Constructing Uncertainty Sets for Robust Risk Measures: A Composition of ϕ-Divergences Approach to Combat Tail Uncertainty

Guanyu Jin[a], Roger J. A. Laeven[a,c,d], Dick den Hertog[b], and Aharon Ben-Tal[e]

[a]Dept. of Quantitative Economics, University of Amsterdam, 1001 NJ Amsterdam, The Netherlands
[b]Dept. of Business Analytics, University of Amsterdam, 1001 NJ Amsterdam, The Netherlands
[c]EURANDOM, 5600 MB Eindhoven, The Netherlands
[d]CentER, Tilburg University, 5000 LE Tilburg, The Netherlands
[e]Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology

May 5, 2025

Abstract

Risk measures, which typically evaluate the impact of extreme losses, are highly sensitive to misspecification in the tails. This paper studies a robust optimization approach to combat tail uncertainty by proposing a unifying framework to construct uncertainty sets for a broad class of risk measures, given a specified nominal model. Our framework is based on a parametrization of robust risk measures using two (or multiple) ϕ-divergence functions, which enables us to provide uncertainty sets that are tailored to both the sensitivity of each risk measure to tail losses and the tail behavior of the nominal distribution. In addition, our formulation allows for a tractable computation of robust risk measures, and elicitation of ϕ-divergences that describe a decision maker's risk and ambiguity preferences.

Keywords: Distributionally robust optimization, Risk measures, Optimized certainty equivalent, Uncertainty sets

1 Introduction

Model misspecification, especially in the tails, is an inevitable issue that underlies the practice of financial and operational risk management. One way to safeguard against model uncertainty is distributionally robust optimization (DRO). The idea is to take a set of plausible models, often referred to as the uncertainty/ambiguity set, and evaluate the worst-case risk among all these models. DRO has its roots in the classical robust optimization pioneered by Ben-Tal et al. [2009], as well as in economics and decision theory, by the works of Gilboa and Schmeidler [1989], Maccheroni et al. [2006], and Hansen and Sargent [2008].

An essential part of distributional robust optimization is to choose a proper uncertainty set. In data-driven optimization, DRO is often used to address sampling errors of empirical risk minimization, where uncertainty sets are calibrated to offer certain statistical guarantees. For this line of research, we refer the readers to works such as Ben-Tal et al. [2013], Esfahani and Kuhn [2018], Lam [2019], Duchi et al. [2021], van Parys et al. [2021]. In risk management, such as the calculation of capital requirement, one typically works with a pre-specified nominal model (e.g., an internal model proposed by the regulator) instead of the empirical distribution. In this case, the goal is to construct uncertainty sets that are sufficiently large to protect against model misspecification.

In much of the DRO literature, the optimization problem is studied in a risk-neutral environment, where a random loss is evaluated under its expected value. In finance, insurance, and economics, it is more natural to model a decision maker's preferences using a risk measure that is nonlinear in the probability, since humans typically do not perceive changes in probabilities linearly, especially for extreme events. This complicates the choice of an uncertainty set, since in some cases the linearity assumption can be crucial for deriving a tractable reformulation of the primal robust problem. Moreover, the sensitivity of a risk measure to the tails of a distribution also determines the specification of an uncertainty set. For example, an evaluation of the entropic risk measure is only finite for light-tailed distributions. This implies that any uncertainty set containing a heavy-tailed distribution is inadequate for the entropic risk measure. Thus, the natural question arises: given a risk measure and a pre-specified nominal model, how should one properly specify an uncertainty set that is not too conservative (and not too restrictive), and that in addition admits a tractable reformulation?

In this paper, we provide an answer to the above question by proposing a simple, yet unifying framework to specify uncertainty sets for a broad class of risk measures, based on ϕ-divergences. Introduced by Csiszár [1975], ϕ-divergences are statistical measures for probability models that have received much attention in the DRO literature since the seminal work of Ben-Tal et al. [2013]. Uncertainty sets that are defined as a ball around a nominal model, measured using a ϕ-divergence, have been extensively studied by Kruse et al. [2019, 2021] in the context of model risk assessment. Interestingly, ϕ-divergences are also strongly connected to risk measures. As shown by Ben-Tal et al. [1991], Ben-Tal and Teboulle [2007], the four major classes of risk measures: Expected utility, optimized certainty equivalent, shortfall risk measure (a.k.a. u-Mean certainty equivalent), and distortion risk measures, all admit a robust representation with a ϕ-divergence penalty function.

Our main contributions can be summarized as follows: by utilizing the pivotal role of ϕ-divergence in both robust optimization and risk theory, we characterize a wide range of distributionally robust risk measures in terms of two ϕ-divergences (ϕ_1, ϕ_2): a ϕ_1-divergence that specifies the risk measure through an *inner* robust representation, and a ϕ_2-divergence that controls an *outer* ambiguity set to address model uncertainty. We call this the *composition* approach, since the parametrization with two ϕ-divergences is formulated as a composite robust optimization problem with two layers of uncertainty. This approach allows us to translate the problem of choosing an uncertainty set for a risk measure into the task of specifying a ϕ_2 function given ϕ_1. By deriving a tractable reformulation of the composite robust problem, we show that our (ϕ_1, ϕ_2) characterization is not only computationally tractable, but also provides a blueprint for how to calibrate the ϕ_2-divergence uncertainty set to address tail uncertainty, given a risk measure and a nominal model. For risk measures such as the Conditional Value-at-Risk and the entropic risk measure, we provide examples showing the inadequacy of standard divergences to evaluate model risk for certain nominal models, and provide explicit construction of new ϕ_2-divergences that address these shortcomings, using our suggested framework. In addition, we show that our composition representation also offers other advantages, such as the elicitation of divergences (ϕ_1, ϕ_2) through queries on decision makers, and natural extension to include a higher level of uncertainty using *globalized robust optimization* (as introduced in Ben-Tal et al. 2017), that also offers more flexibility in the construction of uncertainty sets. These results serve as extra tools for selecting proper divergences that are customized to a decision maker's risk and ambiguity preferences.

Finally, we give a brief review of other related literature. Lam and Mottet [2017] studied the robust optimization approach to address tail uncertainty by constructing uncertainty sets that impose shape constraints on probability densities. However, the tractability of their formulation relies on an explicit representation of the risk measure in terms of density functions, which are not always obtainable for risk measures that are outside the expected utility framework. On the other hand, the tractability of our composite divergences formulation can be easily achieved via convex duality. Breuer and Csiszár [2016] used ϕ-divergences to measure model risk and provided a tractable reformulation of the corresponding worst-case expectation problem. Building upon this work, Kruse et al. [2019, 2021] demonstrated how specific ϕ-divergences can be designed to control the tail behaviors of the distributions in the divergence ball. Our work extends and unifies the results of Breuer and Csiszár [2016] and Kruse et al. [2019, 2021] to risk measures that are not simply expected value, such as the optimized certainty equivalent and shortfall risk measures. Glasserman and Xu [2014], Schneider and Schweizer [2015] have used the Kullback-Leibler divergence to measure model risk for many financial problems such as option hedging and portfolio credit risk management. Bertsimas and Brown [2009] provided a method for constructing uncertainty sets using coherent risk measures. However, while this approach is theoretically sound, it requires the specification of a decision maker's individual preference for risk, which has to be elicited. Robust risk measures with uncertainty sets that are defined using the Wasserstein distance have also been studied by works such as Bartl et al. [2020] and Bernard et al. [2023], where the focus is more on tractability than calibration of uncertainty sets.

The remaining parts of the paper are organized as follows: Section 2 formalizes our two divergences approach in more detail. Section 3 and 4 establish the conditions that we impose on ϕ_2 for different types of risk measures and provide explicit examples of the construction of new divergences, tailored to a given risk measure and nominal model. Section 5 shows how to elicit divergences from robust risk measures, and Section 6 discusses the computational aspects of robust risk measures, as well as an extension of our framework using globalized robust optimization to address more uncertainty. Numerical examples of robust options hedging and inventory planning problems are displayed in Section 7. A concluding remark is given in Section 8. Furthermore, all proofs and additional technical details are contained in an Electronic Companion.

2 ϕ-Divergences and Risk Measures

Let Φ_0 denote the set of all non-negative functions $\phi : \mathbb{R} \to [0, \infty)$ that are convex and normalized, i.e., $\phi(1) = 0$, with an effective domain $\mathrm{dom}(\phi) \triangleq \{t \in \mathbb{R} \mid \phi(t) < \infty\} \subset [0, \infty)$ which also contains a neighborhood around 1. The ϕ-divergence between any two measures $\mathbb{P}$ and $\mathbb{Q}$ on an event space $(\Omega, \mathcal{F})$ is defined as

$$I_\phi(\mathbb{Q}, \mathbb{P}) \triangleq \begin{cases} \int_\Omega \phi\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right) \mathrm{d}\mathbb{P} & \text{if } \mathbb{Q} \ll \mathbb{P} \\ +\infty & \text{else,} \end{cases} \tag{1}$$

where $\mathbb{Q} \ll \mathbb{P}$ denotes absolute continuity with respect to $\mathbb{P}$ (i.e., $\mathbb{P}(A) = 0 \Rightarrow \mathbb{Q}(A) = 0, \forall A \in \mathcal{F}$), and $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ denotes the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$. Furthermore, we adopt the conventions: $\phi(0) \triangleq \lim_{t \downarrow 0} \phi(t)$, $0\phi\,(0/0) = 0$, $0\phi\,(a/0) = \lim_{\epsilon \downarrow 0} \epsilon\phi\,(a/\epsilon) = a \lim_{t\to\infty} \phi(t)/t,\ a > 0$. We note that for any probability $\mathbb{Q}, \mathbb{P}$, we have $I_\phi(\mathbb{Q}, \mathbb{P}) \geq 0$ and $I_\phi(\mathbb{P}, \mathbb{P}) = 0$. If $\mathbb{P}$ and $\mathbb{Q}$ have density functions f, g, respectively, with respect to some σ-finite measure $\mathrm{d}x$ on $\mathbb{R}$,[1] then (1) can be expressed more explicitly in terms of real-valued density functions:

$$I_\phi(g, f) = \int_{\mathbb{R}} \phi\left(\frac{g(x)}{f(x)}\right) f(x)\mathrm{d}x.$$

