value of VaR^+_\alpha is equal to that of VaR_\alpha but is not attained.

Example 4.7. Figures 3 and 4 illustrate the isotonic projections for the worst- and best-case quantile function of RVaR_{\alpha, \beta} with \alpha = 0.6 and different values of \beta \in \{0.61, 0.85, 0.99\}. Specifically, Figure 3 displays \(\gamma(u) + \lambda F^{-1}(u), u \in (0, 1)\) and its isotonic projection \(k_\lambda^I\) onto the set of non-decreasing functions (derived in Proposition 4.3). Note that we chose \varepsilon sufficiently small such that the worst-case quantile functions of RVaR_{\alpha, \beta} are not two-point distributions, and thus case 1 of Theorem 3.7 applies. Figure 4 displays the corresponding graphs for the best-case quantile functions.

The left plot of Figure 5 displays the lower and upper bounds on VaR_\alpha, RVaR_{\alpha, \beta}, and TVaR_\alpha as a function of the Wasserstein distance \varepsilon. When \varepsilon is sufficiently large (case 2 in Theorems 3.7 and 3.12), the Wasserstein distance no longer affects the bounds; these bounds coincide with...
For the risk measure RVaR$_{\alpha,\beta}$, we plot $\gamma(u) + \lambda F^{-1}(u)$, $u \in (0, 1)$ and its isotonic projection $k_{\lambda}$ onto the set of non-decreasing functions (derived in Proposition 4.3). The reference distribution $F$ is standard normal with $\mu = \mu_F = 0$ and $\sigma = \sigma_F = 1$, and we set $\epsilon = 0.2$. [Color figure can be viewed at wileyonlinelibrary.com]
For the risk measure $\text{RVaR}_{\alpha, \beta}$, we plot $\gamma(u) - \lambda F^{-1}(u)$, $u \in (0, 1)$ and its isotonic projection $k_1^+$ onto the set of non-increasing functions (derived in Proposition 4.4). The reference distribution $F$ is standard normal with $\mu = \mu_F = 0$ and $\sigma = \sigma_F = 1$, and we set $\varepsilon = 0.2$. [Color figure can be viewed at wileyonlinelibrary.com]
FIGURE 5  Left: Best- and worst-case values of VaR, RVaR, and TVaR as a function of the tolerance level $\epsilon$. VaR and TVaR are at level $\alpha = 0.6$ and RVaR is at levels $\alpha = 0.6$ and $\beta = 0.85$. Solid lines represent the worst-case values, dashed lines represent the best-case values. Right: Length of the normalized bounds of VaR, RVaR, and TVaR. [Color figure can be viewed at wileyonlinelibrary.com]
the well-known Cantelli bounds. The normalized lengths of the bounds on VaR$_\alpha$, RVaR$_{\alpha,\beta}$, and TVaR$_\alpha$ as a function of $\varepsilon$ are displayed in the right plot of Figure 5. The normalized length of the bounds is the differences between Equations (5b) and (5a) divided by the risk measure evaluated at the reference distribution and have been introduced to assess model risk by Barrieu and Scandolo (2015). The length of the bounds on RVaR$_{\alpha,\beta}$, for example, is $(\sup_{G \in \mathcal{M}_{\varepsilon}(\mu, \sigma)} \text{RVaR}_{\alpha,\beta}(G) - \inf_{G \in \mathcal{M}_{\varepsilon}(\mu, \sigma)} \text{RVaR}_{\alpha,\beta}(G))/\text{RVaR}_{\alpha,\beta}(F)$.

We observe that the normalized length of the bounds on all three risk measures increases with respect to $\varepsilon$. Moreover, the normalized length of the VaR bounds is significantly larger than in the case of TVaR, a fact that is well-known for the case when $\varepsilon = +\infty$ (Embrechts et al., 2015).

### 4.2 Wasserstein uncertainty set

In this section, we consider worst-case concave distortion risk measures under complete moment uncertainty. Specifically, we consider an uncertainty set that contains all distribution functions that lie within an $\sqrt{\varepsilon}$-Wasserstein ball around the reference distribution, that is, without a first and second moment constraint. Thus, the optimization problem becomes

$$\text{arg max}_{G \in \mathcal{M}_{\varepsilon}} H_g(G),$$

where $\mathcal{M}_{\varepsilon} = \{ G \in \mathcal{M}^2 \mid d_W(F, G) \leq \sqrt{\varepsilon} \}$ is the set of all distribution functions that lie within a $\sqrt{\varepsilon}$-Wasserstein ball around the reference distribution $F$.

**Theorem 4.8** (Wasserstein uncertainty). Let Assumptions 2.1 and 2.3 be fulfilled. Further, let $g$ be a concave distortion function; then, the solution to problem (18) exists, is unique, and has quantile function given by

$$h_\lambda(u) = \mu^* + \sigma^* \left( \frac{\gamma(u) + \lambda F^{-1}(u) - a_\lambda}{b_\lambda} \right), \quad 0 < u < 1,$$

with

$$\left( \mu^*, \sigma^* \right) = \left( \mu_F + \sqrt{\frac{\varepsilon}{1 + V}}, \quad \sqrt{\frac{\sigma_F^2 + \frac{2C_{\gamma,F} \sqrt{\varepsilon}}{\sqrt{1 + V} + \varepsilon V}}{1 + V}} \right),$$

where $a_\lambda, b_\lambda, V, \text{ and } C_{\gamma,F}$ are given in Theorem 3.1 and $\lambda > 0$ denotes the unique positive solution to $d_W(F^{-1}, h_\lambda) = \sqrt{\varepsilon}$, which is explicitly given in Theorem 3.1. The corresponding worst-case value is

$$H_g(h_\lambda) = \mu^* + \sigma^* \text{ std}(\gamma(U)) \text{corr}(\gamma(U), \lambda F^{-1}(U)).$$

Note that if $\varepsilon > 0$, then the mean and standard deviation of the worst-case distribution fulfill $\mu^* > \mu_F$ and $\sigma^* > \sigma_F$ (recall that $C_{\gamma,F} \geq 0$). Thus, the worst-case distribution when the uncertainty set is only characterized by a Wasserstein ball has a larger mean and standard deviation than the benchmark. Thus, the uncertainty set $\mathcal{M}_{\varepsilon}(\mu, \sigma)$ with fixed $\mu$ and $\sigma$ results in a different worst-case distribution compared to the uncertainty set characterized solely by the Wasserstein distance.
The tolerance distance \( \varepsilon \), that is, the degree of the uncertainty, should be adequately chosen. Pesenti and Jaimungal (2023) study optimal portfolio choice and assume that the investor is prepared to accept terminal wealth distributions that stay within an \( \varepsilon \)-Wasserstein ball around some benchmark distribution. In this case, the Wasserstein tolerance distance \( \varepsilon \) reflects the investor’s tolerance of deviating from the benchmark portfolio strategy and its value is thus driven by the investor’s risk preferences. In most applications, however, the uncertainty set is constructed to contain all “plausible” distributions, that is, the uncertainty set should be large enough to contain with high probability the true data-generating distribution and at the same time small enough to exclude pathological distributions, which would incentivize overly conservative decisions (Esfahani & Kuhn, 2018). In these settings, the tolerance distance should be driven by data considerations rather than being exogenously specified. In this regard, in various (data) contexts and under various assumptions, Wozabal (2014) chooses \( \varepsilon \) via cross-validation whereas Blanchet et al. (2022) choose \( \varepsilon \) such that the true distribution function lies with a given probability within a Wasserstein ball around the empirical reference distribution. Additionally, including expert opinion in constructing the uncertainty set may provide valuable information, particularly in the case of limited available data (Clemen & Winkler, 1999). In the applications, hereafter, we do not pursue the estimation of \( \varepsilon \) in great length, but rather provide some guidelines for choosing it.

5.1 Portfolio optimization

We consider a portfolio optimization problem in which an investor aims to construct a robust portfolio \( \mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{P} \) among returns \( \mathbf{R} = (R_1, \ldots, R_n) \), \( n \in \mathbb{N} \), in which \( \mathcal{P} \) is a suitable polyhedral set. We denote the multivariate distribution function of \( \mathbf{R} \) by \( G_\mathbf{R} \) and assume that it is subject to uncertainty in that only its mean vector \( \mu \) and covariance matrix \( \Sigma \) are known. For each portfolio \( \mathbf{x} \), we denote by \( G_\mathbf{x} \) the (unknown) distribution function of the aggregate portfolio loss \( -\mathbf{x}^\top \mathbf{R} \) having mean \( -\mu_\mathbf{x} = -\mathbf{x}^\top \mu \) and variance \( \sigma_\mathbf{x}^2 = \mathbf{x}^\top \Sigma \mathbf{x} \). There is a benchmark model under which the investor assumes that the aggregate portfolio loss has distribution function \( F_\mathbf{x} \) belonging to a location-scale family, that is, there is a distribution function \( F \) such that

\[
F_\mathbf{x}^{-1}(u) = -\mu_\mathbf{x} + \sigma_\mathbf{x} F^{-1}(u), \quad 0 < u < 1.
\]

Then, for fixed \( \mathbf{x} \in \mathcal{P} \), the Wasserstein distance between \( F_\mathbf{x} \) and \( G_\mathbf{x} \) is given by

\[
d_W(F_\mathbf{x}, G_\mathbf{x})^2 = 2\sigma_\mathbf{x}^2(1 - \text{corr}(F_\mathbf{x}^{-1}(U), G_\mathbf{x}^{-1}(U))).
\] (19)

The investor now aims to find the portfolio \( \mathbf{x} \) that has the smallest risk—measured via a concave distortion risk measure—among all possible portfolios whose aggregate loss lie within a Wasserstein distance of the benchmark portfolio. Specifically, we consider the optimization problem

\[
\min_{\mathbf{x} \in \mathcal{P}} \max_{G_\mathbf{R} \in \mathcal{M}(\mu, \Sigma)} H_\mathbf{g}(G_\mathbf{x}), \quad d_W(F_\mathbf{x}, G_\mathbf{x}) \leq \sqrt{\tau_\mathbf{x}}.
\] (20)

\[ This assumption is fulfilled, for instance, when an elliptical multivariate distribution function is taken as reference model for \( \mathbf{R} \).
where $\mathcal{M}(\mu, \Sigma)$ is the set of all $n$-dimensional distribution functions with mean vector $\mu$ and covariance matrix $\Sigma$, and where we allow the tolerance distance to depend on $x$, that is, $\varepsilon_x \geq 0$, $x \in P$.

**Proposition 5.1.** Let Assumptions 2.1, 2.3, and 2.4 be fulfilled. Further, let $g$ be a concave distortion function. Then, problem (20) is equivalent to

$$
\min_{x \in P} \left\{ -\mu_x + \left( \sigma_x - \frac{\varepsilon_x}{2\sigma_x} \right) \sqrt{Vc_0} + \sqrt{\varepsilon_x - \frac{\varepsilon_x^2}{4\sigma_x^2}} \sqrt{V(1 - c_0^2)} \right\}. 
$$

(21)

where $\varepsilon_x = \min\{\varepsilon_x, 2\sigma_x^2(1 - c_0)\}$, $V = \text{var}(\gamma(U))$, and $c_0 = \text{corr}(F^{-1}(U), \gamma(U))$.

Note that problem formulation (20) is general in that the tolerances $\varepsilon_x$ may depend on the portfolio $x \in P$ at hand. As a consequence, it is not always possible to obtain the minimum in Equation (21) in explicit form, and numerical procedures may be required.

From Equation (19), however, we observe that the Wasserstein distance scales linearly with the standard deviation of the portfolio loss. Hence, it appears natural to take the tolerance $\varepsilon_x$ of the form $^4 \varepsilon_x = 2\sigma_x^2 A$ for $0 \leq A \leq 1$. The next result considers this case. Its proof follows from Proposition 5.1 via direct calculation and is omitted.