Let $X : \Omega \to \mathbb{R}$ be a random payoff.[2] As shown in Ben-Tal and Teboulle [1987, 2007] and Ben-Tal et al. [1991], many types of risk measure/certainty equivalents admit a robust representation with a ϕ-divergence penalization. This is a consequence of the dual relationship between a utility function and a ϕ-divergence function through the Legendre-Fenchel transformation (a.k.a. convex/concave conjugate: for any ψ, its convex and concave conjugate are respectively $\psi^*(\mathbf{y}) \triangleq \sup_{\mathbf{x}\in\mathbb{R}^d}\{\mathbf{y}^T\mathbf{x} - \psi(\mathbf{x})\}$ and $\psi_*(\mathbf{y}) \triangleq -\psi^*(-\mathbf{y})$). For example, the optimized certainty equivalents (OCE) defined with respect to a probability measure $\mathbb{P}$ and a utility function u, is equal to the following robust problem where all probability measures $\mathbb{Q} \ll \mathbb{P}$ are penalized by $\phi = -u_*$:

$$\begin{aligned} \rho_{\mathrm{oce},\mathbb{P}}(X) &\triangleq \sup_{\{\mathbb{Q}:\mathbb{Q}\ll\mathbb{P}\}} \mathbb{E}_\mathbb{Q}[-X] - I_\phi(\mathbb{Q}, \mathbb{P}) \\ &= \inf_{\eta\in\mathbb{R}} \{\eta - \mathbb{E}_\mathbb{P}[u(X + \eta)]\}. \end{aligned} \tag{2}$$

Therefore, the OCE is also known as the *divergence risk measure*, which includes examples such as:

- Conditional Value-at-Risk (CVaR_α) at $\alpha \in (0, 1)$, where $\phi(t) = \delta_{[0,1/(1-\alpha)]}(t)$ is the indicator function on $[0, 1/(1-\alpha)]$ [3].
- Entropic risk measure $\rho_{e,\gamma}(X) = \log\left(\mathbb{E}[\exp(-\gamma X)]\right)/\gamma$, $\gamma > 0$, where $\phi(t) = (t\log(t) - t + 1)/\gamma$, for $t \geq 0$.

The class of utility-based shortfall risk measures, a.k.a. u-Mean certainty equivalents, can also

[1] Such as the Lebesgue measure or the counting measure or a combination of both

[2] Negative payoffs are considered as losses

[3] For a set $\mathcal{C} \in \mathbb{R}^n$, the indicator function is defined as $\delta_\mathcal{C}(x) = \begin{cases} 0 & x \in \mathcal{C} \\ +\infty & x \notin \mathcal{C}. \end{cases}$

be represented by a ϕ-divergence:

$$\begin{aligned}\rho_{\mathrm{sf},\mathbb{P}}(X) &\triangleq \sup_{\lambda>0}\sup_{\{\mathbb{Q}:\mathbb{Q}\ll\mathbb{P}\}} \mathbb{E}_{\mathbb{Q}}[-X] - I_{\phi^\lambda}(\mathbb{Q},\mathbb{P}) \\ &= \inf_{\eta\in\mathbb{R}}\{\eta \mid \mathbb{E}_{\mathbb{P}}[-u(X+\eta)] \leq 0\},\end{aligned} \tag{3}$$

where $\phi^\lambda(t) = \lambda\phi(t/\lambda)$ and $\phi(t) = -u_*(t)$.

For expected utility, the ϕ-divergence robust representation is a maximization over all finite measures, instead of only probability measures:

$$\mathbb{E}_{\mathbb{P}}[-u(X)] = \sup_{\{\mathbb{Q}:\mathbb{Q}\ll\mathbb{P},\mathbb{Q}(\Omega)<\infty\}} \mathbb{E}_{\mathbb{Q}}[-X] - I_\phi(\mathbb{Q},\mathbb{P}), \tag{4}$$

where again $\phi = -u_*$.

Therefore, ϕ-divergences are basic building blocks for expected utility, optimized certainty equivalents and utility-based shortfall risk measures. Using a second ϕ_2-divergence, we can thus parametrize robust risk measures of these classes by a composition of ϕ-divergences. We can address ambiguity in two ways: Either by means of the *multiple prior* approach via an uncertainty set (which we denote using ρ^s), or by means of the *multiplier preference* approach through penalization (which we denote using ρ^l). In the case of OCE risk measures, this leads to the following composite robust formulations, where the maximization is taken over two variables $(\mathbb{Q},\bar{\mathbb{Q}})$:

$$\begin{aligned}\rho^s_{\mathrm{oce},\mathbb{P}_0}(X) \triangleq \sup_{\{\mathbb{Q}:I_{\phi_2}(\mathbb{Q},\mathbb{P}_0)\leq r\}}\ &\sup_{\{\bar{\mathbb{Q}}:\bar{\mathbb{Q}}\ll\mathbb{Q}\}} \mathbb{E}_{\bar{\mathbb{Q}}}[-X] \\ &- I_{\phi_1}(\bar{\mathbb{Q}},\mathbb{Q}),\end{aligned} \tag{5}$$

and

$$\begin{aligned}\rho^l_{\mathrm{oce},\mathbb{P}_0}(X) \triangleq \sup_{\{\mathbb{Q}:\mathbb{Q}\ll\mathbb{P}_0\}}\ &\sup_{\{\bar{\mathbb{Q}}:\bar{\mathbb{Q}}\ll\mathbb{Q}\}} \mathbb{E}_{\bar{\mathbb{Q}}}[-X] \\ &- I_{\phi_1}(\bar{\mathbb{Q}},\mathbb{Q}) - I_{\phi_2}(\mathbb{Q},\mathbb{P}_0).\end{aligned} \tag{6}$$

Other classes of robust risk measures are defined similarly. We note that for any constant $\lambda > 0$, we have $\lambda I_\phi(\mathbb{Q},\mathbb{P}) = I_{\lambda\phi}(\mathbb{Q},\mathbb{P})$. Hence, the representations in (5) and (6) also include penalization constants that are not equal to one. Throughout this paper, we fix $\mathbb{P}_0$ as a *given* nominal distribution. $\mathbb{P}_0$ may be a distribution that is already specified in an economic model, or a distribution calibrated from a parametric family. Furthermore, we assume that $X \in L^1(\mathbb{P}_0)$ (i.e., $\int |X| \mathrm{d}\mathbb{P}_0 < \infty$). In the following section, we extensively study the robust OCE risk measures as the prime example of this paper. We show that by reformulating the problems (5) and (6) into finite-dimensional dual problems, we can obtain a blueprint for constructing ϕ_2 that is specifically tailored to a given risk measure ϕ_1 and the nominal model $\mathbb{P}_0$.

3 Divergence Choices for Robust OCE Risk Measures

To illustrate the importance of choosing a proper ϕ-divergence for a robust risk measure, consider the divergence function associated with the CVaR_α risk measure as defined in (2), where $\phi_\alpha(t) = \delta_{[0,1/(1-\alpha)]}(t)$, for some $\alpha \in (0,1)$. Then, any alternative distributions with a density function g that has a heavier tail than the nominal density f_0 (i.e., $\lim_{x\to\infty} g(x)/f_0(x) = \infty$) will have an infinite divergence value with respect to ϕ_α. Indeed, the likelihood ratio $g(x)/f_0(x)$ exceeds $1/(1-\alpha)$ for sufficiently large values of x, eventually lying outside the effective domain of ϕ_α. Therefore, the ϕ-divergence uncertainty set induced by ϕ_α does not include any distributions that have a heavier right-side tail than the nominal distribution, while heavy-tailed losses are often the real threats in financial risk management that one wishes to be robust against.

The choice of ϕ-divergence also depends on the sensitivity of a risk measure to tail losses. For example, the entropic risk measure $\rho_{e,1}(X) = \log(\mathbb{E}[\exp(-X)])$ is only finite for distributions where the tail of their density functions decays faster than $\exp(-|x|)$. However, if the modified chi-squared divergence $\phi_2(x) = (x-1)^2$ is used to construct an uncertainty set where the nominal model for $Y \triangleq -X$ is the exponential distribution $f_0(y) = \lambda \exp(-\lambda y)$ for $y \geq 0, \lambda > 1$, then this can lead to an infinitely pessimistic robust risk evaluation. Indeed, any exponential distribution $g(y) = \eta \exp(-\eta y)$ with $\eta < 1 < \lambda < 2\eta$ has a finite modified chi-squared divergence with respect to f_0. Therefore, the modified chi-squared divergence ball will contain the distribution g for sufficiently large radius r. However, since $1 - \eta > 0$, the entropic risk measure is infinite under the distribution of g. In fact, for any radius $r > 0$, one can construct a density $\tilde{g}$, that has the same asymptotic tail behavior as $\exp(-\eta y)$, and such that $I_{\phi_2}(\tilde{g}, f_0) < r$ (see Kruse et al. 2021). Therefore, in this particular example, the modified chi-squared divergence ball will always lead to an infinite robust entropic risk measure evaluation, for all radius $r > 0$.

Motivated by these examples, it is thus important to establish the necessary and sufficient conditions that one must impose on ϕ_2 to ensure a finite evaluation of the robust risk measures. For robust OCE risk measures $\rho^s_{\mathrm{oce},\mathbb{P}_0}$ and $\rho^l_{\mathrm{oce},\mathbb{P}_0}$, these conditions can be identified by reformulating $\rho^s_{\mathrm{oce},\mathbb{P}_0}$ and $\rho^l_{\mathrm{oce},\mathbb{P}_0}$ into finite-dimensional problems using convex duality. This is stated in the following theorem, where for any divergence function ϕ, we define $0\phi^*(s/0) \triangleq 0$ if $s \leq 0$ and $0\phi^*(s/0) \triangleq \infty$ if $s > 0$.

Theorem 1. *Let $X \in L^1(\mathbb{P}_0)$ and $\phi_1, \phi_2 \in \Phi_0$ be lower-semicontinuous. Then, we have the following equalities:*

$$\begin{aligned} \rho^l_{\mathrm{oce},\mathbb{P}_0}(X) = \inf_{\theta_1,\theta_2\in\mathbb{R}} & -\theta_1 - \theta_2 \\ & + \mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*(\phi_1^*(\theta_2 - X) + \theta_1)\right] \end{aligned} \tag{7}$$

$$\begin{aligned} \rho^s_{\mathrm{oce},\mathbb{P}_0}(X) = \inf_{\substack{\lambda\geq 0\\ \theta_1,\theta_2\in\mathbb{R}}} & -\theta_1 - \theta_2 + \lambda r \\ & + \mathbb{E}_{\mathbb{P}_0}\left[\lambda \phi_2^*\left(\frac{\phi_1^*(\theta_2 - X) + \theta_1}{\lambda}\right)\right]. \end{aligned} \tag{8}$$

Furthermore, if the following conditions are satisfied:

1. *ϕ_1^* and ϕ_2^* are differentiable on $\mathbb{R}$,*

2. *Each domain set $\mathcal{F}(\boldsymbol{\theta}) \triangleq \{\boldsymbol{\theta} \in \mathbb{R}^2 \mid \int_\Omega \beta^*(\omega, \boldsymbol{\theta}) \mathrm{d}\mathbb{P}_0(\omega) < \infty\}$ and $\tilde{\mathcal{F}}(\boldsymbol{\theta}, \lambda) \triangleq \{(\boldsymbol{\theta}, \lambda) \in \mathbb{R}^2 \times [0, \infty) \mid \int_\Omega \tilde{\beta}^*(\omega, \boldsymbol{\theta}, \lambda) \mathrm{d}\mathbb{P}_0(\omega) < \infty\}$ has non-empty interior, where $\beta^*(\omega, \boldsymbol{\theta}) \triangleq \phi_2^*(\phi_1^*(\theta_2 - X(\omega)) + \theta_1)$ and $\tilde{\beta}^*(\omega, \boldsymbol{\theta}, \lambda) \triangleq \lambda \phi_2^* \left(\frac{\phi_1^*(\theta_2 - X(\omega)) + \theta_1}{\lambda} \right)$.*

Then, there exists dual solutions $\boldsymbol{\theta}^*$ and $(\tilde{\boldsymbol{\theta}}^*, \lambda^*)$ that attain respectively the minimum (7) *and* (8)*. Moreover, if $\boldsymbol{\theta}^* \in \mathrm{int}(\mathcal{F}(\boldsymbol{\theta}))$ and $(\tilde{\boldsymbol{\theta}}^*, \lambda^*) \in \mathrm{int}(\tilde{\mathcal{F}}(\boldsymbol{\theta}, \lambda))$, then the partial derivatives of $\beta^*(\omega, .)$ and $\tilde{\beta}^*(\omega, ., \lambda)$ with respect to the variable $\boldsymbol{\theta}$, evaluated at the dual solutions $\boldsymbol{\theta}^*$ and $(\tilde{\boldsymbol{\theta}}^*, \lambda^*)$, are respectively the worst-case densities of the measure $\mathbb{Q}^*, \bar{\mathbb{Q}}^*$ with respect to $\mathbb{P}_0$ that attain the maximum of the primal problems* (6) *and* (5)*.*

As a direct consequence of Theorem 1, we derive the following necessary and sufficient conditions for the finiteness of $\rho^s_{\mathrm{oce}, \mathbb{P}_0}(X)$ and $\rho^l_{\mathrm{oce}, \mathbb{P}_0}(X)$.

Corollary 1. *Let $X \in L^1(\mathbb{P}_0)$ and $\phi_1, \phi_2 \in \Phi_0$ be lower-semicontinuous. Suppose there exists a real-valued density function $f_0(x)$ for the distribution of X under $\mathbb{P}_0$. Then, we have that $\rho^s_{\mathrm{oce}, \mathbb{P}_0}(X) < \infty$, if and only if there exists $\theta_1, \theta_2 \in \mathbb{R}, \lambda \geq 0$, such that*

$$\int_{\mathbb{R}} \lambda \phi_2^* \left(\frac{\phi_1^*(\theta_2 - x) + \theta_1}{\lambda} \right) f_0(x) \mathrm{d}x < \infty. \tag{9}$$

Similarly, we have that $\rho^l_{\mathrm{oce}, \mathbb{P}_0}(X) < \infty$, if and only if there exists $\theta_1, \theta_2 \in \mathbb{R}$, such that

$$\int_{\mathbb{R}} \phi_2^* \left(\phi_1^*(\theta_2 - x) + \theta_1 \right) f_0(x) \mathrm{d}x < \infty. \tag{10}$$

The integral conditions (9) and (10) displayed in Corollary 1 show that the finiteness of a robust OCE risk measure is completely determined by the tail behaviors of three components $(\phi_2^*, \phi_1^*, f_0)$, where ϕ_2^*, f_0 control the content of the ϕ_2-uncertainty set, and ϕ_1^* encodes the sensitivity of the risk measure to losses in the tail. This provides a guideline on how to calibrate an uncertainty set to a given risk measure and nominal model. Namely, one must choose the ϕ_2-divergence function such that its conjugate ϕ_2^* satisfies the integral condition in (9) or (10).

We note that the (9) and (10) are only verifiable if we have information on the tail behavior of the nominal density f_0. In practice, X might depend on many underlying risk factors $Z_1, \ldots, Z_I$, where only the nominal distributions of the marginals are specified and have an explicit form. In the following proposition, which is adapted from Kruse et al. [2019], we provide a sufficient condition under which one can choose the divergences based on the nominal distributions of the marginals Z_i's.

Proposition 1. *Let $X \in L^1(\mathbb{P}_0)$. Suppose there exists a constant $C > 0$, such that*

$$|X| \leq C \left(1 + \sum_{i=1}^m |Z_i| \right), \tag{11}$$

holds $\mathbb{P}_0$-almost surely. If there exists $\theta_1, \theta_2 \in \mathbb{R}$, such that for all $i = 1, \ldots, I$,

$$\mathbb{E}_{\mathbb{P}_0} \left[\phi_2^* \left(\theta_1 + \phi_1^*(\theta_2 + C(1 + m \cdot |Z_i|)) \right) \right] < \infty.$$

Then, the integral conditions (9) *and* (10) *are also satisfied.*

Using Corollary 1, we perform several tail analyses and illustrate how most standard choices of ϕ-divergences are not suitable to address tail uncertainty for some common risk measures and nominal distributions. Throughout this paper, we use the big O notation $f_1(x) = O(f_2(x))$ if $\limsup_{x\to\infty} |f_1(x)|/|f_2(x)| < \infty$, and the small o notation $f_1(x) = o(f_2(x))$, if we have $\lim_{x\to\infty} |f_1(x)|/|f_2(x)| = 0$.

Example 1. Consider the CVaR_α risk measure with a log-normal nominal model. A common choice of a ϕ-divergence is the Kullback-Leibler divergence, where $\phi(t) = t\log(t) - t + 1$. Its conjugate is given by $\phi^*(s) = \exp(s) - 1$. The CVaR_α risk measure (2) is an OCE risk measure with $\phi_1^*(s) = \max\{s/(1-\alpha), 0\}$. However, examining the integral condition (9) reveals that for any $\theta_1, \theta_2 \in \mathbb{R}, \lambda > 0$, the integrand has a tail behavior of $O\left(\exp\left\{|x|/(\lambda\alpha)\right\}\right)$, which diverges to $+\infty$ as $x \to -\infty$. Hence, the uncertainty set induced by the Kullback-Leibler divergence is too conservative for the CVaR_α risk measure under a log-normal nominal model. In Table 1, we give an overview of the finiteness status of robust CVaR with other nominal distributions and ϕ-divergences.

Example 2. The χ^2-distance is another popular ϕ-divergence where $\phi(t) = (t-1)^2/t$. However, its conjugate $\phi^*(s) = 2 - 2\sqrt{1-s}, s < 1$ is only finite on the domain on $(-\infty, 1)$, i.e., $\phi^*(s) = +\infty$, for $s > 1$. Hence, the integral condition (9) can never be satisfied if the χ^2-distance is used to define a robust OCE risk measure, for any nominal distribution that has unbounded support towards $-\infty$.