**Corollary 5.2.** Let Assumptions 2.1, 2.3, and 2.4 be fulfilled. Further, let $g$ be a concave distortion function and assume that $\varepsilon_x = 2\sigma_x^2 A$ for $A \in [0, 1]$. Let $V = \text{var}(\gamma(U))$ and $c_0 = \text{corr}(F^{-1}(U), \gamma(U))$. Then, the following holds:

1. If $A = 0$, then problem (20) is equivalent to

$$
\min_{x \in P} \left\{ -\mu_x + \sigma_x \sqrt{Vc_0} \right\}.
$$

2. If $0 < A < 1 - c_0$, then problem (20) is equivalent to

$$
\min_{x \in P} \left\{ -\mu_x + \sigma_x \sqrt{V \left( c_0(1 - A) + \sqrt{A(2 - A)} \sqrt{1 - c_0^2} \right)} \right\}.
$$

3. If $1 - c_0 \leq A \leq 1$, then problem (20) is equivalent to

$$
\min_{x \in P} \left\{ -\mu_x + \sigma_x \sqrt{V} \right\}.
$$

$^4$If historical asset returns are available, the value of $A$ can be determined as follows: using the historical returns we estimate for a given portfolio $x = (x_1, \ldots, x_n) \in P$, the empirical distribution $\hat{G}_x$ of the portfolio loss, its standard deviation $\hat{\sigma}_x$, and the Wasserstein distance $\hat{d}_x$ between $\hat{G}_x$ and the (discretised version of the) reference distribution $F_x$. By doing so for a set of portfolios $x^{(i)} \in P$, $i \in \{1, \ldots, N\}$, one can estimate $A$ by $\hat{A} = \max \{ \frac{d_x}{\hat{\sigma}_x^2} \mid x^{(i)} \in P, i = 1, \ldots, N \}$. This procedure in particular guarantees that for each portfolio $x^{(i)}$, $i \in \{1, \ldots, N\}$ its corresponding empirical distribution $\hat{G}_x^{(i)}$ lies in the uncertainty set.
In all cases of Corollary 5.2, the robust portfolio optimization problem reduces to solving a second-order cone program. Specifically, for $A = 0$, there is no ambiguity and the investor aims to obtain the optimal portfolio under the assumption that for each portfolio $x$, the portfolio (loss) has known quantile function $F_x^{-1}(u) = -\mu_x + \sigma_x F_Z^{-1}(u)$. The optimal choice $x^*$ is the portfolio\(^{5}\)\(^{6}\) for which $H_g(F_{x^*})$ attains
\[
\min_{x \in P} \{ -\mu_x + \sigma_x H_g(F) \},
\]
where we recall that $\sqrt{V} c_0 = H_g(F)$. When $1 - c_0 \leq A \leq 1$, the Wasserstein distance becomes irrelevant and the optimal portfolio problem reduces to the one considered in Li (2018), that is, to
\[
\min_{x \in P} \{ -\mu_x + \sigma_x \sqrt{V} \}.
\]
In this regard, $\sqrt{V} c_0 = H_g(F) \leq \sqrt{V}$ implies that in the case of no ambiguity ($A = 0$), more emphasis is placed on the expected return component than in the case in which the Wasserstein distance constraint is irrelevant ($1 - c_0 \leq A \leq 1$). Moreover, in this instance, the optimal terminal wealth has a quantile function that is linear in the (non-negative) weight function $\gamma$ (see Corollary 3.3), and thus is bounded from below, a feature that might be considered as not very realistic. The cases $0 < A < (1 - c_0)$ deal with situations that are between these extreme scenarios. Observe that $c_0(1 - A) + \sqrt{A(2 - A)} \sqrt{1 - c_0^2}$ is increasing in $A$ on the interval $[0, 1 - c_0]$ taking value $c_0$ when $A = 0$, and value 1 when $A = 1 - c_0$. As $A$ and thus the degree of ambiguity increases, less emphasis is placed on the expected return leading to optimal portfolios that become more conservative. Note that in all intermediate cases, the optimal quantile function inherits structure of the reference quantile function $F_{x^*}^{-1}$ (see Theorem 3.1).

### 5.2 Insurance portfolio of risks

This section illustrates the best- and worst-case values of VaR on a simulated portfolio of dependent risks, that could, for example, arise from insurance activities. We consider the Pareto–Clayton model that offers a flexible way of modeling portfolios with dependent risks, see Oakes (1989) and Albrecher et al. (2011) for applications in insurance and Dacorogna et al. (2016) and Bernard et al. (2018) for applications in finance. In the Pareto-Clayton model, the portfolio components $X_1, \ldots, X_d$ are, given $\Theta = \theta > 0$, independent and exponentially distributed with parameter $\theta$, that is drawn from a Gamma distribution, that is, $\Theta \sim Gamma(a, b)$. The aggregate portfolio risk $S = \sum_{i=1}^d X_i$ thus follows, conditionally on $\Theta = \theta$, a Gamma($d, \theta$) distribution. By Dubey (1970), Equations (1.1) and (1.3), the scaled aggregate portfolio risk $S/b$ is a Beta distribution of the second kind. Thus, the quantile function of the aggregate portfolio risk $S$ and the first five moments are

---

\(^{5}\)If an elliptical multivariate distribution function is taken as the reference model for $R$, then $F_{x^*}$ is also an elliptical distribution.

\(^{6}\)Note that when $\varepsilon_x \geq 2\sigma_x^2(1 - c_0)$, that is, when $\varepsilon_x = 2\sigma_x^2(1 - c_0)$, it follows that $K_x = \sigma_x^2 c_0$ and thus that $\lambda_x = 0$. 
TABLE 2 Alternative models for the aggregate portfolio risk. The last column reports the Wasserstein distance between the Pareto–Clayton and the alternative model.

| Model               | Abb. | Parameters | \(d_W\) |
|---------------------|------|------------|---------|
| Lognormal model     | LN   | \(\mu_{LN} = 2.4\) \(\sigma_{LN} = 0.36\) | 0.298   |
| Gamma               | \(\Gamma\) | \(\alpha_{\Gamma} = 7.3\) \(\beta_{\Gamma} = 1.5\) | 0.637   |
| Weibull             | W    | \(\lambda_W = 12\) \(k_W = 2.9\) | 3.787   |
| Inverse Gaussian    | IG   | \(\mu_{IG} = 11\) \(\lambda_{IG} = 82\) | 3.802   |
| Inverse Gamma       | IG\(\Gamma\) | \(\alpha_{IG} = 9.3\) \(\beta_{IG} = 93\) | 3.792   |
| Inverse-Weibull     | IW   | \(\lambda_{IW} = 9.3\) \(k_{IW} = 4.4\) | 3.868   |
| Log-Logistic        | LL   | \(\alpha_{LL} = 10\) \(\beta_{LL} = 5.3\) | 0.345   |

explicitly given by

\[
P^{-1}_S(u) = b \frac{G^{-1}_B(u)}{1 - G^{-1}_B(u)} \quad \text{and} \quad \mathbb{E}(S^k) = \frac{b^k B(a-k, d+k)}{B(d, a)}, \quad \text{for} \quad k = 1, \ldots, 5.
\]

where \(B(\cdot, \cdot)\) denotes the Beta function and \(G^{-1}_B\) the quantile function of the Beta distribution with parameters \(d\) and \(a\).

We consider the situation where modelers or experts may not fully agree on the reference Pareto–Clayton model, as available data or estimation may induce uncertainties or the distribution of \(\Theta\) might not seem appropriate. As alternative models for the distribution of the aggregate portfolio risk \(S\), we consider two-parameter distributions with support in \((0, +\infty)\) that are commonly utilized for modeling insurance portfolios. Specifically, we consider a Lognormal model (LN), a Gamma model (\(\Gamma\)), a Weibull model (W), an Inverse Gaussian model (IG), an Inverse Gamma model (IG\(\Gamma\)), an Inverse-Weibull model (IW), and a Log-Logistic model (LL). The different models are described in Table 2 and additional information is provided in Appendix C. For the numerical simulations of the Pareto–Clayton model, we use \(a = 10\), \(b = 1\), and \(d = 100\), as in Bernard et al. (2018). The parameters of the alternative models, reported in Table 2, are obtained by matching the first two moments to those of the reference Pareto–Clayton model, that is, \(\mathbb{E}(S) = 11.11\) and \(\mathbb{E}(S^2) = 140.28\).

We illustrate the reference Pareto–Clayton model (solid black) and the alternative models in Figure 6 via their densities. It is apparent that the considered alternative models can be grouped into light-tailed models (W, IG, IG\(\Gamma\), IW) and heavy-tailed models (LN, \(\Gamma\), LL).

Next, we discuss the choice of Wasserstein distance by means of model uncertainty. Specifically, we let \(\varepsilon\) be the maximum of all Wasserstein distances between the reference and a selection of alternative models. The constructed Wasserstein ball then includes all distribution functions whose Wasserstein distance is smaller or equal to the maximum of all Wasserstein distances between the reference and the considered alternative models and, in particular, all considered alternative models. We report the Wasserstein distances between the Pareto–Clayton model and all alternative models in Table 2 and observe that the Wasserstein distances between the reference distribution and the alternatives from the heavy-tailed models are significantly smaller compared to the alternatives from the light-tailed models. This illustrates that model uncertainty is considered with respect to the reference distribution, which in our case is heavy-tailed. Thus, in a model uncertainty context, changing from the heavy-tailed reference model, that is, Pareto–Clayton, to a light-tailed model, for example, Weibull, entails a higher level of uncertainty compared to other models from the heavy-tailed family, for example, Lognormal.
The best- and worst-case values of VaR$_\alpha$, for $\alpha = 0.9$, 0.95, 0.99, for various uncertainty sets are report in Table 3. The values of VaR$_\alpha$ of the Pareto–Clayton model are, respectively, $\text{VaR}_{0.90} = 16.29$, $\text{VaR}_{0.95} = 18.75$, and $\text{VaR}_{0.99} = 24.79$. The uncertainty sets we consider are $\mathcal{M}(\mu, \sigma)$ with $\mu = \mu_F = 11.11$, $\sigma^2_F = \sigma^2 = 16.82$, $\varepsilon = 0.637$, and $\varepsilon = 3.868$. The tolerance $\varepsilon = 0.637$ corresponds to model uncertainty with respect to heavy-tailed alternative models (LN, $\Gamma$, LL), whereas $\varepsilon = 3.868$ allows for alternative models that are light-tailed (W, IG, IG, IW). To compare these bounds with results in the literature, we calculate the lower and upper bounds on VaR when only the first $k$, $k = 1, \ldots, 5$, moments are known (Bernard et al., 2018). Note that the forth row in Table 3 ($k = 2$) corresponds to $\mathcal{M}(\mu, \sigma)$, that is, with Wasserstein tolerance $\varepsilon = +\infty$. It is apparent that the bounds with a tolerance distance that contain the heavy-tailed alternative models ($\varepsilon = 0.637$) are significantly smaller than the ones corresponding to both heavy- and light-tailed alternative models ($\varepsilon = 3.868$). Furthermore, we report in the last row of Table 3 the bounds on VaR for an uncertainty set with fixed mean, bounded standard deviation, and where the uncertainty set further contains only unimodal distributions (Bernard et al., 2020). We observe in Table 3 that the Wasserstein constraint significantly reduces the bounds and that information on higher moments only affects the bounds in a minor way. Moreover, when only heavy-tailed models are considered as alternatives ($\varepsilon = 0.637$), the length of the bounds is $\frac{18.8 - 12.8}{21.6 - 10.7} = 55\%$ of the length when both light- and heavy-tailed are valid alternative models ($\varepsilon = 0.3868$).
6 | CONCLUSION

We derive quasi explicit best- and worst-case values of a large class of distortion risk measures when the underlying loss distribution has a given mean and variance and lies within a $\sqrt{\varepsilon}$-Wasserstein ball around a given reference distribution. We find that for small Wasserstein tolerance distances, the worst-case distribution function (the distribution function attaining the worst-case value) is probabilistically close to the reference distribution, whereas for large tolerance distances, the worst-case distribution function is no longer affected by the reference distribution. We also derive the worst-case distribution function for VaR and TVaR and show that for small Wasserstein tolerance distances, the worst-case distribution functions of VaR and TVaR are no longer two-point distributions, thus making the bounds attractive for risk management applications.

Our results generalize findings in Li (2018) and Zhu and Shao (2018), which correspond to a Wasserstein tolerance of $\varepsilon = +\infty$. Furthermore, we illustrate the risk bounds on two applications, to portfolio optimization and to model risk assessment of an insurance portfolio.

ACKNOWLEDGMENTS

S. Pesenti acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) with funding reference numbers DGECR-2020-00333 and RGPIN-2020-04289. C. Bernard and S. Vanduffel acknowledge funding from FWO (FWOODYSSII1).

DATA AVAILABILITY STATEMENT

N/A. This is a theory paper which does not make use of data.

ORCID

Carole Bernard https://orcid.org/0000-0002-3996-9260
Silvana M. Pesenti https://orcid.org/0000-0002-6661-6970
Steven Vanduffel https://orcid.org/0000-0003-0511-1044

REFERENCES

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. Journal of Banking & Finance, 26(7), 1505–1518.
Albrecher, H., Constantinescu, C., & Loisel, S. (2011). Explicit ruin formulas for models with dependence among risks. Insurance: Mathematics and Economics, 48(2), 265–270.
Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. Mathematical Finance, 9(3), 203–228.
Barlow, R. E., Bartholomew, D., Bremner, J. M., & Brunk, H. D. (1972). Statistical inference under order restrictions: The theory and application of isotonic regression. Wiley.
Barrieu, P., & Scandolo, G. (2015). Assessing financial model risk. European Journal of Operational Research, 242(2), 546–556.
Bartl, D., Drapeau, S., & Tangpi, L. (2020). Computational aspects of robust optimized certainty equivalents and option pricing. Mathematical Finance, 30(1), 287–309.
BCBS. (2013, October). Consultative document. Fundamental review of the trading book: A revised market risk framework (Technical Report). Basel Committee on Banking Supervision, Bank for International Settlements.
Bernard, C., Denuit, M., & Vanduffel, S. (2018). Measuring portfolio risk under partial dependence information. Journal of Risk and Insurance, 85(3), 843–863.
Bernard, C., Kazzi, R., & Vanduffel, S. (2020). Range value-at-risk bounds for unimodal distributions under partial information. Insurance: Mathematics and Economics, 94(1), 9–24.
Bernard, C., Rüschendorf, L., Vanduffel, S., & Wang, R. (2017). Risk bounds for factor models. *Finance and Stochastics, 21*(3), 631–659.

Blanchet, J., Chen, L., & Zhou, X. Y. (2022). Distributionally robust mean-variance portfolio selection with Wasserstein distances. *Management Science, 68*(9), 6382–6410.

Blanchet, J., & Murthy, K. (2019). Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research, 44*(2), 565–600.

Brighi, B., & Chipot, M. (1994). Approximated convex envelope of a function. *SIAM Journal on Numerical Analysis, 31*(1), 128–148.

Brunk, H. (1965). Conditional expectation given a σ-lattice and applications. *The Annals of Mathematical Statistics, 36*(5), 1339–1350.

Cai, J., Li, J., & Mao, T. (2023). Distributionally robust optimization under distorted expectations. *Operations Research, forthcoming.*

Chen, Z., & Xie, W. (2021). Sharing the value-at-risk under distributional ambiguity. *Mathematical Finance, 31*(1), 531–559.

Clemen, R. T., & Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis, 19*(2), 187–203.

Cont, R., Deguest, R., & Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance, 10*(6), 593–606.

Cornilly, D., Rüschendorf, L., & Vanduffel, S. (2018). Upper bounds for strictly concave distortion risk measures on moment spaces. *Insurance: Mathematics and Economics, 82*, 141–151.

Dacorogna, M. M., Elbahtouri, L., & Kratz, M. (2016). *Explicit diversification benefit for dependent risks* [SCOR papers] 38(1).

Dall’Aglio, G. (1956). Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze, 10*(1-2), 35–74.

De Schepper, A., & Heijnen, B. (2010). How to estimate the value at risk under incomplete information. *Journal of Computational and Applied Mathematics, 233*(9), 2213–2226.

Denuit, M., Genest, C., & Marceau, É. (1999). Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics, 25*(1), 85–104.

Dhaene, J., Kukush, A., Linders, D., & Tang, Q. (2012). Remarks on quantiles and distortion risk measures. *European Actuarial Journal, 2*(2), 319–328.

Dubey, S. D. (1970). Compound gamma, beta and F distributions. *Metrika, 16*(1), 27–31.

EIOPA. (2009). *Directive 2009/138/EC of the European parliament and of the council* (Technical Report). European Insurance and Occupational Pensions Authority.

Embretch, P., & Puccetti, G. (2006). Bounds for functions of dependent risks. *Finance and Stochastics, 10*(3), 341–352.

Embretch, P., Puccetti, G., & Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking & Finance, 37*(8), 2750–2764.

Embretch, P., Wang, B., & Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics, 19*(4), 763–790.

Esfahani, P. M., & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming, 171*(1-2), 115–166.

Glaser, P., & Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance, 14*(1), 29–58.

Grundy, B. D. (1991). Option prices and the underlying asset’s return distribution. *The Journal of Finance, 46*(3), 1045–1069.

Hürlimann, W. (2002). Analytical bounds for two value-at-risk functionals. *ASTIN Bulletin, 32*(2), 235–265.

Hürlimann, W. (2005). Improved analytical bounds for gambler’s ruin probabilities. *Methodology and Computing in Applied Probability, 7*(1), 79–95.

Kaas, R., & Goovaerts, M. J. (1986). Best bounds for positive distributions with fixed moments. *Insurance: Mathematics and Economics, 5*(1), 87–92.

Kou, S., Peng, X., & Heyde, C. C. (2013). External risk measures and Basel accords. *Mathematics of Operations Research, 38*(3), 393–417.

Kusuoka, S. (2001). On law invariant coherent risk measures. In *Advances in mathematical economics* (Vol. 3, pp. 83–95). Springer.
Lam, H. (2016). Robust sensitivity analysis for stochastic systems. *Mathematics of Operations Research, 41*(4), 1248–1275.

Li, J. Y.-M. (2018). Closed-form solutions for worst-case law invariant risk measures with application to robust portfolio optimization. *Operations Research, 66*(6), 1533–1541.

Li, L., Shao, H., Wang, R., & Yang, J. (2018). Worst-case range value-at-risk with partial information. *SIAM Journal on Financial Mathematics, 9*(1), 190–218.

Lo, A. W. (1987). Semi-parametric upper bounds for option prices and expected payoffs. *Journal of Financial Economics, 19*(2), 373–387.

Lux, T., & Rüschendorf, L. (2019). Value-at-Risk bounds with two-sided dependence information. *Mathematical Finance, 29*(3), 967–1000.

Natarajan, K., Sim, M., & Uichanco, J. (2010). Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance, 20*(4), 695–731.

Németh, A. (2003). Characterization of a Hilbert vector lattice by the metric projection onto its positive cone. *Journal of Approximation Theory, 123*(2), 295–299.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association, 84*(406), 487–493.

Pesenti, S. M. (2022). Reverse sensitivity analysis for risk modelling. *Risks, 10*(7), 141.

Pesenti, S. M., & Jaimungal, S. (2023). Portfolio optimisation within a Wasserstein ball. *SIAM Journal on Financial Mathematics, 3744994*, forthcoming.

Pesenti, S. M., Millossovich, P., & Tsanakas, A. (2016). Robustness regions for measures of risk aggregation. *Dependence Modeling, 4*(1), 348–367.

Pesenti, S. M., Wang, Q., & Wang, R. (2020). Optimizing distortion riskmetrics with distributional uncertainty. Available at SSRN, 3728638. [https://arxiv.org/abs/2011.04889](https://arxiv.org/abs/2011.04889)

Pflug, G., & Wozabal, D. (2007). Ambiguity in portfolio selection. *Quantitative Finance, 7*(4), 435–442.

Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research, 55*(1), 98–112.

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

Wang, B., & Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis, 102*(10), 1344–1360.

Wang, Q., Wang, R., & Wei, Y. (2020). Distortion riskmetrics on general spaces. *ASTIN Bulletin, 50*(3), 827–851.

Wang, R., Xu, Z. Q., & Zhou, X. Y. (2019). Dual utilities on risk aggregation under dependence uncertainty. *Finance and Stochastics, 23*(4), 1025–1048.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin, 26*(1), 71–92.

Wozabal, D. (2014). Robustifying convex risk measures for linear portfolios: A nonparametric approach. *Operations Research, 62*(6), 1302–1315.

Zhu, W., & Shao, H. (2018). Closed-form solutions for extreme-case distortion risk measures and applications to robust portfolio management. Available at SSRN.

---

**How to cite this article:** Bernard, C., Pesenti, S. M., & Vanduffel, S. (2024). Robust distortion risk measures. *Mathematical Finance, 34*, 774–818. [https://doi.org/10.1111/mafi.12414](https://doi.org/10.1111/mafi.12414)

---

**APPENDIX A: ISOTONIC PROJECTION**

Here, we collect properties of the metric projection defined in Equation (9). For this, we denote by \( \mathcal{N} = \{ \ell | \ell : (0, 1) \to \mathbb{R}, \int_0^1 \ell(u)^2 \, du < +\infty \} \) the space of square-integrable functions on (0,1).
and consider the following metric projection from $\mathcal{N}$ to $\mathcal{K}$ (set defined by Equation 8):

$$P_K(\ell) = \arg \min_{k \in \mathcal{K}} ||\ell - k||^2, \quad \ell \in \mathcal{N}.$$ 

A metric projection is called *isotone* if the projection preserves the order induced by $\mathcal{K}$ (Németh, 2003). The partial order $\preceq_{\mathcal{K}}$ on $\mathcal{N}$ induced by $\mathcal{K}$ is given, for any $\ell_1, \ell_2 \in \mathcal{N}$, by

$$\ell_1 \preceq_{\mathcal{K}} \ell_2 \quad \text{if} \quad \ell_2 - \ell_1 \in \mathcal{K}.$$ 

**Proposition A.1** Theorem 3. Németh (2003). The metric projection $P_K$ is isotone and subadditive; that is, $P_K$ fulfills for all $\ell_1, \ell_2 \in \mathcal{N}$:

1. **Isotone:** If $\ell_1 \preceq_{\mathcal{K}} \ell_2$, then $P_K(\ell_1) \preceq_{\mathcal{K}} P_K(\ell_2)$.
2. **Subadditive:** $P_K(\ell_1 + \ell_2) \preceq_{\mathcal{K}} P_K(\ell_1) + P_K(\ell_2)$.

In this paper, we compare only projections of functions of the type $\gamma + \lambda F^{-1}$, for fixed $F$ and $\gamma$. Thus, we view the projection of $\gamma + \lambda F^{-1}$ as a function of $\lambda$ and define $k_\lambda = P_K(\gamma + \lambda F^{-1})$, $\lambda \geq 0$.

**Proposition A.2.** For fixed $\gamma$ and $F$, the metric projection $k_\lambda$ is an isotonic projection with the following properties:

1. **Isotone:** For $\lambda_1 \leq \lambda_2$, it holds that $k_{\lambda_1} \preceq_{\mathcal{K}} k_{\lambda_2}$.
2. **Continuous:** If $\lambda_n \to \lambda$, then $\lim_{n \to +\infty} ||k_{\lambda_n} - k_{\lambda}||^2 = 0$.

**Proof of Proposition A.2.** Property 1. (isotonicity of $k_\lambda$) follows by noting that $\lambda_1 \leq \lambda_2$ implies $\gamma + \lambda_2 F^{-1} - (\gamma + \lambda_1 F^{-1}) = (\lambda_2 - \lambda_1)F^{-1} \in \mathcal{K}$. Thus, by isotonicity of $P_K$, we obtain

$$k_{\lambda_1} = P_K(\gamma + \lambda_1 F^{-1}) \preceq_{\mathcal{K}} P_K(\gamma + \lambda_2 F^{-1}) = k_{\lambda_2}.$$ 

To prove continuity of $k_\lambda$ (Property 2.), denote $\lambda_n = \lambda + e_n$, for any sequence $e_n$ with $e_n \to 0$, as $n \to +\infty$. Then by Theorem 2.3 in Brunk (1965), it holds that

$$||k_{\lambda_n} - k_{\lambda}||^2 \leq ||\gamma + (\lambda + e_n) F^{-1} - \gamma - \lambda F^{-1}||^2 = |e_n|^2 ||F^{-1}||^2 \to 0, \quad \text{for} \quad n \to +\infty. \qed$$

The next proposition collects a few properties of the isotonic projection $P_K$ and thus of $k_\lambda$. These properties follow from Brunk (1965, Theorem 2.2. and Corollary 2.3) and we provide short proofs adapted to our setting. For this, we denote by $\langle \cdot, \cdot \rangle$ the inner product on $\mathcal{N}$—that is, $\langle \ell_1, \ell_2 \rangle = \int_0^1 \ell_1(u)\ell_2(u)du$, for $\ell_1, \ell_2 \in \mathcal{N}$.