Similarly, other canonical examples of ϕ-divergences such as Burg entropy $\phi(t) = -\log(t) + t - 1$, total variation distance $\phi(t) = |t-1|$, and polynomial divergence $\phi(t) = (t^p - p(t-1) - 1)/(p(p-1))$ with degree $p < 1$, all have a conjugate function ϕ^* that is infinite on certain interval $[a, \infty)$ (see Table 2 of Ben-Tal et al. 2013 and page 6 of Pardo 2006). Hence, they all lead to an infinite robust OCE risk evaluation for a nominal distribution with unbounded support towards $-\infty$.

Example 3. The entropic risk measure $\rho_{e,\gamma}(X) = \log\left(\mathbb{E}[\exp(-\gamma X)]\right)/\gamma$ is an OCE risk measure with $\phi_1^*(x) = (\exp(\gamma x) - 1)/\gamma$. Due to its exponential growth behavior, the entropic risk measure has infinite evaluation for heavy-tailed distributions and therefore also its robust counterpart. For the Gaussian distribution, the entropic risk measure is finite and equal to $\mu + \gamma\sigma^2/2$. However, if one chooses the KL-divergence to define a robust entropic risk measure with a Gaussian nominal model, then condition (9) is not satisfied for any $\theta_1, \theta_2 \in \mathbb{R}, \lambda > 0$, since the integrand has a tail asymptotic of the order

$$O\left(\exp\left\{\frac{1}{\lambda}\left(\frac{1}{\gamma}\left(\exp(\gamma(\theta_2 - x)) - 1\right) + \theta_1\right) - \left(\frac{x-\mu}{\sigma}\right)^2\right\}\right),$$

which diverges to ∞ as $x \to -\infty$. Therefore, the KL-divergence is not adequate to address model uncertainty for the entropic risk measure, under many nominal distributions such as the Gaussian and Weibull distribution, as shown in Table 2.

Divergence $\phi(t)$	$N(\mu,\sigma^2)$	$W(\lambda_0,k)$	$\ln(\mu,\sigma^2)$	$P(x_m,p_0)$	$t(\nu)$
$t\log t - t + 1$ *(Kullback-Leibler)*	$<\infty$	$<\infty \Leftrightarrow k \geq 1$	∞	∞	∞
$(t^p - pt + p - 1)/(p(1-p))$ *(Polynomial, $p > 1$)*	$<\infty$	$<\infty$	$<\infty$	$<\infty \Leftrightarrow \frac{p}{p-1} < p_0$	$<\infty \Leftrightarrow \frac{p}{p-1} < \nu$
$(t-1)^2$ *(Modified χ^2-distance)*	$<\infty$	$<\infty$	$<\infty$	$<\infty \Leftrightarrow p_0 > 2$	$<\infty \Leftrightarrow \nu > 2$
$(t-1)^2/t$ *(χ^2-distance)*	∞	∞	∞	∞	∞
$\lvert t-1 \rvert$ *(Variation distance)*	∞	∞	∞	∞	∞
$(\sqrt{t}-1)^2$ *(Hellinger distance)*	∞	∞	∞	∞	∞
$-\log t + t - 1$ *(Burg entropy)*	∞	∞	∞	∞	∞

Table 1: Finiteness status of robust CVaR_α risk measures (5) and (6) for different divergence measures versus nominal distributions. $N(\mu,\sigma^2)$ and $\ln(\mu,\sigma^2)$ are respectively the normal and lognormal distribution with mean μ and variance σ^2. $W(\lambda_0,k)$ denotes the Weibull distribution with scale λ_0 and shape k. $P(x_m,p_0)$ is the Pareto distribution with scale x_m and shape p_0. $t(\nu)$ denotes the Student's t-distribution with ν degrees of freedom. Details of derivations can be found in Section EC.4.

From these examples, we observe that many standard choices of ϕ-divergence uncertainty sets often lead to infinitely conservative robust risk evaluations. On the other hand, the polynomial divergence $\phi(t) = (t^p - p(t-1) - 1)/(p(p-1))$ can often be made to satisfy the finiteness conditions (9) and (10), by choosing a sufficiently large p. Indeed, an examination of its conjugate function $\phi^*(s) = \max\{1 + s(p-1), 0\}^{\frac{p}{p-1}}/p - 1/p$ shows that as $p \to \infty$, ϕ^* approaches the linear function $\phi^*(s) = s$, which is the least conservative conjugate function of the ϕ-divergence that only considers the nominal model (i.e., $\phi(1) = 0$, $\phi(t) = +\infty$ elsewhere). However, the drawback of a polynomial divergence uncertainty set is that it can be very restrictive. For example, consider two Weibull distributions $g(x) = kx^{k-1}\exp(-x^k)$ and $f(x) = lx^{l-1}\exp(-x^l)$ with $k < l$. Then, any polynomial divergence $\phi(t) = O(t^p)$ with $p > 1$ would give $I_\phi(g,f) = \infty$, since $\phi(g(x)/f(x))\, f(x) = O(x^{k-l}\exp((p-1)x^l - px^k))$. Therefore, when the nominal distribution belongs to the Weibull class, an uncertainty set induced by the polynomial divergence does not include any other Weibull distribution that has a heavier tail. Hence, apart from the conditions in Corollary 1, we need to impose additional constraints on the conjugate function ϕ_2^*, such that the corresponding divergence ϕ_2 is not too restrictive. We will address this issue in the next subsection.

3.1 Controlling the Tail Properties of Distributions in a Divergence Ball

As illustrated by previous examples, many classical divergences fail to guarantee the finiteness of a robust risk measure, due to improper matching of tail behaviors between ϕ_2^*, ϕ_1^* and f_0. Therefore, to construct an alternative ϕ_2-divergence that addresses this issue, we can specify a new tail function $\tilde{\psi}(s), s \geq 0$ that is compatible with (ϕ_1^*, f_0) in the sense of Corollary 1, and then set $\phi_2^*(s) = \tilde{\psi}(s)$, for $s \geq 0$. Since the integral conditions in Corollary 1 depend only on the tail behavior of $\phi_2^*(s)$ as $s \to \infty$, we may simply set $\phi_2^*(s) = \exp(s) - 1$ for $s \leq 0$. Finally, the new tail function $\tilde{\psi}(s)$ must be normalized such that $(\phi_2^*)^*$ becomes a ϕ-divergence function (i.e., a convex, non-negative function satisfying $\phi(1) = 0$ and $\mathrm{dom}(\phi) \subset [0,\infty)$). This can be

Divergence $\phi(t)$	$N(\mu,\sigma^2)$	$W(\lambda_0,k)$	$\ln(\mu,\sigma^2)$	$P(x_m,p_0)$	$t(\nu)$
$t\log t - t + 1$ *(Kullback-Leibler)*	∞	∞	∞	∞	∞
$(t^p - pt + p - 1)/(p(1-p))$ *(Polynomial, $p > 1$)*	$< \infty$	$< \infty \Leftrightarrow \begin{cases} k > 1 \\ k = 1 \text{ and } \frac{p}{p-1} < \frac{\gamma}{\lambda_0} \end{cases}$	∞	∞	∞
$(t-1)^2$ *(Modified χ^2-distance)*	$< \infty \Leftrightarrow \gamma > 2\sigma^2$	$< \infty \Leftrightarrow \begin{cases} k > 2 \\ k = 2 \text{ and } \gamma > \lambda_0 \end{cases}$	∞	∞	∞
$(t-1)^2/t$ *(χ^2-distance)*	∞	∞	∞	∞	∞
$\lvert t-1 \rvert$ *(Variation distance)*	∞	∞	∞	∞	∞
$(\sqrt{t}-1)^2$ *(Hellinger distance)*	∞	∞	∞	∞	∞
$-\log t + t - 1$ *(Burg entropy)*	∞	∞	∞	∞	∞

Table 2: Finiteness status of robust entropic risk measures (5) and (6) with parameter $\gamma > 0$, for different divergence measures versus nominal distributions, with exponential utility OCE. $N(\mu, \sigma^2)$ and $\ln(\mu, \sigma^2)$ are respectively the normal and lognormal distribution with mean μ and variance σ^2. $W(\lambda_0, k)$ denotes the Weibull distribution with scale λ_0 and shape k. $P(x_m, p_0)$ is the Pareto distribution with scale x_m and shape p_0. $t(\nu)$ denotes the Student's t-distribution with ν degrees of freedom. Details of derivations can be found in Section EC.4.

done by choosing $\tilde{\psi}(s)$ to be increasing and convex with $\tilde{\psi}(0) = 0$ and $\tilde{\psi}'(0) = 1$. To conclude, we propose the following construction: let ψ be an increasing, convex differentiable function on $[0, \infty)$ such that $\psi''(0) \neq 0$. Set, for $s \geq 0$,

$$\tilde{\psi}(s) = \frac{1}{\psi''(0)} \left(\psi(s) + \left(\psi''(0) - \psi'(0)\right) s - \psi(0)\right), \tag{12}$$

and define the function

$$\phi_2^*(s) = \begin{cases} \tilde{\psi}(s) & s \geq 0 \\ \exp(s) - 1 & s \leq 0. \end{cases} \tag{13}$$

By construction, we have $(\phi_2^*)'(0) = 1$ and $(\phi_2^*)''(0) = 1$. These are normalization procedures that ensure $\phi_2(1) = 0, \phi_2'(1) = 0$ and $\phi_2''(1) = 1$, where $\phi_2 = \phi_2^{**}$ (the first two conditions are needed for ϕ_2 to be a ϕ-divergence, and $\phi_2''(1) = 1$ is often desirable in the statistical application of ϕ-divergences, see Remark 1 of Kruse et al. 2019 and Pardo 2006). The following proposition, which is an application of Corollary 23.5.1 of Rockafellar [1970], verifies that ϕ_2^*, as defined in (13), is indeed the conjugate function of a ϕ-divergence.