**Proposition A.3.** The isotonic projection $P_K$ satisfies the following properties for all $\ell \in \mathcal{N}$ and $k \in \mathcal{K}$:

1. $\langle \ell - P_K(\ell), P_K(\ell) - k \rangle \geq 0$;
2. $\langle \ell, P_K(\ell) \rangle = \langle P_K(\ell), P_K(\ell) \rangle$;
3. $\langle \ell, k \rangle \leq \langle P_K(\ell), k \rangle$.

**Proof of Proposition A.3.** We follow the arguments of the proof of Theorem 1.3.2 by Barlow et al. (1972).
Property 1. For $\ell \in \mathcal{N}$, $f \in \mathcal{K}$, and $w \in [0, 1]$ define the function $(1 - w)P_K(\ell) + wf$. By definition of the isotonic projection, it holds that $(1 - w)P_K(\ell) + wf \in \mathcal{K}$. Next, we consider the function 

$$w \mapsto \langle \ell - (1 - w)P_K(\ell) - wf, \ell - (1 - w)P_K(\ell) - wf \rangle,$$

which obtains its minimum in $w = 0$, by definition of the isotonic projection of $P_K(\ell)$. The first-order condition at $w = 0$ implies that 

$$2 \left( \ell - (1 - w)P_K(\ell) - wf, P_K(\ell) - f \right) |_{w=0} = 2 \left( \ell - P_K(\ell), P_K(\ell) - f \right) \geq 0.$$

Property 2. By the definition of an isotonic projection, it holds that 

$$\langle P_K(\ell) - \ell, P_K(\ell) - \ell \rangle \leq \langle k - \ell, k - \ell \rangle$$

for all $\ell, k \in \mathcal{N}$. Choosing $k = cP_K(\ell)$ with $c > 0$, we obtain 

$$2(c - 1) \langle P_K(\ell), \ell \rangle \leq (c^2 - 1) \langle P_K(\ell), P_K(\ell) \rangle.$$

Thus, for $0 < c_1 < 1$ and $c_2 > 1$, the above inequality becomes 

$$\frac{1-c^2}{2(1-c_1)} \langle P_K(\ell), P_K(\ell) \rangle \leq \langle P_K(\ell), \ell \rangle \leq \frac{c^2-1}{2(c_2-1)} \langle P_K(\ell), P_K(\ell) \rangle.$$

Taking limits with $c_1 \nearrow 1$ and $c_2 \searrow 1$ concludes the statement of the property.

Property 3. For $\ell \in \mathcal{N}$ and $k \in \mathcal{K}$, we obtain, by first applying Property 2 and then Property 1.

$$\langle \ell - P_K(\ell), P_K(\ell) - k \rangle = \langle \ell, P_K(\ell) \rangle - \langle \ell, k \rangle - \langle P_K(\ell), P_K(\ell) \rangle + \langle k, P_K(\ell) \rangle$$

$$= -\langle \ell, k \rangle + \langle k, P_K(\ell) \rangle \geq 0.$$  

**APPENDIX B: PROOFS**

We denote by $\mathcal{M}_c^{-1}(\mu, \sigma) = \{G^{-1} \mid G \in \mathcal{M}_c(\mu, \sigma)\}$ the corresponding set of quantile functions in $\mathcal{M}_c(\mu, \sigma)$. Further, we denote $\mathcal{M}^{-1}(\mu, \sigma) = \{G^{-1} \mid G \in \mathcal{M}(\mu, \sigma)\}$.

**Lemma B.1.** If $\int_0^1 \gamma(u) \, du = +\infty$, then $\inf_{G \in \mathcal{M}_c(\mu, \sigma)} H_{\phi}(G) = \sup_{G \in \mathcal{M}_c(\mu, \sigma)} H_{\phi}(G) = +\infty$.

**Proof of Lemma B.1.** For $v \in (0, 1)$ define 

$$h_{v,\lambda}(u) = \mu + \sigma \left( \frac{\min \left\{ k_\lambda^1(u), v \right\} - a_{v,\lambda}}{b_{v,\lambda}} \right), \quad 0 < u < 1,$$

where $k_\lambda^1$ is defined in Equation (9), $a_{v,\lambda} = \mathbb{E}(\min\{k_\lambda^1(U), v\})$, and $b_{v,\lambda} = \text{std}(\min\{k_\lambda^1(U), v\})$. By construction $h_{v,\lambda} \in \mathcal{M}^{-1}(\mu, \sigma)$ for all $v \in (0, 1)$. Moreover,

$$\int_0^1 h_{v,\lambda}(u)\gamma(u) \, du = \mu + \frac{\sigma}{b_{v,\lambda}} a_{v,\lambda} \left( \int_0^1 \frac{\min \left\{ k_\lambda^1(u), v \right\}}{a_{v,\lambda}} \gamma(u) \, du - 1 \right)$$

$$\geq \mu + \frac{\sigma}{b_{v,\lambda}} a_{v,\lambda} \left( \int_0^1 \left( \frac{\min \left\{ k_\lambda^1(u), v \right\}}{a_{v,\lambda}} \right)^2 \, du - 1 \right) = \mu + \frac{b_{v,\lambda}}{a_{v,\lambda}},$$
which diverges to $+\infty$, for $v \nearrow 1$.

### B.1 Proof of Theorem 3.1 and Corollary 3.3

For the proof of Theorem 3.1, we need the following lemmas.

**Lemma B.2.** Let $\lambda \geq 0$, and $\gamma$ be non-decreasing, then

$$\sup_{h \in \mathcal{M}^{-1}(\mu,\sigma)} \langle \gamma + \lambda F^{-1}(u), h \rangle$$

has a unique solution given by

$$h_\lambda(u) = \mu + \sigma \left( \frac{\gamma(u) + \lambda F^{-1}(u) - a_\lambda}{b_\lambda} \right), \quad 0 < u < 1,$$

where $a_\lambda = \mathbb{E}(\gamma(U) + \lambda F^{-1}(U))$ and $b_\lambda = \text{std}(\gamma(U) + \lambda F^{-1}(U))$. Moreover, $\text{corr}(F^{-1}(U), h_\lambda(U))$ is continuous in $\lambda$, for $\lambda \geq 0$.

**Proof of Lemma B.2.** The function $h_\lambda$, defined in Equation (B.1), belongs to $\mathcal{M}^{-1}(\mu,\sigma)$. Let $h \in \mathcal{M}^{-1}(\mu,\sigma)$, then, application of the Cauchy–Schwarz inequality implies that $\langle \gamma + \lambda F^{-1}, h \rangle \leq \langle \gamma + \lambda F^{-1}, h_\lambda \rangle$, with equality if and only if $h$ is linear in $\gamma + \lambda F^{-1}$. Continuity of $\text{corr}(F^{-1}(U), h_\lambda(U))$ is immediate.

**Lemma B.3.** The solution $\lambda$ to $d_W(F^{-1}, h_\lambda)^2 = \mathbb{E}((F^{-1}(U) - h_\lambda(U))^2) = \varepsilon$, where $h_\lambda(\cdot)$ is defined in Lemma B.2 is explicit and is given by

$$\lambda = \frac{\sqrt{\Delta - C_{\gamma,F}}}{\sigma_F^2}, \quad \text{where} \quad \Delta = \frac{K^2 \left( C_{\gamma,F}^2 - V \sigma_F^2 \right)}{K^2 - \sigma^2 \sigma_F^2},$$

with $V = \text{var}(\gamma(U))$, $C_{\gamma,F} = \text{cov}(F^{-1}(U), \gamma(U))$, and

$$K = \frac{1}{2} \left( \mu_F^2 + \sigma_F^2 + \mu^2 + \sigma^2 - 2\mu \mu_F - \varepsilon \right) \geq 0.$$

**Proof of Lemma B.3.** Let $h_\lambda(u) \in \mathcal{M}^{-1}(\mu,\sigma)$ defined in Equation (B.1). Solving for $\lambda$ such that $d_W(F^{-1}, h_\lambda)^2 = \varepsilon$ amounts to solving

$$\varepsilon = \mu_F^2 + \sigma_F^2 + \mu^2 + \sigma^2 - 2 \text{cov}(F^{-1}(U), h_\lambda(U)) - 2\mu \mu_F.$$

Using the expression of $K$ in Equation (B.3), Equation (B.4) is simply

$$\text{cov}(F^{-1}(U), h_\lambda(U)) = K,$$

which ensures that $K \geq 0$. Using the expression of $h_\lambda$,

$$\text{cov}(F^{-1}(U), h_\lambda(U)) = \frac{\sigma}{b_\lambda} \text{cov}(F^{-1}(U), \gamma(U) + \lambda F^{-1}(U)) = \frac{\sigma}{b_\lambda} (\text{cov}(F^{-1}(U), \gamma(U)) + \lambda \sigma_F^2).$$

Rearranging, we obtain

$$K b_\lambda = \sigma \text{cov}(F^{-1}(U), \gamma(U)) + \lambda \sigma_F^2.$$
Next, denote by $V = \text{var}(\gamma(U))$ and $C_{\gamma,F} = \text{cov}(F^{-1}(U), \gamma(U))$, then the expression of $b_\lambda = \text{std}(\gamma(U) + \lambda F^{-1}(U))$ becomes $b_\lambda^2 = V + 2\lambda C_{\gamma,F} + \lambda^2 \sigma_F^2$ and the square of Equation (B.6) becomes

$$K^2(V + 2\lambda C_{\gamma,F} + \lambda^2 \sigma_F^2) = (\sigma C_{\gamma,F} + \lambda \sigma_F^2)^2.$$ 

This is a second degree equation in $\lambda$, which can also be written as

$$\sigma_F^2 \lambda^2 + 2C_{\gamma,F} \lambda + \frac{K^2 V - \sigma^2 C_{\gamma,F}^2}{K^2 - \sigma^2 \sigma_F^2} = \sigma_F^2 \Delta - \frac{1}{\sigma_F^2} \left( \Delta - C_{\gamma,F}^2 \right) = 0, \quad \text{as} \quad \sigma_F^2 K^2 - \sigma^2 \sigma_F^2,$$

(B.7)

$$\frac{\sigma_F^2 C_{\gamma,F}^2 - K^2 V}{K^2 - \sigma^2 \sigma_F^2} + C_{\gamma,F}^2 = \Delta.$$ 

(B.8)

Observe that if $\gamma(\cdot) \neq F^{-1}(\cdot)$, then $C_{\gamma,F} > 0$ as the covariance between two comonotonic variables. $K^2 - \sigma^2 \sigma_F^2 < 0$ by definition of $K$ as a covariance between two variables with respective variances $\sigma^2$ and $\sigma_F^2$ and that are not perfectly correlated, and $C_{\gamma,F}^2 - V \sigma_F^2 \leq 0$, by definition of $C_{\gamma,F}$ as a covariance between two variables with respective variances $V$ and $\sigma_F^2$. Thus, from the expression (B.8), we have that $\Delta > 0$. As a consequence, Equation (B.7) then has two roots, of which only one, $\lambda = \frac{\sqrt{\Delta - C_{\gamma,F}^2}}{\sigma_F^2}$, is positive.

Proof of Theorem 3.1. For concave distortion risk measures, Equation (5b) is equivalent to

$$\sup_{h \in M_1^{-1}(\mu, \sigma)} \langle \gamma, h \rangle,$$ 

(B.9)

where $\gamma$ is non-decreasing. For $\lambda \geq 0$, let $h_\lambda \in M^{-1}(\mu, \sigma)$ denote the quantile function defined in Equation (B.1). The Wasserstein distance between $F^{-1}$ and $h_\lambda$ is given by

$$d_W(F^{-1}, h_\lambda) = (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F (1 - c_\lambda),$$

where $c_\lambda = \text{corr}(F^{-1}(U), h_\lambda(U))$. Note that by Lemma B.2, $c_\lambda : [0, +\infty) \rightarrow [-1, 1]$ is continuous with $c_0 = \text{corr}(F^{-1}(U), h_0(U)) = \text{corr}(F^{-1}(U), \gamma(U))$ and $\lim_{\lambda \rightarrow +\infty} c_\lambda = 1$.

Case 1. Let $(\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F (1 - c_0)$. We split this part into three steps. First, we show that the optimal quantile function has to be of the form $h_\lambda$ for some $\lambda \geq 0$. Second, we show that the solution lies at the boundary of the Wasserstein ball. Third, we show uniqueness.

For the first step, let $h_1 \in M_1(\mu, \sigma)$ be a solution to problem (B.9) with $d_W(F^{-1}, h_1) = \sqrt{\varepsilon_1}$ and $(\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon_1 \leq \varepsilon$. Then, by continuity of $c_\lambda$, there exists $\lambda_1 > 0$ and a corresponding $h_{\lambda_1} \in M_1(\mu, \sigma)$ such that $d_W(F^{-1}, h_{\lambda_1}) = \sqrt{\varepsilon_1}$. Moreover, it holds that

$$d_W(F^{-1}, h_1) = d_W(F^{-1}, h_{\lambda_1}) \quad \Leftrightarrow \quad \langle F^{-1}, h_1 \rangle = \langle F^{-1}, h_{\lambda_1} \rangle.$$

Applying Lemma B.2, we obtain $\langle \gamma, h_1 \rangle \leq \langle \gamma, h_{\lambda_1} \rangle$; thus, $h_{\lambda_1}$ improves upon $h_1$ and we conclude the first step.

For the second step, let $\lambda \geq 0$ be such that $h_\lambda$ satisfies $d_W(F^{-1}, h_\lambda) = \sqrt{\varepsilon}$. By construction, it holds that

$$d_W(F^{-1}, h_{\lambda_1}) \leq d_W(F^{-1}, h_\lambda) \quad \Leftrightarrow \quad \langle F^{-1}, h_{\lambda_1} \rangle \geq \langle F^{-1}, h_\lambda \rangle.$$

(B.10)
Furthermore, by Lemma B.2, we have that \( \langle \gamma + \lambda F^{-1}, h_\lambda \rangle \geq \langle \gamma + \lambda F^{-1}, h_{\lambda_1} \rangle \), which, together with Equation (B.10), implies
\[
\langle \gamma, h_\lambda - h_{\lambda_1} \rangle \geq \lambda \langle F^{-1}, h_\lambda - h_{\lambda_1} \rangle \geq 0.
\]
Therefore, we obtain that \( \langle \gamma, h_\lambda \rangle \succeq \langle \gamma, h_{\lambda_1} \rangle \), and hence that the solution lies at the boundary of the Wasserstein ball. Uniqueness follows from the uniqueness of \( h_{\lambda} \) as the solution of Lemma B.2.

The worst-case value \( H_\delta(h_\lambda) \) is
\[
H_\delta(h_\lambda) = \int_0^1 h_\lambda(u)\gamma(u)du = \mu + \frac{\sigma}{b_\lambda} \left( \int_0^1 \gamma(u)(\gamma(u) + \lambda F^{-1}(u))du - a_\lambda \right)
\]
\[
= \mu + \sigma \text{std}(\gamma(U)) \text{corr}(\gamma(U), \gamma(U) + \lambda F^{-1}(U)).
\]

The explicit expression for \( \lambda \) then follows from Lemma B.3 and is given in Equation (B.2).

Case 2. Let \( \left( \mu_F - \mu \right)^2 + (\sigma_F - \sigma)^2 + 2\sigma\sigma_F(1 - c_0) \leq \varepsilon; \) then \( h_0 \in \mathcal{M}^{-1}(\mu, \sigma) \) and thus is admissible. Moreover, it is straightforward that for all \( \lambda > 0, \langle \gamma, h_0 \rangle \succeq \langle \gamma, h_\lambda \rangle \). Hence, \( h_0 \) solves problem (B.9). □

**Proof of Corollary 3.3.** This result follows directly from Theorem 3.1, case 2. □

**B.2 | Proof of Theorem 3.7 and Corollary 3.9**

The following lemma is needed to prove Theorem 3.7.

**Lemma B.4.** Let \( \lambda > 0 \) and \( g \) be a distortion function with weight function \( \gamma \). Then,
\[
\sup_{h \in \mathcal{M}^{-1}(\mu, \sigma)} \langle \gamma + \lambda F^{-1}, h \rangle
\]
has a unique solution given by
\[
h_\lambda^\dagger(u) = \mu + \sigma \left( \frac{k_\lambda^\dagger(u) - a_\lambda}{b_\lambda} \right), \quad 0 < u < 1,
\]
where \( a_\lambda = \mathbb{E}(k_\lambda^\dagger(U)) \) and \( b_\lambda = \text{std}(k_\lambda^\dagger(U)) \), and \( k_\lambda^\dagger \) is given in Equation (9). Moreover, \( \text{corr}(F^{-1}(U), h_\lambda^\dagger(U)) \) is continuous in \( \lambda \), for \( \lambda > 0 \).

**Proof of Lemma B.4.** For \( \lambda > 0 \), note that the isotonic projection \( k_\lambda^\dagger \) is nonconstant, which implies that \( h_\lambda^\dagger \), as defined in Equation (B.11), fulfills \( h_\lambda^\dagger \in \mathcal{M}^{-1}(\mu, \sigma) \). Let \( h \in \mathcal{M}^{-1}(\mu, \sigma) \); then \( k(u) = b_\lambda (h(u) - \mu)/\sigma + a_\lambda \) is a non-decreasing function. Moreover, it holds that \( ||k_\lambda^\dagger||^2 = ||k||^2 \). Thus, the following inequalities are equivalent:
\[
||\gamma + \lambda F^{-1} - k_\lambda^\dagger||^2 \leq ||\gamma + \lambda F^{-1} - k||^2 \iff \langle \gamma + \lambda F^{-1}, k_\lambda^\dagger \rangle \geq \langle \gamma + \lambda F^{-1}, k \rangle \]
\[
\iff \langle \gamma + \lambda F^{-1}, h_\lambda^\dagger \rangle \geq \langle \gamma + \lambda F^{-1}, h \rangle.
\]

Note that unless \( k_\lambda^\dagger = k \), the inequalities are strict, which implies uniqueness of the solution. Finally, continuity of \( \text{corr}(F^{-1}(U), h_\lambda^\dagger(U)) = \text{corr}(F^{-1}(U), k_\lambda^\dagger(U)) \) follows from continuity of the isotonic projection \( k_\lambda \), Property 2 in Proposition A.2. □
Proof of Theorem 3.7. For \( \lambda > 0 \), define \( h_\lambda^1 \) as in Equation (B.11) and note that \( h_\lambda^1 \in \mathcal{M}^{-1}(\mu, \sigma) \). The Wasserstein distance between \( F^{-1} \) and \( h_\lambda^1 \) is given by (see also Theorem 3.1) \( d_W(F^{-1}, h_\lambda^1)^2 = (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F(1 - c_\lambda) \), where \( c_\lambda = \text{corr}(F^{-1}(U), h_\lambda^1(U)) \). By Lemma B.4, \( c_\lambda : (0, +\infty) \rightarrow [-1,1] \) is continuous in \( \lambda > 0 \), with \( \lim_{\lambda \to +\infty} c_\lambda = 1 \).

Case 1. Let \( (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 < \varepsilon < (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F(1 - c_0) \). Similar to the proof of Theorem 3.1, we split the proof into three steps: first, the worst-case quantile function is of the form \( h_\lambda \), second, the solution lies at the boundary of the Wasserstein ball, and third, we show that it is unique. The first two steps follow along the lines of the proof of Theorem 3.1 case 1, replacing \( h_\lambda \) with \( h_\lambda^1 \) and using in the second step the properties of the isotonic projection \( h_\lambda^1 \). Uniqueness follows from the uniqueness of \( h_\lambda^1 \) as the solution to Equation (B.4) in Lemma B.4.

Case 2. If \( (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma_F(1 - c_0) \leq \varepsilon \) and \( \gamma^1 \) is not constant, then Lemma B.4 applies for all \( \lambda \geq 0 \). Let \( h_0^1 \) be as defined in Equation (B.11) with \( \lambda = 0 \). Then, \( h_0^1 \in \mathcal{M}^{-1}(\mu, \sigma) \). Moreover, for all \( h \in \mathcal{M}^{-1}(\mu, \sigma) \) it holds, applying Lemma B.4 with \( \lambda = 0 \), that \( \langle \gamma, h \rangle \leq \langle \gamma, h_0^1 \rangle \). Hence, \( h_0^1 \) cannot be improved and thus yields the unique optimal solution.

If \( \gamma^1 \) is constant, then we find that for all \( h \in \mathcal{M}^{-1}(\mu, \sigma) \), \( \langle \gamma, h \rangle \leq \langle \gamma^1, h \rangle = \mu \), where the first inequality follows from the properties of the isotonic projection (Proposition A.3, Property 3) and the equality follows from the fact that \( \gamma^1 = 1 \) as \( \langle \gamma, 1 \rangle = \langle \gamma^1, 1 \rangle \) (applying Proposition A.3, Property 3). Finally, we show that \( \mu \) is indeed the supremum. Consider the case in which \( \mu \geq 0 \) and define \( t_\alpha = \alpha \mu \gamma^1 + (1 - \alpha)z_\alpha \), for \( 0 < \alpha < 1 \), and where \( z_\alpha \) is a non-negative non-decreasing function such that \( \mathbb{E}(z_\alpha(U)) = \mu \) and \( \text{std}(z_\alpha(U)) = \sigma/(1 - \alpha) \). Hence, \( t_\alpha \in \mathcal{M}^{-1}(\mu, \sigma) \), for \( 0 < \alpha < 1 \). Moreover, for every \( \varepsilon > 0 \) that are sufficiently small, there exists \( 0 < \alpha < 1 \) such that

\[
\langle \gamma, t_\alpha \rangle = \langle \gamma, \alpha \mu \gamma^1 \rangle + \langle \gamma, (1 - \alpha)z_\alpha \rangle \geq \alpha \mu + \mu - \varepsilon.
\]

For \( \mu < 0 \), the proof is similar and is omitted.

Proof of Corollary 3.9. Corollary 3.9 follows from Theorem 3.7. We prove only that equation (11) (when \( \gamma^1 \) is not constant) is equivalent to the worst-case value in Theorem 3.7 (case 2). From the characterization of the isotonic projection—that is, applying Property 3 of Proposition A.3 with \( k = \pm 1 \)—we obtain that \( \langle \gamma^1, 1 \rangle = \langle \gamma, 1 \rangle = 1 \); thus, \( \mathbb{E}(\gamma^1(U)) = \mathbb{E}(\gamma(U)) = 1 \). Further, by Property 2 of Proposition A.3, it holds that \( \mathbb{E}(\gamma(U))^2 = \mathbb{E}(\gamma(U)\gamma(U)) \). Hence,

\[
\text{corr}(\gamma(U), \gamma^1(U)) = \frac{\text{cov}(\gamma^1(U), \gamma(U))}{\text{std}(\gamma^1(U))\text{std}(\gamma(U))} = \frac{\text{var}(\gamma^1(U))}{\text{std}(\gamma^1(U))\text{std}(\gamma(U))} = \frac{\text{std}(\gamma^1(U))}{\text{std}(\gamma(U))}.
\]

B.3  Proof of Theorem 3.12
The following lemma is needed for the proof of Theorem 3.12.

Lemma B.5. Let \( \lambda > 0 \) and \( g \) be a distortion function. Then,

\[
\inf_{h \in \mathcal{M}^{-1}(\mu, \sigma)} \langle \gamma - \lambda F^{-1}, h \rangle
\]

has a unique solution given by

\[
h_\lambda^1(u) = \mu + \sigma \left( \frac{a_\lambda - k_\lambda(u)}{b_\lambda} \right), \quad 0 < u < 1,
\]

(B.12)
where \( a_\lambda = \mathbb{E}(k^\lambda(U)) \) and \( b_\lambda = \text{std}(k^\lambda(U)) \), and \( k^\lambda \) is given in Equation (12). Moreover, \( \text{corr}(F^{-1}(U), h^\lambda(U)) \) is continuous in \( \lambda \), for \( \lambda > 0 \).