Proposition 2. *The function $\phi_2 \triangleq (\phi_2^*)^*$, where ϕ_2^* is defined in* (13)*, is a convex, non-negative function on $[0, \infty)$ with $\phi_2(1) = 0$. If ψ in* (12) *is strictly convex and continuously differentiable with $\lim_{s\to\infty} \psi'(s) = \infty$, then ϕ_2 is finite on $[0, \infty)$ and we have that $\phi_2'(t) = ((\phi_2^*)')^{-1}(t)$, for $t > 0$.*

Besides the finiteness of a robust risk measure, it is also desirable to have a divergence that is not too restrictive, and allows for distributions with heavier tails than the nominal one in the divergence ball. Intuitively, the degree of penalization of ϕ_2 is determined by its growth rate as $t \to \infty$, which is governed by the derivative $\phi_2'(t)$. Therefore, one must control $\phi_2'(t)$ to adjust the content of a divergence ball. However, if the divergence is constructed through the conjugate as in (13), then there is not always an explicit form of ϕ_2' available that allows us to examine its growth behavior as $t \to \infty$. Indeed, although the equality $\phi_2' = ((\phi_2^*)')^{-1}$ holds, the inverse of $(\phi_2^*)'$ cannot always be calculated explicitly. Therefore, it would seem more convenient to specify ϕ_2 through its derivative instead of the conjugate, and to derive sufficient conditions for (9) and (10) in terms of ϕ_2' (e.g., by bounding the conjugate using $\phi_2^*(s) \leq s(\phi_2')^{-1}(s)$, similarly as in Kruse et al. 2019). This creates a dilemma, since the latter construction does not give an analytical expression for the conjugate function and its derivatives, which is important for the use of optimization techniques to compute the robust risk measures via their dual problems, as we elaborate in Section 6. Nevertheless, we can resolve this dilemma as follows: first, use the divergence construction as given in (13) to obtain an explicit conjugate function ϕ_2^*. Second, we calculate the derivative $(\phi_2^*)'$, which gives $(\phi_2')^{-1}$ by Proposition 2. Finally, we may find functions with similar tail properties as $(\phi_2^*)'$ that provide respectively upper and lower bound on $(\phi_2^*)'$, for which the inverse can be explicitly calculated and used to examine the content of a divergence ball (using Proposition 2 and 3 in Kruse et al. 2019). We state this more precisely in the following proposition.

Proposition 3. *Let ϕ_2 be a continuous divergence function with a strictly increasing derivative ϕ_2'. Suppose ψ_1' is a continuous function such that its inverse satisfies $(\psi_1')^{-1}(y) \leq (\phi_2')^{-1}(y)$ for all $y \geq y_0$, for some $y_0 \in \mathbb{R}$. If for some $d > 1$, we have that*

$$\limsup_{|x| \to \infty} \psi_1' \left(\frac{1}{f_0(x)} \right) |x|^{-d} < \infty, \tag{14}$$

then for any probability density g such that the likelihood g/f_0 is bounded on any compact subset of $(-\infty, 0]$, we have that if g has a finite d-th moment: $\int_{-\infty}^{\infty} g(x)|x|^d \mathrm{d}x < \infty$, then we also have $I_{\phi_2}(g, f_0) < \infty$.

On the other hand, if there exists a divergence function $\psi_2 \in \Phi_0$ such that $(\phi_2')^{-1}(y) \leq (\psi_2')^{-1}(y)$ for all $y \geq 0$, we have that $I_{\phi_2}(g, f_0) = \infty$, if $I_{\psi_2}(g, f_0) = \infty$, for any density function g.

3.2 Examples of Tailored Divergences

In this subsection, we provide examples of new ϕ_2-divergences that are tailored to a given risk measure and nominal model. We will only specify $\phi_2^*(s)$ for $s \geq 0$ in these examples, since we always set $\phi_2^*(s) = \exp(s) - 1$ as in (13).

Example 4. Consider the CVaR_α risk measure with the generalized log-normal nominal distribution, where the density function is given by

$$f_0(x) = \frac{1}{C\sigma|x|} \exp\left\{ -\frac{1}{p\sigma^p} (\log|x| - \mu)^p \right\}, \ x \leq 0,$$

where $\mu \in \mathbb{R}, \sigma > 0, p \geq 2$, and $C > 0$ is a normalization constant. As shown in Example 1, the KL-divergence is not suitable for this combination of nominal model and risk measure. Hence, using the construction method in (13), we propose the following divergence specified by its conjugate function:

$$\phi_2^*(s) = c_1(s+e)\exp\left\{\frac{1}{p\,(\sigma d)^p}\log^p(s+e)\right\} \\ + c_2 s + c_3,\ \ s \geq 0, \tag{15}$$

where the constants are $c_1 = 1/(p^2(a^2+a)\exp(a-1)), c_2 = 1-\exp(a)(ap+1)/c_1, c_3 = -\exp(a+1)/c_1$ with $a = 1/(p\,(\sigma d)^p)$ and $d > 1$. We note that the proposed divergence (15) contains the parameters (σ, p), due to its dependence on the log-normal nominal model. It also contains a parameter $d > 1$, which controls the moments of the distributions in its divergence ball. Notably, the divergence (15) is independent of the parameter α of the Conditional Value-at-Risk, which is due to the linearity of the ϕ_1^* function corresponding to the CVaR_α risk measure, that does not determine the asymptotic behavior of the integrand in (10).

Lemma 1. *The function ϕ_2^* defined in* (15) *is strictly convex and twice continuously differentiable on $\mathbb{R}$. Its conjugate $\phi_2 \triangleq (\phi_2^*)^*$ belongs to Φ_0. Moreover, ϕ_2^* satisfies* (9) *and* (10)*, for the* CVaR_α *risk measure and the generalized log-normal nominal model.*

In addition, for any continuous probability density g defined on $(-\infty, 0]$ that has finite d-th moment: $\int_{-\infty}^0 g(x)|x|^d < \infty$, we have that $I_{\phi_2}(g, f_0) < \infty$. Conversely, for any density function g such that $\liminf_{x\to-\infty} |x|^{t+1} g(x) > 0$, where $t \in (1, d)$, we have $I_{\phi_2}(g, f_0) = \infty$.

Example 5. Consider now the CVaR_α risk measure with a Weibull nominal distribution, where the density f_0 is given by:

$$f_0(x) = \frac{k}{\lambda}\left(\frac{|x|}{\lambda}\right)^{k-1}\exp\left\{-\left(\frac{|x|}{\lambda}\right)^k\right\},\ \ x \leq 0,$$

where $\lambda, k > 0$. If $k < 1$, the Weibull distribution becomes heavy-tailed, and the robust CVaR_α risk measure defined by the KL-divergence will lead to an infinite evaluation. To match the exponential tail of f_0, we propose the following divergence:

$$\phi_2^*(s) = c_1(s+1)\exp\{(s+1)^{k/d}\} \\ + c_2 s + c_3,\ \ s \geq 0. \tag{16}$$

where $d > 1$, and the normalization constants are $c_1 = 1/(ep(2p+1))$, $c_2 = 1 - (1+p)/(2p+1), c_3 = 1/(p(2p+1))$ with $p = k/d$.

Lemma 2. *The function ϕ_2^* defined in* (16) *is strictly convex and twice continuously differentiable on $\mathbb{R}$. Its conjugate $\phi_2 \triangleq \phi_2^{**}$ belongs to Φ_0. Moreover, ϕ_2^* satisfies the finiteness conditions* (9) *and* (10)*, for the* CVaR_α *risk measure and the Weibull nominal model.*

In addition, for any continuous probability density g defined on $(-\infty, 0]$ that has finite d-th moment: $\int_{-\infty}^0 g(x)|x|^d < \infty$, we have that $I_{\phi_2}(g, f_0) < \infty$. Conversely, for any density function g such that $\liminf_{x\to-\infty} |x|^{t+1} g(x) > 0$, where $t \in (1, d)$, we have $I_{\phi_2}(g, f_0) = \infty$.

Example 6. Consider the entropic risk measure $\rho_{e,\gamma}$ with the Weibull nominal distribution, where $k > 1$. Table 2 shows that the KL-divergence is generally not suitable for the entropic risk measure. Moreover, as mentioned previously, the polynomial divergence is too restrictive since its uncertainty set does not include any Weibull distributions with a heavier tail than the nominal one. Therefore, we propose the following divergence:

$$\begin{aligned}\phi_2^*(s) = c_1(s+e)\exp\left\{\frac{1}{(2\gamma\lambda)^k}\log^k(s+e)\right\}\\ + c_2 s + c_3,\ \ s\geq 0,\end{aligned} \tag{17}$$

where the constants are $c_1 = 1/(k^2(a^2+a)\exp(a-1))$, $c_2 = 1-\exp(a)(ak+1)/c_1$, $c_3 = -\exp(a+1)/c_1$ with $a = 1/\left(2\gamma\lambda\right)^k$.

Lemma 3. *The function ϕ_2^* defined in* (17) *is strictly convex and twice continuously differentiable on $\mathbb{R}$. Its conjugate $\phi_2 \triangleq \phi_2^{**}$ belongs to Φ_0. Moreover, ϕ_2^* satisfies the finiteness conditions* (9) *and* (10)*, for the entropic risk measure (with parameter $\gamma > 0$) and the Weibull nominal distribution f_0 with parameters (λ, k). Furthermore, for any Weibull density function g with parameters (λ, l) such that $1 < l < k$, we have $I_{\phi_2}(g, f_0) < \infty$.*

4 Other Types of Robust Risk Measures

In this section, we show that the formulation with two ϕ-divergences (ϕ_1, ϕ_2) provides a general framework that encompasses other families of robust risk measures. Similar to the previous sections, we establish the sufficient and necessary conditions for finiteness and derive the finite-dimensional dual problems of the corresponding robust risk measures.