**Proof of Lemma B.5.** For \( \lambda > 0 \), note that \( k^\lambda \) is not constant and that \( h^\lambda \), as defined in Equation (B.12), fulfills \( h^\lambda \in \mathcal{M}^{-1}(\mu, \sigma) \). Let \( h \in \mathcal{M}^{-1}(\mu, \sigma) \), then \( k(u) = b_\lambda(\mu - h(u))/\sigma + a_\lambda \) is a non-increasing function. Moreover, it holds that \( ||k^\lambda||^2 = ||k||^2 \). Thus, the following inequalities are equivalent:

\[
||y - \lambda F^{-1} - k^\lambda||^2 \leq ||y - \lambda F^{-1} - k||^2 \iff \langle y - \lambda F^{-1}, k^\lambda \rangle \geq \langle y - \lambda F^{-1}, k \rangle
\]

\[
\iff \langle y - \lambda F^{-1}, h^\lambda \rangle \leq \langle y - \lambda F^{-1}, h \rangle.
\]

Note that unless \( k^\lambda = k \), the inequalities are strict, which implies that the solution is unique.

Finally, continuity of \( \text{corr}(F^{-1}(U), h^\lambda(U)) = \text{corr}(F^{-1}(U), k^\lambda(U)) \) follows from continuity of the isotonic projection \( k^\lambda \), that is, Proposition A.2, Property 2. \( \square \)

**Proof of Theorem 3.12.** For \( \lambda > 0 \), define \( h^\lambda \) as in Equation (B.12) and observe that \( h^\lambda \in \mathcal{M}^{-1}(\mu, \sigma) \). Then, \( d_W(F^{-1}, h^\lambda)^2 = (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 + 2\sigma\sigma_F(1 - c_\lambda) \), where \( c_\lambda = \text{corr}(F^{-1}(U), h^\lambda(U)) \). By Lemma B.5, \( c_\lambda : (0, +\infty) \to [-1, 1] \) is continuous in \( \lambda \), \( \lambda > 0 \), with \( \lim_{\lambda \to +\infty} c_\lambda = 1 \). The remainder of the proof follows along the arguments of the proof of Theorem 3.7. \( \square \)

### B.4 | Proofs of Corollaries 4.1 and 4.2 and Propositions 4.3 and 4.4.

**Proof of Corollary 4.1.** These results are well-known and we provide here a proof using the statements derived in this paper. First, we prove the worst-case value of the TVaR. Recall that its weight function \( \gamma(u) = \frac{1}{1-\alpha} \mathbb{1}_{(\alpha, 1)}(u) \) is non-decreasing and, hence, \( \gamma^\uparrow(u) = \gamma(u) \) for all \( u \in (0, 1) \). Applying Corollary 3.9, we obtain \( \int_0^1 (\gamma^\uparrow(u) - 1)^2 \, du = \frac{\alpha}{1-\alpha} \), and therefore

\[
\sup_{G \in \mathcal{M}(\mu, \sigma)} \text{TVaR}_\alpha(G) = \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}}.
\]

Next, we calculate the worst-case value of RVaR\( _{\alpha, \beta} \) with weight function \( \gamma(u) = \frac{1}{\beta - \alpha} \mathbb{1}_{(\alpha, \beta]}(u) \). To apply Corollary 3.9, we first derive the isotonic projection of \( \gamma \). Using similar arguments as in the proof of Proposition 4.3 below, we find that the isotonic projection of \( \gamma(u) \) is equal to \( \gamma(u) \) on \( u \in (0, \alpha] \) and constant on \( (\alpha, 1] \), on which it takes the value \( c > 0 \), which must be such that \( \langle \gamma, 1 \rangle = \langle \gamma^\uparrow, 1 \rangle = 1 \). Hence, \( c = \frac{1}{1-\alpha} \), and the worst-case RVaR\( _{\alpha, \beta} \) equals that of TVaR\( _{\alpha} \). Consider the right-continuous quantile function \( H^{-1} \) taking the value \( \mu - \sigma \sqrt{\frac{1-\alpha}{\alpha}} \) on \( (0, \alpha] \), and the value \( \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}} \) on \( [\alpha, 1] \), and observe that it has mean \( \mu \) and standard deviation \( \sigma \). As \( \text{VaR}^+_{\alpha}(H^{-1}) = \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}} \), it follows that

\[
\sup_{G \in \mathcal{M}(\mu, \sigma)} \text{VaR}^+_{\alpha}(G) = \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}}.
\]

Finally, assume that \( \sup_{G \in \mathcal{M}(\mu, \sigma)} \text{VaR}_{\alpha}(G) = \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}} - \delta \), for some \( \delta > 0 \). One can easily construct a left-continuous quantile function \( H^{-1} \) having mean \( \mu \) and standard deviation \( \sigma \), such
that $\text{VaR}_\alpha(H) > \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}} - \delta$. Hence,

$$\sup_{G \in M(\mu, \sigma)} \text{VaR}_\alpha(G) = \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}}. \square$$

**Proof of Corollary 4.2.** The results for the case of RVaR $\alpha$, VaR $\alpha$, and VaR $\alpha^+$, $0 < \alpha < \beta < 1$ follow from Remark 3.14 and Corollary 3.9 in a straightforward manner.

For the lower bound on TVaR $\alpha$, $0 < \alpha < 1$, observe that the isotonic projection $\gamma^1(u)$ of $\gamma(u) = \frac{1}{1-\alpha} \mathbb{1}_{(\alpha,1]}(u)$ is constant, and thus that $\gamma^1 = 1$, as $\langle \gamma, 1 \rangle = \langle \gamma^1, 1 \rangle$ (applying Proposition A.3, Property 3 with $k \equiv 1$ and $k \equiv -1$). Hence,

$$\inf_{G \in M(\mu, \sigma)} \text{TVaR}_\alpha(G) = \mu. \square$$

**Proof of Proposition 4.3.** By Theorem 3.7 case 1, we have to calculate the isotonic projection of $\frac{1}{\beta-\alpha} \mathbb{1}_{(\alpha,\beta]} + \lambda F^{-1}$ onto the space of non-decreasing functions. According to Lemma 5.1 in Brighi and Chipot (1994), the isotonic projection has the form

$$k^*_\lambda(u) = \left( \frac{1}{\beta-\alpha} \mathbb{1}_{(\alpha,\beta]}(u) + \lambda F^{-1}(u) \right) \mathbb{1}_{(\bigcup_j I_j) \setminus \bigcup_j \mathbb{1}_{I_j}}(u) + \sum_{j \in J} c_j \mathbb{1}_{I_j}(u), \quad 0 < u < 1, \quad (B.13)$$

where $J$ is a countable index set, and $I_j \subset [0,1], j \in J$, are disjoint intervals, and for any $j \in J$, $I_j^c = [0,1] \setminus I_j$ denotes the complement of the set $I_j$.

We first show that $J$ contains only one element. For this, let $k^*$ be an optimal solution with representation given by Equation (B.13). By definition, $k^*$ minimizes $\| \frac{1}{\beta-\alpha} \mathbb{1}_{(\alpha,\beta]} + \lambda F^{-1} - k \|_2^2$ over all non-decreasing functions $k$. Moreover,

$$\left\| \frac{1}{\beta-\alpha} \mathbb{1}_{(\alpha,\beta]} + \lambda F^{-1} - k^* \right\|_2^2 = \sum_{j \in J} \int_{I_j} \left( \frac{1}{\beta-\alpha} \mathbb{1}_{(\alpha,\beta]}(u) + \lambda F^{-1}(u) - c_j \right)^2 du. \quad (B.14)$$

**Case 1:** If $I_j \cap (\alpha, \beta] = \emptyset$, then the $j$th summand in Equation (B.14) can be improved by choosing $k^*(u) = \min\{ \lambda F^{-1}(u), k^*(u_j^+) \}$, for all $u \in I_j$, and where $u_j^+$ denotes the right endpoint of $I_j$.

**Case 2:** If $I_j \subset [\alpha, \beta]$, then the $j$th summand in Equation (B.14) can be improved by setting $k^*(u) = \min\{ \frac{1}{\beta-\alpha} + \lambda F^{-1}(u), k^*(u_j^+) \}$ on $I_j$, and where $u_j^+$ denotes the right endpoint of $I_j$. Therefore, $|J| = 1$ and the isotonic projection $k^*$ is constant on one interval $I = (w_0, w_1]$, with $\alpha \leq w_0 \leq \beta \leq w_1$. Define $c = k^*(u)$, $u \in I$. Then, $c$ crosses $\lambda F^{-1}(u)$ at $u = w_1$ and, moreover, $c$ crosses $1/(\beta - \alpha) + \lambda F^{-1}(u)$ when $u = w_0$, provided that the jump $1/(\beta - \alpha)$ is not too large. More precisely, for a given $c$, $w_0$ and $w_1$ satisfy

$$\lambda F^{-1}(w_0) = \begin{cases} c - \frac{1}{\beta-\alpha}, & \text{if } \frac{1}{\beta-\alpha} \leq c - \lambda F^{-1}(\alpha), \\ \lambda F^{-1}(\alpha), & \text{otherwise,} \end{cases}$$

and $$\lambda F^{-1}(w_1) = \begin{cases} \lambda F^{-1}(1) & \text{if } c > F^{-1}(1), \\ c & \text{otherwise}, \end{cases}$$
The optimal $c$, which minimizes Equation (B.14), fulfills

$$
\arg\min_{w_0, w_1, c} \int_{w_0}^{w_1} \left( \frac{1}{\beta - \alpha} \mathbb{1}_{(\alpha, \beta]}(u) + \lambda F^{-1}(u) - c \right)^2 du
$$

$$
= \arg\min_{w_0, w_1, c} -2c \left( \frac{\beta - w_0}{\beta - \alpha} + \lambda \int_{w_0}^{\beta} F^{-1}(u) du \right) + c^2(w_1 - w_0).
$$

Thus, the optimal $c, w_0,$ and $w_1$ fulfill

$$
c = \frac{1}{w_1 - w_0} \frac{\beta - w_0}{\beta - \alpha} + \frac{\lambda}{w_1 - w_0} \int_{w_0}^{w_1} F^{-1}(u) du.
$$

\(\square\)

Proof of Proposition 4.4. The proof follows analogously to the proof of Proposition 4.3 and is omitted.\(\square\)

Proof of Corollary 4.5. For the worst-case value, let $G^{-1,*}$ denote the quantile function given in Proposition 4.3 with $\beta = 1$, $w_0 = 1$, and $w_1 = 1$. Direct calculations show that $G^{-1,*}$ indeed attains the supremum given in Corollary 3.2. For the best-case value, recall that TVaR$_\alpha = \lim_{\beta \to 1} RVaR_{\alpha, \beta}$ and that RVaR$_{\alpha, \beta}$ is an increasing function in $\beta$. Thus, by Theorem 3.12

$$
\inf_{G \in M_{\mu, \sigma}} TVaR_{\alpha} \geq \lim_{\beta \to 1} \inf_{G \in M_{\mu, \sigma}} RVaR_{\alpha, \beta} = \lim_{\beta \to 1} \left( \mu \frac{\sigma}{b_\beta} \left( b_\beta \mathbb{1}_{[U \in (\alpha, \beta)]} k_\beta^1(U) \right) - a_\beta \right),
$$

(B.15)

where $a_\beta = \mathbb{E}(k_\beta^1(U))$, $b_\beta = \text{std}(k_\beta^1(U))$, and $k_\beta^1$ is given in Proposition 4.4. Tedious calculations show that the limit in Equation (B.15) exists and is equal to TVaR$_\alpha(G^*)$, where $G^*$ has a quantile function given in Proposition 4.4 with $\beta = 1$.\(\square\)

Proof of Proposition 4.6. We first consider the worst-case of VaR$_\alpha^+$ and second that of VaR$_\alpha$.

To show the worst-case value of VaR$_\alpha^+$, we split the proof into three steps.

**Step 1:** We show that for any $\eta, \delta > 0$ and $m \in \mathbb{R}$, the unique quantile function attaining

$$
\sup_{d_W(G^{-1, \eta F^{-1} - m}) \leq \sqrt{\delta}} \text{VaR}_\alpha^+(G).
$$

(B.16)

is given by

$$
H_{\eta, \delta, m}^{\alpha,-1}(u) = (\eta F^{-1}(u) - m) + (c - \eta F^{-1}(u) + m) \mathbb{1}_{[U \in (\alpha, w_1)]},
$$

(B.17)

where $w_1 \in (0, 1]$ and $c \in \mathbb{R}$ are the unique solutions to

$$
d_{W}(G^{-1, \eta F^{-1} - m}) = \sqrt{\delta}.
$$

First, we show that the solution to Equation (B.16), denoted here by $G^*$, lies at the boundary of the Wasserstein ball, that is, $d_W(G^*, F^{-1}) = \sqrt{\delta}$. Assume by contradiction that $G^*$ lies in the interior of the Wasserstein ball, then, there exists a $\delta_0 > 0$ such that $d_W(G^* + \delta_0, F^{-1}) > \delta$ and $VaR_{\alpha}^+(G^* + \delta_0) > VaR_{\alpha}^+(G^* + \delta) > VaR_{\alpha}^+(G^*)$, which is a contradiction to the optimality of the solution. Next, we show that the optimal solution to Equation (B.16) has to be of the form (B.17). For this, assume that $G^*$ is a solution with $VaR_{\alpha}^+(G^*) > VaR_{\alpha}^+(H_{\eta, \delta, m}^{\alpha,-1}) = c$. By non-decreasingness of $G^*$, it holds that $G^{\alpha,-1}(u) > c \geq \eta F^{-1}(u) - m$, for all $u \in [\alpha, w_1]$. Thus, the
Wasserstein distance has a lower bound given by
\[
d_W(G^{\ast,-1}, \eta F^{-1} - m)^2 \geq \int_\alpha^{u_1} (G^{\ast,-1}(u) - \eta F^{-1}(u) + m)^2 \, du > \int_\alpha^{u_1} (c - \eta F^{-1}(u) + m)^2 \, du
\]
which is a contradiction to the fact that the optimal solution lies at the boundary of the Wasserstein ball. Uniqueness of the solution follows from the uniqueness of \( \delta \) and \( u_1 \).

Step 2: Denote by \( \eta^*, \delta^* \), and \( m^* \) the unique parameters such that \( H_{\eta^*,\delta^*,m^*}^{\ast,-1} \) has mean \( \mu \), variance \( \sigma^2 \), and Wasserstein distance \( d_W(H_{\eta^*,\delta^*,m^*}^{\ast,-1}, F) = \sqrt{\epsilon} \). These indeed exist as the three equations for the optimal \( \eta^*, \delta^* \), and \( m^* \) are not linearly dependent. Next, we claim that \( H_{\eta^*,\delta^*,m^*}^{\ast,-1} \) is the solution to our problem:

\[
\sup_{G \in \mathcal{M}(\mu, \sigma)} \text{VaR}^+_\alpha(G).
\] (B.18)

To show this, note that problem (B.18) is equivalent to

\[
\sup_{G \in \mathcal{M}(\mu, \sigma)} \text{VaR}^+_\alpha(G).
\] (B.19)

Indeed, a quantile function \( H^{-1} \) is admissible to Equation (B.18) if and only if it is admissible to Equation (B.19). This is because \( H^{-1} \) and \( H_{\eta^*,\delta^*,m^*}^{\ast,-1} \) have the same mean and variance, which implies that \( d_W(H^{-1}, F^{-1}) \leq d_W(H_{\eta^*,\delta^*,m^*}^{\ast,-1}, F^{-1}) = \sqrt{\epsilon} \) if and only if \( \text{corr}(H_{\eta^*,\delta^*,m^*}^{\ast,-1}(U), F^{-1}(U)) \leq \text{corr}(H^{-1}(U), F^{-1}(U)) \) and only if \( \text{corr}(H_{\eta^*,\delta^*,m^*}^{\ast,-1}(U), \eta^* F^{-1}(U) - m^*) \leq \text{corr}(H^{-1}(U), \eta^* F^{-1}(U) - m^*) \) if and only if \( d_W(H^{-1}, \eta^* F^{-1} - m^*) \leq d_W(H_{\eta^*,\delta^*,m^*}^{\ast,-1}, \eta^* F^{-1} - m^*) = \sqrt{\delta^*} \). However, \( H_{\eta^*,\delta^*,m^*}^{\ast,-1} \) is the unique solution to Equation (B.19) (because it solves Equation B.16 with \( \eta = \eta^*, m = m^*, \) and \( \delta = \delta^* \)). Hence, \( H^{-1} \) cannot be a solution to Equation (B.18) unless it coincides with \( H_{\eta^*,\delta^*,m^*}^{\ast,-1} \).

Step 3: Finally, we can verify that \( H_{\eta^*,\delta^*,m^*}^{\ast,-1} \) coincides with the solution as stated in the statements of the corollary. For this, note first that both \( H_{\eta^*,\delta^*,m^*}^{\ast,-1} \) and \( k_\lambda^1 \) are affine functions of \( F^{-1} \), and we only need to show that for every \( c_0 \leq \rho < 1 \), there exist \( \lambda > 0 \) such that \( \text{corr}(F^{-1}(U), k_\lambda^1(U)) = \rho \). Let us first consider the case in which \( \rho \) is close to one. First, assume that \( F^{-1} \) is constant on \((\alpha, u_1^\ast)\), where \( u_1^\ast = \sup\{\beta \in (\alpha, 1)|F^{-1}(\beta) = F^{-1}(\alpha)\} \), then we take \( w_1 = u_1^\ast \). By increasing \( \lambda \), we can bring \( k_\lambda^1 \) arbitrarily close to \( \lambda F^{-1} \) on \((\alpha, u_1^\ast)\), and \( k_\lambda^1 \) will coincide with \( \lambda F^{-1} \) on \((w_1^\ast, 1] \) as well as on \((0, \alpha] \). Second, assume that \( F^{-1} \) is not constant on the right of \( \alpha \), then for any \( w_1 > \alpha \) close to \( \alpha \), we can find \( \lambda > 0 \) such that \( k_\lambda^1 \) will coincide with \( \lambda F^{-1} \) on \((w_1, 1] \) as well as on \((0, \alpha] \). Hence, for any value of \( \rho \) close to 1, it is possible to specify \( k_\lambda^1 \) such that \( \text{corr}(F^{-1}(U), k_\lambda^1(U)) = \rho \).

When \( \rho = c_0 \), we can take \( \lambda = 0 \), that is, \( k_0^1(u) = c_1[u \in (\alpha, 1]] \) and \( \text{corr}(F^{-1}(U), k_0^1(U)) = c_0 \). By continuity of \( \text{corr}(F^{-1}(U), k_\lambda^1(U)) \), it thus follows that for all \( c_0 \leq \rho < 1 \), there exist \( \lambda > 0 \) such that \( \text{corr}(F^{-1}(U), k_\lambda^1(U)) = \rho \).

Next, we consider the worst-case \( \text{VaR}_\alpha \). Clearly, the worst-case values of \( \text{VaR}_\alpha \) and \( \text{VaR}_\alpha^+ \) are equal. The worst-case value of \( \text{VaR}_\alpha \) however, is not attained. To see this, assume by contradiction that the worst-case \( \text{VaR}_\alpha \) is attained by \( G^\ast \). Then, there exists \( \delta_0 > 0 \) such that \( G^{\ast,-1}(u) >
Bernard et al.\] $\eta F^{-1}(u) - m$, for all $u \in (\alpha - \delta, \alpha]$. This, however, implies that $d_W(G^*, F) > d_W(\hat{G}, F) = \sqrt{\delta}$, where $\hat{G}$ attains the worst-case VaR $\text{VaR}^\alpha$.

The proof of the best-case values of $\text{VaR}_\alpha$ and $\text{VaR}^+_\alpha$ follows similar steps to the worst-case values. Note that the best-case $\text{VaR}_\alpha$ is attained while the best-case $\text{VaR}^+_\alpha$ is not. \hfill $\square$

**B.5 | Proof of Theorem 4.8**

**Lemma B.6.** Optimization problem (18) can be equivalently written as

$$ \sup_{G \in \mathcal{M}_\varepsilon} H_g(G) = \sup_{(\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 \leq \varepsilon, \sigma > 0} \mu + \sigma \text{std}(\gamma(U)) \text{corr}(\gamma(U), \gamma(U) + \lambda F^{-1}(U)), $$

where $\lambda$ is given in Theorem 3.1.

**Proof of Lemma B.6.** First, we define the set

$$ \mathcal{N}_\varepsilon = \{ G \in \mathcal{M}_\varepsilon(\mu, \sigma) | (\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 \leq \varepsilon, \sigma > 0 \} $$

and show that $\mathcal{N}_\varepsilon = \mathcal{M}_\varepsilon$. Clearly, if $G \in \mathcal{N}_\varepsilon$, then $G \in \mathcal{M}_\varepsilon$. Conversely, let $G \in \mathcal{M}_\varepsilon$ with mean $\mu_G$ and standard deviation $\sigma_G > 0$, then, the Wasserstein distance between $F$ and $G$ fulfills $(\mu_F - \mu_G)^2 + (\sigma_F - \sigma_G)^2 \leq d_W(F, G)^2 \leq \varepsilon$. Thus, $G \in \mathcal{N}_\varepsilon$ and we indeed obtain that $\mathcal{M}_\varepsilon = \mathcal{N}_\varepsilon$.

Applying Theorem 3.1 case 1 concludes the proof. \hfill $\square$

**Proof of Theorem 4.8.** By Lemma B.6, we only have to solve

$$ \sup_{(\mu_F - \mu)^2 + (\sigma_F - \sigma)^2 \leq \varepsilon, \sigma > 0} \Theta(\mu, \sigma), \quad \text{where} \quad \Theta(\mu, \sigma) = \mu + \sigma \text{std}(\gamma(U)) \text{corr}(\gamma(U), \gamma(U) + \lambda F^{-1}(U)). $$

We can rewrite $\Theta$ as follows, using the expression of $\lambda$ in Equation (B.2):

$$ \Theta(\mu, \sigma) = \mu + \frac{\sigma}{b_\lambda} \left( V + \frac{\sqrt{\Delta - C_{\gamma,F}}}{\sigma_F^2} C_{\gamma,F} \right), $$

where $b_\lambda$ can be simplified as $b_\lambda = \frac{\sigma \sqrt{\Delta}}{K}$ using Equations (B.2) and (B.6). Thus,

$$ \Theta(\mu, \sigma) = \mu + \frac{K \left( V \sigma_F^2 + C_{\gamma,F} \sqrt{\Delta - C_{\gamma,F}^2} \right)}{\sqrt{\Delta} \sigma_F^2}. $$

Furthermore, recall that $\Delta = \frac{K^2 (C_{\gamma,F}^2 - V \sigma_F^2)}{K^2 - \sigma^2 \sigma_F^2}$ and that $K \geq 0$ since $K$ is the covariance between two comonotonic variables in Equation (B.5). We, therefore, obtain that

$$ \Theta(\mu, \sigma) = \mu + \frac{C_{\gamma,F} K}{\sigma_F^2} + \sqrt{V \frac{C_{\gamma,F}^2}{\sigma_F^2}} \sqrt{\sigma^2 - \frac{K^2}{\sigma_F^2}}. $$
Observe that from the expression of $K$ in Equation (B.3), $K = \frac{\mu^2 + \sigma_F^2 + \mu^2 + \sigma_F^2 - 2\mu\mu_F - \varepsilon}{2}$, we have

$$\frac{\partial K}{\partial \mu} = \mu - \mu_F, \quad \frac{\partial K}{\partial \sigma^2} = \frac{1}{2}, \quad \text{and} \quad \frac{\partial \Theta(\mu, \sigma)}{\partial K} = C_{\gamma,F} \frac{K\sqrt{V - \frac{C_{\gamma,F}^2}{\sigma_F^2}}}{\sigma_F^2 \sqrt{\sigma^2 - \frac{K^2}{\sigma_F^2}}}.$$  

Thus, when differentiating $\Theta$ with respect to $\mu$ interpreted as a function of $(\mu, K(\mu))$, we obtain

$$\frac{\partial \Theta(\mu, \sigma)}{\partial \mu} = 1 + \left( \frac{C_{\gamma,F}}{\sigma_F^2} - \frac{K\sqrt{V - \frac{C_{\gamma,F}^2}{\sigma_F^2}}}{\sigma_F^2 \sqrt{\sigma^2 - \frac{K^2}{\sigma_F^2}}} \right)(\mu - \mu_F), \quad (B.20)$$

and when differentiating $\Theta$ with respect to $\sigma^2$, we obtain

$$\frac{\partial \Theta(\mu, \sigma)}{\partial \sigma^2} = \left( \frac{1}{2} - \frac{K}{2\sigma_F^2} \right) \frac{\sqrt{V - \frac{C_{\gamma,F}^2}{\sigma_F^2}}}{\sqrt{\sigma^2 - \frac{K^2}{\sigma_F^2}}} + \frac{C_{\gamma,F}}{2\sigma_F^2} \sqrt{\frac{\varepsilon - (\mu - \mu_F)^2}{V}} \quad (B.21)$$