We first note that the robust expected utility risk measure can also be formulated in terms of two ϕ-divergences (ϕ_1, ϕ_2). Indeed, as shown in (4), expected utility can be represented by the divergence $\phi_1 = -u_*$. The dual reformulation of robust expected utility risk measures has already been derived in Breuer and Csiszár [2016], and its implication for divergence choices has been studied in Kruse et al. [2019, 2021]. Therefore, in this section, we focus on the class of robust shortfall risk measures, which can be characterized as follows:

$$\begin{aligned}\rho^s_{\mathrm{sf},\mathbb{P}}(X) \triangleq \sup_{\lambda>0}\ \sup_{\{\mathbb{Q}: I_{\phi_2}(\mathbb{Q},\mathbb{P}_0)\leq r\}}\ \sup_{\{\bar{\mathbb{Q}}:\bar{\mathbb{Q}}\ll\mathbb{Q}\}} \mathbb{E}_{\bar{\mathbb{Q}}}[-X]\\ - I_{\phi_1^\lambda}(\bar{\mathbb{Q}}, \mathbb{Q}).\end{aligned} \tag{18}$$

By examining the dual of (18), we can once again establish the finiteness conditions for (18) in terms of the tail behavior of ϕ_1^*, ϕ_2^* and f_0.

Theorem 2. *Let $X \in L^1(\mathbb{P}_0)$. Let $\phi_1, \phi_2 \in \Phi_0$ be lower-semicontinuous. Then, we have that $\rho^s_{\mathrm{sf},\mathbb{P}}(X) < \infty$, if and only if there exists $\theta_1, \theta_2 \in \mathbb{R}, \eta \geq 0$, such that*

$$\int_{\mathbb{R}} \eta\phi_2^*\left(\frac{\phi_1^*(-\theta_2 - x) + \theta_1}{\eta}\right) f_0(x)\mathrm{d}x \leq \theta_1 - \eta r. \tag{19}$$

Moreover, we have the duality relation:

$$\rho^s_{\mathrm{sf},\mathbb{P}}(X) = \inf_{\substack{\eta\geq 0\\ \theta_1,\theta_2\in\mathbb{R}}} \{\theta_2 \mid \mathbb{E}_{\mathbb{P}_0}\left[\beta^*(X,\theta_1,\theta_2,\eta)\right] \\ \leq \theta_1 - \eta r\}, \tag{20}$$

where $\beta^*(X,\theta_1,\theta_2,\eta) = \eta\phi_2^*\left(\frac{\phi_1^*(-\theta_2-X)+\theta_1}{\eta}\right)$.

However, a similar duality result cannot be obtained if the robust shortfall risk measure is defined by adding a second ϕ_2 penalization (such as in (6)), since the non-convex product term $\lambda\mathbb{Q}$ in $I_{\phi_1^\lambda}(\bar{\mathbb{Q}},\mathbb{Q})$ is hard to handle.[4] Nevertheless, as an alternative, we note that one can instead also define a robust shortfall risk measure by adding the ϕ_2 penalization to the expectation constraint in (3), and then apply the reformulation in Breuer and Csiszár [2016] for the robust expectation. This would lead to the dual formulation in (27).

The necessary and sufficient conditions for $\rho^s_{\mathrm{sf},\mathbb{P}}(X) < \infty$ as established in (19) are not easy to verify analytically. Indeed, the integral on the left-hand side of (19) requires an explicit calculation, whereas in (9) for example, one only needs to perform an asymptotic tail analysis. To work around this, we note that the right-hand side of (19) is independent of θ_2, whereas the left-hand side is monotonically decreasing as $\theta_2 \to \infty$. This allows us to develop a set of tests presented in the following corollary, where (19) can be verified without the need for an explicit calculation of the integral in (19).

Corollary 2. *Let the assumptions of Theorem 2 be fulfilled, and suppose that* $\phi_1^*(-\infty) \triangleq \lim_{s\to-\infty}\phi_1^*(s)$ *exists, either finite or equal to* $-\infty$*. Then, the following hold:*

1. *If there exists no* $\theta_1,\theta_2\in\mathbb{R}$, $\eta\geq 0$, *such that*

$$\int_{\mathbb{R}} \eta\phi_2^*\left(\frac{\phi_1^*(-\theta_2-x)+\theta_1}{\eta}\right) f_0(x)\mathrm{d}x < \infty, \tag{21}$$

   then $\rho^s_{\mathrm{sf},\mathbb{P}}(X) = \infty$.

2. *If for all* $\theta_1,\theta_2\in\mathbb{R}$, $\eta\geq 0$ *that satisfy* (21), *we have*

$$\eta\phi_2^*\left(\frac{\phi_1^*(-\infty)+\theta_1}{\eta}\right) > \theta_1 - \eta r, \tag{22}$$

   then $\rho^s_{\mathrm{sf},\mathbb{P}}(X) = \infty$ *(if* $\phi_1^*(-\infty) = -\infty$*, we define* $\eta\phi_2^*\left(\frac{-\infty+\theta_1}{\eta}\right) \triangleq \lim_{s\to-\infty}\eta\phi_2^*\left(\frac{s}{\eta}\right)$*, for any* $\eta\geq 0, \theta_1\in\mathbb{R}$*).*

3. *If there exists some* $\theta_1,\theta_2\in\mathbb{R}$, $\eta\geq 0$, *such that* (21) *holds, and*

$$\eta\phi_2^*\left(\frac{\phi_1^*(-\infty)+\theta_1}{\eta}\right) < \theta_1 - \eta r, \tag{23}$$

   then, $\rho^s_{\mathrm{sf},\mathbb{P}}(X) < \infty$.

[4] In fact, this is also the case for (18), but the set formulation in (18) allows us to circumvent the non-convexity by a change of variable.

Example 7. Consider a shortfall risk measure with a quadratic $\phi_1(t) = (t-1)^2$. Then,

$$\phi_1^*(s) = \begin{cases} s + s^2/4 & s \geq -2 \\ -1 & s < -2. \end{cases}$$

We show that the robust shortfall risk measure with the following combinations of divergence ϕ_2 and nominal distribution f_0 is finite for $r \leq 1$:

(1) Polynomial divergence with $p = 3$ and Log-normal distribution

(2) KL-divergence and Gaussian distribution (with $\sigma \leq 1$)

(3) Modified χ^2-divergence and Student-t distribution (with $\nu > 4$)

(1): We apply Corollary 2. First, for any $\theta_1, \theta_2 \in \mathbb{R}$, $\eta > 0$, as $x \to -\infty$,

$$\eta\phi_2^*\left(\frac{\phi_1^*(-\theta_2 - x) + \theta_1}{\eta}\right) f_0(x) = O\left(|x|^{2p/(p-1)} \exp\left(-\log^2|x|/(2\sigma^2)\right)\right),$$

which is integrable for any $p > 1$. Hence, (21) is satisfied for any $\theta_1, \theta_2 \in \mathbb{R}$, $\eta > 0$. It remains to verify (23). This can be verified numerically. We find that for $\theta_1 = 0.1098$, $\eta = 0.01$, (23) is satisfied for all $r \leq 1$.

(2): For any $\theta_1, \theta_2 \in \mathbb{R}$ and $\eta > \sigma^2/2$, we have that for $x \leq 2 - \theta_2$ (note that the remaining part of the integrand below for $x > 2 - \theta_2$ is integrable):

$$\eta\phi_2^*\left(\frac{\phi_1^*(-\theta_2 - x) + \theta_1}{\eta}\right) f_0(x) = O\left(\exp\left(\left(\frac{1}{4\eta} - \frac{1}{2\sigma^2}\right) x^2\right)\right).$$

We note that $1/(4\eta) - 1/(2\sigma^2) < 0$ since $\eta > \sigma^2/2$. Hence, (21) is satisfied for any $\theta_1, \theta_2 \in \mathbb{R}$ and $\eta > \sigma^2/2$. It remains to verify (23). We find numerically that for $\theta_1 = 0.2098$, $\eta = 0.501$, (23) is satisfied for all $\sigma \leq 1, r \leq 1$.

(3): For any $\theta_1, \theta_2 \in \mathbb{R}$, $\eta > 0$, we have that as $x \to -\infty$,

$$\eta\phi_2^*\left(\frac{\phi_1^*(-\theta_2 - x) + \theta_1}{\eta}\right) f_0(x) = O\left(|x|^{4-\nu-1}\right),$$

which is only integrable if $\nu > 4$. It remains to verify (23). We find numerically that for $\theta_1 = 0.1098$, $\eta = 0.501$, (23) is satisfied for all $r \leq 1$.