To solve $\frac{\partial \Theta(\mu, \sigma)}{\partial \sigma^2} = 0$, we first use that $\sigma^2 = 2K + \varepsilon - (\mu - \mu_F)^2$ and then square the equation we aim to solve. There are two roots for $K$ characterized by $K = \sigma_F^2 \pm \frac{C_{\gamma,F}}{\sqrt{V}} \sqrt{\varepsilon - (\mu - \mu_F)^2}$. But only one solves the original equation (B.21), which is given by

$$K = \sigma_F^2 + \frac{C_{\gamma,F}}{\sqrt{V}} \sqrt{\varepsilon - (\mu - \mu_F)^2}, \quad (B.22)$$

which then corresponds to $\sigma^2 = \varepsilon - (\mu - \mu_F)^2 + \sigma_F^2 + \frac{2C_{\gamma,F}}{\sqrt{V}} \sqrt{\varepsilon - (\mu - \mu_F)^2}$. From the fact that $\frac{\partial \Theta(\mu, \sigma)}{\partial \sigma^2} = 0$ in Equation (B.21), we have that

$$\sqrt{V - \frac{C_{\gamma,F}^2}{\sigma_F^2}} = \frac{C_{\gamma,F}}{K - \sigma_F^2} \sqrt{\sigma^2 - \frac{K^2}{\sigma_F^2}}.$$

Replacing this expression in Equation (B.20), we find that $\frac{\partial \Theta(\mu, \sigma)}{\partial \mu} = 0$ is equivalent to $C_{\gamma,F}(\mu_F - \mu) + K - \sigma_F^2 = 0$. Using the expression of $K$ in Equation (B.22), this gives $\sqrt{\varepsilon - (\mu - \mu_F)^2} = (\mu - \mu_F)\sqrt{V}$, which can be solved explicitly and has a unique root:

$$\mu = \mu_F + \sqrt{\frac{\varepsilon}{1 + V}}.$$
Recall the equation \( \sigma^2 = \varepsilon - (\mu - \mu_F)^2 + \sigma_F^2 + \frac{2\gamma \varepsilon}{\sqrt{V}} \sqrt{\varepsilon - (\mu - \mu_F)^2} \), which can be simplified to the unique point

\[
(\mu^*, \sigma^*) = \left( \mu_F + \sqrt{\frac{\varepsilon}{1 + V}}, \sqrt{\frac{\sigma_F^2}{1 + V} + \frac{2\gamma \varepsilon}{1 + V}} \right),
\]

in which \( \frac{\partial \Theta(\mu^*, \sigma^*)}{\partial \sigma} = \frac{\partial \Theta(\mu^*, \sigma^*)}{\partial \mu} = 0 \). Next, we verify that the Hessian matrix is negative-definite at this point (negative diagonal coefficients and positive determinant), which ensures that it is a global maximum.

In fact, the Hessian matrix at any point \((\mu, \sigma)\) can be written as

\[
\begin{pmatrix}
\frac{\partial^2 \Theta(\mu, \sigma)}{\partial \mu^2} & \frac{\partial^2 \Theta(\mu, \sigma)}{\partial \sigma^2} \\
\frac{\partial^2 \Theta(\mu, \sigma)}{\partial \mu \partial \sigma} & \frac{\partial^2 \Theta(\mu, \sigma)}{\partial \sigma^2}
\end{pmatrix} = \begin{pmatrix}
\frac{C_{\gamma F}}{\sigma_F^2} - M \frac{(\mu - \mu_F)^2}{\sigma_F^2} & \frac{M}{\sigma_F^2} \frac{(\mu - \mu_F)^2}{\sigma_F^2 - K^2} \\
\frac{M}{\sigma_F^2} \frac{(\mu - \mu_F)^2}{\sigma_F^2 - K^2} & \frac{M}{\sigma_F^2} \frac{(\mu - \mu_F)^2}{\sigma_F^2 - K^2}
\end{pmatrix},
\]

where \( M = \sqrt{V - \frac{c^2_{\gamma F}}{\sigma_F^2}}/\sqrt{\sigma^2 - K^2} \). Evaluating this Hessian matrix at \((\mu^*, (\sigma^*)^2)\), it can be simplified to

\[
\begin{pmatrix}
\frac{\partial^2 \Theta(\mu, \sigma)}{\partial \mu^2} & \frac{\partial^2 \Theta(\mu, \sigma)}{\partial \sigma^2} \\
\frac{\partial^2 \Theta(\mu, \sigma)}{\partial \mu \partial \sigma} & \frac{\partial^2 \Theta(\mu, \sigma)}{\partial \sigma^2}
\end{pmatrix} = \begin{pmatrix}
K(\sigma^*)^2 & V \sqrt{\varepsilon} \sigma_F^2 (\sigma^*)^2 - K^2 \\
V \sqrt{\varepsilon} \sigma_F^2 (\sigma^*)^2 - K^2 & V \sqrt{\varepsilon} \sigma_F^2 (\sigma^*)^2 - K^2
\end{pmatrix}.
\]

The diagonal coefficients are clearly negative and the determinant

\[
\frac{V}{4(\sigma_F^2 (\sigma^*)^2 - K^2)} + \frac{\varepsilon (V \sigma_F^2 - C_{\gamma F}^2)}{4(\sigma_F^2 (\sigma^*)^2 - K^2)^2 (1 + V)}
\]

is positive, which ensures that \( \Theta(\mu, \sigma) \) is locally concave at \((\mu^*, (\sigma^*)^2)\), which completes the proof.

Proof of Proposition 5.1. The general projection property (Theorem 1) in Popescu (2007) shows that for every portfolio \( \mathbf{x} \in \mathcal{P} \) and every distribution function \( G_x \) having mean \(-\mu_x = -\mathbf{x}^T \mu\) and variance \( \sigma_x^2 = \mathbf{x}^T \Sigma \mathbf{x} \) there exists a multivariate distribution function \( G_R \in \mathcal{M}(\mu, \Sigma) \) such that the portfolio loss \(-\mathbf{x}^T \mathbf{R}\) is distributed with \( G_x \). As a consequence, optimization (20) is equivalent to

\[
\min_{\mathbf{x} \in \mathcal{P}} \quad \max_{\gamma \in M(\mathbf{x}^*, \mu, \sigma)} H_\gamma(G_x) \tag{B.23}
\]

where \( M(-\mu_x, \sigma_x) \) denotes the set of univariate distributions with mean \(-\mu_x\) and standard deviation \( \sigma_x \). Applying Theorem 3.1 yields that Equation (B.23) is equivalent to the minimization problem

\[
\min_{\mathbf{x} \in \mathcal{P}} \left\{ -\mu_x + \frac{\sigma_x}{b_{\lambda_x}} (V + \lambda_x C_x) \right\} \tag{B.24}
\]
in which \( \lambda_x = \frac{K_x}{\sigma_x^2} \sqrt{\frac{\sigma_x^2 - C_x^2}{\sigma_x^2 - K_x^2}} - \frac{C_x}{\sigma_x^2} \) with \( C_x = \text{cov}(y(U), F^{-1}(U)) \), \( K_x = \sigma_x^2 - \frac{\epsilon_x}{2} \), and \( b_{\lambda_x} = \sigma_x \sqrt{\frac{\sigma_x^2 - C_x^2}{\sigma_x^2 - K_x^2}} \), where we used Equation (B.6) to obtain the expression for \( b_{\lambda_x} \). Note that \( C_x = \sigma_x \tilde{C} \), where \( \tilde{C} = \text{cov}(y(U), F^{-1}(U)) \). Then, minimization (B.24) becomes

\[
\min_{x \in \mathcal{P}} \left\{ -\mu_x + \frac{V \sqrt{\sigma_x^4 - K_x^2}}{\sigma_x V - \tilde{C}^2} + \frac{\sqrt{\sigma_x^4 - K_x^2}}{\sigma_x V - \tilde{C}^2} \left( \frac{K_x \sqrt{V - \tilde{C}^2}}{\sigma_x^2} \right) \right\},
\]

which reduces to

\[
\min_{x \in \mathcal{P}} \left\{ -\mu_x + \frac{V \sqrt{\sigma_x^4 - K_x^2}}{\sigma_x V - \tilde{C}^2} + \frac{K_x \tilde{C}}{\sigma_x} - \frac{\sqrt{\sigma_x^4 - K_x^2}}{\sigma_x V - \tilde{C}^2} \right\}.
\]

Substituting \( K_x = \sigma_x^2 - \frac{\epsilon_x}{2} \) and noting that \( \tilde{C} = \sqrt{\mathcal{C}_0} \) concludes the proof.

APPENDIX C: APPLICATION TO A PORTFOLIO OF RISKS

Here, we collect the models used in Section 5.2. We provide the first two moments, which are used to calculate the models’ parameters by matching the first two moments to those of the reference Pareto–Clayton model.

1. **Lognormal model (LN)**: We write \( S = e^Z \), where \( Z \) is a normally distributed random variable with mean \( \mu \) and variance \( \sigma^2 \). The first two moments of \( S \) are \( \mathbb{E}(S) = e^{\mu + \frac{1}{2}\sigma^2} \) and \( \mathbb{E}(S^2) = e^{2\mu + 2\sigma^2} \).

2. **Gamma model (Γ)**: The portfolio loss \( S \) is Gamma distributed with parameters \( \alpha, \beta > 0 \), \( \mathbb{E}(S) = \alpha \beta \), \( \mathbb{E}(S^2) = (1+\alpha)\alpha \beta^2 \), and density \( f_S(s) = \frac{\beta^\alpha s^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta s} \), \( s > 0 \), where \( \Gamma(\cdot) \) denotes the Gamma function.

3. **Weibull model (W)**: The portfolio loss \( S \) follows a Weibull distribution with parameters \( \lambda, k \in (0, +\infty) \), distribution function \( F_S(s) = 1 - e^{-(s/\lambda)^k} \), \( s > 0 \), \( \mathbb{E}(S) = \lambda \Gamma(1 + 1/k) \), and \( \mathbb{E}(S^2) = \lambda^2 \Gamma(1 + 2/k) \).

4. **Inverse Gaussian model (IG)**: The distribution of \( S \) is Inverse Gaussian with parameters \( \mu, \lambda > 0 \), distribution function \( F_S(s) = \Phi(\sqrt{\frac{1}{s} \left( \frac{1}{\mu} - 1 \right)} ) + e^{\frac{2\lambda}{\mu}} \Phi(-\sqrt{\frac{4}{s} \left( \frac{1}{\mu} + 1 \right)} ) \), \( s > 0 \), \( \mathbb{E}(S) = \mu \), and \( \mathbb{E}(S^2) = \mu^2 + \mu^2 \).

5. **Inverse Gamma model (IG)**: The portfolio loss \( S \) follows an Inverse Gamma distribution with parameters \( \alpha, \beta > 0 \), density function \( f_S(s) = \frac{\beta^\alpha (1/s)^{\alpha+1} e^{-\beta/s}}{\Gamma(\alpha)} \), \( s > 0 \), \( \mathbb{E}(S) = \frac{\beta}{\alpha-1} \), and \( \mathbb{E}(S^2) = \frac{\beta^2}{(\alpha-1)(\alpha-2)} \).

6. **Inverse Weibull model (IW)**: The distribution function of \( S \) is Inverse Weibull with parameters \( \lambda, k > 0 \), and density \( f_S(s) = e^{-\lambda(s/k)^k} \), \( s > 0 \), \( \mathbb{E}(S) = \lambda \Gamma(1 - 1/k) \), and \( \mathbb{E}(S^2) = \lambda^2 \Gamma(1 - 2/k) \).
7. **Log-Logistic model (LL):** The portfolio loss $S$ follows a Log-Logistic distribution with parameters $\alpha, \beta > 0$, quantile function $F^{-1}_S(s) = \alpha\left(\frac{p}{1-p}\right)^{1/\beta}$, $s > 0$, $\mathbb{E}(S) = \frac{\alpha\pi/\beta}{\sin(\pi/\beta)}$, and $\mathbb{E}(S^2) = \frac{\alpha^2\pi^2/\beta}{\sin(2\pi/\beta)}$. 