5 Elicitation of Divergences from Robust Risk Measures

In the previous sections, we have shown that the robust risk measures characterized by (ϕ_1, ϕ_2) can be reformulated as finite-dimensional dual problems in terms of the conjugates (ϕ_1^*, ϕ_2^*). In this section, we study the inverse problem of recovering the (ϕ_1^*, ϕ_2^*) from the robust risk measures. In the non-robust case where ambiguity is not present, it is shown by Ben-Tal and Ben-Israel [1991], Ben-Tal and Teboulle [2007] that utility functions which are normalized and strongly risk-averse (i.e., $u(0) = 0, u(x) < x, \forall x \in \mathbb{R}$) can be elicited by querying the certainty equivalent values (EU, OCE, and u-Mean) of random variables X_p, which for any $x \in \mathbb{R}$ is defined as

$$X_p = \begin{cases} x & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases} \tag{24}$$

More precisely, for any certainty equivalent CE $\in \{\text{EU}, \text{OCE}, \text{u-Mean}\}$, we have the limit relation

$$\lim_{p \downarrow 0} \frac{\text{CE}(X_p)}{p} = u(x). \tag{25}$$

Our goal is to extend the identity (25) to robust risk measures, which we show are possible for robust risk measures $\{\rho^l_{\text{eu},\mathbb{P}}, \rho^l_{\text{oce},\mathbb{P}}, \rho^l_{\text{sf},\mathbb{P}}\}$ that are defined through penalization. In the case of robust expected utility $\rho^l_{\text{eu},\mathbb{P}}$, such an extension is already evident due to its duality relation to OCE (see (2)):

$$\rho^l_{\text{eu},\mathbb{P}}(X) \triangleq \inf_{\theta \in \mathbb{R}} -\theta + \mathbb{E}_{\mathbb{P}}[\phi_2^* \left(\phi_1^*(-X) + \theta\right)],$$

where the right-hand side is equal to $-\text{OCE}_{v,\mathbb{P}}(u(X))$, for $u(x) \triangleq -\phi_1^*(-x)$ and $v(x) = -\phi_2^*(-x)$. Since identity (25) holds for the OCE, it follows that if $\phi_2^*(x) < x, \forall x \in \mathbb{R} \setminus \{0\}$, then we have that

$$\lim_{p \downarrow 0} \frac{\rho^l_{\text{eu},\mathbb{P}}(X_p)}{p} = \phi_2^*(\phi_1^*(-x)). \tag{26}$$

In Proposition 4, we state that the same relation (26) can also be derived for the robust OCE risk measure $\rho^l_{\text{oce},\mathbb{P}}$ and the robust shortfall risk measure $\rho^l_{\text{sf},\mathbb{P}}$ through their dual formulation. Recall that the dual of $\rho^l_{\text{oce},\mathbb{P}}$ is given in (7). The robust risk measure $\rho^l_{\text{sf},\mathbb{P}}$, which is defined by replacing the expected utility constraint in (3) with its robust counterpart, has the following dual formulation:

$$\begin{aligned} \rho^l_{\text{sf},\mathbb{P}}(X) = \inf_{\theta_1, \theta_2 \in \mathbb{R}} & \{\theta_2 \mid \mathbb{E}_{\mathbb{P}}[\phi_2^*(\theta_1 - u(X + \theta_2))] \\ & \leq \theta_1\}, \end{aligned} \tag{27}$$

Proposition 4. *Let $\phi_1, \phi_2 \in \Phi_0$ be lower-semicontinuous, twice continuously differentiable, and $\text{dom}(\phi_1^*) = \text{dom}(\phi_2^*) = \mathbb{R}$. Assume furthermore that $\phi_1^*(s) > s$, $\phi_2^*(s) > s$ for all $s \neq 0$. Then, for X_p defined in* (24) *with any $x \in \mathbb{R}$, we have*

$$\lim_{p \downarrow 0} \frac{\rho^l_{\text{oce},\mathbb{P}}(X_p)}{p} = \phi_2^*(\phi_1^*(-x)), \tag{28}$$

and

$$\lim_{p \downarrow 0} \frac{\rho^l_{\text{sf},\mathbb{P}}(X_p)}{p} = \phi_2^*(\phi_1^*(-x)). \tag{29}$$

Therefore, Proposition 4 suggests a method for measuring the divergences ϕ_1^*, ϕ_2^*, by conducting the following experiment. One can first measure ϕ_1^*, by retrieving the certainty equivalent for a sequence of X_p, where p is known and tends to zero. In this case, the decision maker is not ambiguous and (25) gives a measurement of $u(x) = -\phi_1^*(-x)$. In a second experiment, we determine the robust certainty equivalent for X_p, where p is now hidden from the decision maker. This will yield $\phi_2^*(\phi_1^*(-x))$. Of course, this only allows us to recover the values of ϕ_2^* on the image set of ϕ_1^*. On the other hand, the values of ϕ_2^* outside this image set do not contribute to the values of robust risk measures and can be thus taken arbitrarily.

Finally, we note that besides the elicitation of divergences, the formulation of the robust OCE and robust shortfall risk measures in terms of (ϕ_1^*, ϕ_2^*) also allows us to characterize risk aversion by simple convexity conditions imposed on ϕ_1^*, ϕ_2^*. We refer the readers to the Electronic Companion for more details.

6 Computing the Robust OCE Risk Measures

In this section, we discuss the computation of a robust OCE risk measure using sample average approximation (SAA). Theorem 1 states that robust OCE risk measures can be reformulated into the finite-dimensional dual problems (7) and (8), which are stochastic programming problems. The SAA method approximates these stochastic optimization problems by replacing $\mathbb{P}_0$ with the empirical distribution constructed from samples $X_1, \ldots, X_N$ from $\mathbb{P}_0$. This leads to the following optimization problems:

$$\min_{\theta_1, \theta_2 \in \mathbb{R}} -\theta_1 - \theta_2 + \frac{1}{N} \sum_{i=1}^{N} \phi_2^* \left(\phi_1^*(\theta_2 - X_i) + \theta_1 \right), \tag{30}$$

and

$$\begin{aligned} \min_{\substack{\lambda \geq 0 \\ \theta_1, \theta_2 \in \mathbb{R}}} \; & -\theta_1 - \theta_2 + \lambda r \\ & + \frac{1}{N} \sum_{i=1}^{N} \lambda \phi_2^* \left(\frac{\phi_1^*(\theta_2 - X_i) + \theta_1}{\lambda} \right). \end{aligned} \tag{31}$$

The formulations in (30) and (31) are convex optimization problems with only two or three variables. Therefore, they can be efficiently solved using the ellipsoid method (see, e.g., Bland et al. 1981). This is particularly useful if ϕ_1^* is only subdifferentiable, such as the case of the CVaR function where $\phi_1^*(x) = \max\{x/\alpha, 0\}$. If both ϕ_1^* and ϕ_2^* are conic representable, then (30) and (31) can also be solved using conic optimization software. Otherwise, one can also consider using solvers such as CVXOPT and Ipopt. In practice, we have observed that the ellipsoid method and conic solvers are the most promising ones for the problems (30) and (31).

The consistency of the SAA estimation method, both in objective value and solutions, is established in Theorem 5.4 of Shapiro et al. [2009], which requires the assumption that the optimal solutions of the true stochastic problems (7) and (8) must be contained in a compact set, for any nominal distribution $\mathbb{P}$. In the following proposition, we show that this is indeed the case.

Proposition 5. *Let $X \in L^1(\mathbb{P}_0)$, and $\phi_1, \phi_2 \in \Phi_0$ be lower semi-continuous. Suppose that condition* (10) *holds. Then, the decision variables θ_1, θ_2 in the dual problems* (7) *and* (8) *can be restricted to a compact set, without changing the optimum.*

Together with Proposition 5 and Theorem 5.4 of Shapiro et al. [2009], we can now state the consistency of the SAA method for robust OCE risk measures.

Theorem 3. *Let $X \in L^1(\mathbb{P}_0)$ and ϕ_1^*, ϕ_2^* be proper, non-decreasing convex functions that satisfy the following conditions*

1. $\mathrm{dom}(\phi_1^*) = \mathrm{dom}(\phi_2^*) = \mathbb{R}$
2. $\phi_2^*(s) > s$, $\phi_1^*(s) > s$, *for all* $s \neq 0$ *and* $\phi_1^*(0) = \phi_2^*(0) = 0$.

Assume furthermore that there exists a $(\theta_1^0, \theta_2^0, \lambda_0)$ such that for all $(\theta_1, \theta_2, \lambda)$ in a neighborhood of $(\theta_1^0, \theta_2^0, \lambda_0)$, we have

$$\mathbb{E}_{\mathbb{P}_0}\left[\lambda\phi_2^*\left(\frac{\phi_1^*(\theta_2 - X) + \theta_1}{\lambda}\right)\right] < \infty.$$

Then, as the sample size $N \to \infty$, the SAA approximations (30) *and* (31) *converge both in optimal objective value and in optimal solutions respectively to those of the true problems* (7) *and* (8)*, with probability 1.*

6.1 Importance Sampling

Although theoretically the SAA method converges, it might be very slow in practice due to a lack of data in the tails of the nominal distribution $\mathbb{P}_0$ of X. As noted by Smith and Winkler [2006], empirical risk minimization often underestimates the true risk. If we have a closed-form of the nominal density function f_0, then importance sampling can also be employed to improve the performance of SAA through a change of measure. For example, let g be the density function of a measure $\tilde{\mathbb{P}}_0$ such that $\mathbb{P}_0 \ll \tilde{\mathbb{P}}_0$, and let $Y_1, \ldots, Y_N$ be i.i.d. samples of $\tilde{\mathbb{P}}_0$. Then, through a multiplication of a likelihood factor, we can also express the SAA of (31) as

$$\begin{aligned} \min_{\substack{\lambda \geq 0 \\ \theta_1, \theta_2 \in \mathbb{R}}} \; & -\theta_1 - \theta_2 + \lambda r \\ & + \frac{1}{N}\sum_{i=1}^{N} \lambda\phi_2^*\left(\frac{\phi_1^*(\theta_2 - Y_i) + \theta_1}{\lambda}\right)\frac{f(Y_i)}{g(Y_i)}. \end{aligned} \tag{32}$$

Note that the likelihood factor $f(Y_i)/g(Y_i)$ does not depend on the decision variables, and thus does not affect the tractability of (32).

6.2 Addressing Sampling Errors and Model Misspecification Using Globalized Robust Optimization

Another way to address the sampling errors of SAA is to add an extra layer of robustification. Therefore, we formulate the following *globalized* robust OCE risk measure, by introducing a third divergence ϕ_3:

$$\begin{aligned} \rho^g_{\mathrm{oce},\mathbb{P}_0}(X) = & \sup_{\{\mathbb{P}: I_{\phi_3}(\mathbb{P}, \mathbb{P}_0) \leq r\}} \; \sup_{\{\mathbb{Q}: \mathbb{Q} \ll \mathbb{P}\}} \; \sup_{\{\bar{\mathbb{Q}}: \bar{\mathbb{Q}} \ll \mathbb{Q}\}} \\ & \mathbb{E}_{\bar{\mathbb{Q}}}[-X] - I_{\phi_1}(\bar{\mathbb{Q}}, \mathbb{Q}) - I_{\phi_2}(\mathbb{Q}, \mathbb{P}). \end{aligned} \tag{33}$$

We call (33) the globalized robust OCE risk measure since it utilizes the idea of globalized robust optimization introduced by Ben-Tal et al. [2017]. We note that (33) is a robust version of (5), where extra robustness is added to address the sampling error of SAA approximation for computing a robust risk measure. In the following theorem, we show that (33) can also be reformulated as a finite-dimensional dual optimization problem.

Theorem 4. *Let $X \in L^1(\mathbb{P}_0)$ and $\phi_1, \phi_2, \phi_3 \in \Phi_0$ be lower-semicontinuous. Then, we have the duality:*

$$\begin{aligned} \rho^g_{\mathrm{oce},\mathbb{P}_0}(X) = \inf_{\substack{\lambda \geq 0 \\ \theta_1,\theta_2,\theta_3 \in \mathbb{R}}} \quad & -\theta_1 - \theta_2 - \theta_3 \\ & + \lambda r + \mathbb{E}_{\mathbb{P}_0}\left[\gamma^*(X, \theta_1, \theta_2, \theta_3, \lambda)\right], \end{aligned} \tag{34}$$

where $\gamma^*(X, \theta_1, \theta_2, \theta_3, \lambda) = \lambda \phi_3^*\left(\frac{\phi_2^*(\phi_1^*(\theta_3 - X) + \theta_2) + \theta_1}{\lambda}\right)$.

We note that the globalized robust OCE risk measure also offers a new alternative for constructing uncertainty sets to address tail misspecification. Indeed, the duality relation (34) also leads to the following integral condition that guarantees the finiteness of the globalized robust OCE risk measure under the nominal $\mathbb{P}_0$, namely that there must exist $\theta_1, \theta_2, \theta_3 \in \mathbb{R}$ and $\lambda \geq 0$ such that

$$\int_{\mathbb{R}} \lambda \phi_3^*\left(\frac{\phi_2^*(\phi_1^*(\theta_3 - x) + \theta_2) + \theta_1}{\lambda}\right) f_0(x)\mathrm{d}x < \infty. \tag{35}$$

The presence of an extra divergence ϕ_3^* in condition (35), compared to the non-globalized case (9), offers certain flexibility advantages. Instead of defining a single, more complicated divergence to tailor the tails of the conjugate functions and the nominal density, we can also use a composition of two elementary divergences $\phi_3^* \circ \phi_2^*$. For example, a composition of the KL-divergence $\phi_3^*(x) = O(e^x)$ and the polynomial divergence with $\phi_2^*(x) = O(x^{\tilde{p}})$ for some $\tilde{p} > 1$, would result in a tail behavior $\phi_3^*(\phi_2^*(-x)) = O(\exp(|x|^{\tilde{p}}))$ as $x \to -\infty$, which could be a suitable divergence for CVaR_α risk measures and a Weibull nominal distribution with shape parameter $k > 1$. The advantage of using a combination of KL-divergence ϕ_3 and a polynomial divergence ϕ_2 is that both divergences have an explicit form of ϕ and ϕ^* functions.

7 Numerical Examples

In this section, we present three numerical experiments where we utilize results in Section 3 to construct uncertainty sets for robust CVaR_α risk measures. The Python codes for each experiment are provided on https://github.com/GuanJinNL/Uncertainty-Sets-for-Robust-Risk-Measures.git.

7.1 Toy Examples

We investigate a toy example where a payoff variable X follows a distribution supported on $[-\infty, -1)$ with density $g(x) = 2/|x|^3$ for $x \leq -1$ (i.e., $-X$ is Pareto variable with scale 1 and shape 2). We aim to measure its $\mathrm{CVaR}_{0.975}$ risk at 0.975 level, aligning with the Basel III regulation. The true $\mathrm{CVaR}_{0.975}(X)$ value under g is equal to 12.649 (see Norton et al. 2021). Since the true distribution $g(x)$ is often unknown, we compute the robust $\mathrm{CVaR}_{0.975}$, which we define as in (5) by means of a ϕ-divergence uncertainty set with a radius $r > 0$. We let the nominal distribution be $f_0(x) = 2.2/|x|^{3.2}$ for $x \leq -1$, which is slightly different from $g(x)$

and thus there is model misspecification. We investigate two polynomial divergences $p = 3$ and $p = 11$, and their ability to address this model misspecification. We note that according to Table 1, both divergences will ensure a finite robust $\mathrm{CVaR}_{0.975}$ evaluation. However, the polynomial divergence set with $p = 3$ contains $g(x)$ for r sufficiently large, whereas for $p = 11$ it does not contain $g(x)$ for any $r > 0$.[5]. Therefore, we expect the robust $\mathrm{CVaR}_{0.975}$ risk measure with $p = 11$ tends to underestimate the true $\mathrm{CVaR}_{0.975}$ risk.

For each radius r considered, we generate 10 times 500 samples from the nominal distribution f_0 and calculate the sample average approximation of the robust $\mathrm{CVaR}_{0.975}$ value for the two polynomial divergences $p = 3$ and $p = 11$. The SAA of the robust $\mathrm{CVaR}_{0.975}$ with polynomial divergence is given by the following conic optimization problem (31), where $\phi_1^*(s) = \max\{s/(1-\alpha), 0\}$ and $\phi_2^*(s) = \max\{1 + s(p-1), 0\}^{\frac{p}{p-1}} /p - 1/p$. The results are displayed in Table 7.1. As we can observe, the polynomial divergence uncertainty set with $p = 3$ is able to account for model misspecification when $r > 0.02$, where the robust $\mathrm{CVaR}_{0.975}$ value provides an upper estimate on the true $\mathrm{CVaR}_{0.975}$. On the other hand, the polynomial divergence $p = 11$ is too restrictive and still underestimates the true risk.

Radius	Robust $\mathrm{CVaR}_{0.975}$ ($p = 3$)	Robust $\mathrm{CVaR}_{0.975}$ ($p = 11$)
0.001	9.445	9.043
0.005	10.589	9.448
0.01	11.320	9.652
0.02	12.212	9.874
0.05	13.278	10.180
0.1	14.190	10.417

Table 3: SAA of Robust $\mathrm{CVaR}_{0.975}$ risk measure with the polynomial divergence of degree $p = 3$ and $p = 11$, calculated for 500 samples drawn from the nominal f_0, for each radius r. The SAA values are taken as the median value over 10 repetitions, for each r.

In a second experiement, we also investigate the effect on SAA of robust $\mathrm{CVaR}_{0.975}$ risk measure, when we choose a ϕ_2-divergence that does not ensure a finite evaluation. According to Table 1, we have infinite evaluation under f_0, when we use the KL-divergence. Therefore, we calculate the SAA (31) for both the polynomial divergence of $p = 3$ and the KL-divergence across different sample sizes generated from f_0, where we fix $r = 0.05$. The results are shown in Table 4. As we can see, the SAA of the robust $\mathrm{CVaR}_{0.975}$ risk measure defined by the KL-divergence provides a much more conservative upper estimation on true $\mathrm{CVaR}_{0.975}(X)$ than the one defined by the polynomial divergence. To make this difference even more visible, we use importance sampling to draw more data from the tail by sampling under the Pareto distribution $\tilde{g}(x) = 1/x^2$. The results are given in Table 5. Since we are sampling from the tails, it is expected that the values in Table 5 are larger than the ones in Table 4. However, in contrast to the polynominal divergence, the SAA of the robust $\mathrm{CVaR}_{0.975}$ risk measure defined by the KL-divergence now provides overly-conservative upper bounds on the exact $\mathrm{CVaR}_{0.975}(X)$ value. Moreover, it is sensitive to extreme values, as is displayed by the drastic increase in SAA values when the sample size changes from 2500 to 3000. This is because the robust $\mathrm{CVaR}_{0.975}(X)$ risk measure under the KL-divergence is in fact infinite.

[5]Indeed, $g(x)/f_0(x) = O(|x|^{0.2})$. For $\phi_2(t) = O(t^p)$, $\int_{\infty}^{-1} \phi_2\left(g(x)/f_0(x)\right) f_0(x)dx < \infty$ if and only if $p < 11$.

Sample Size	Robust $\text{CVaR}_{0.975}$ Polynomial	Robust $\text{CVaR}_{0.975}$ KL
500	11.388	14.765
1000	11.452	14.650
1500	12.748	18.349
2000	13.298	18.775
2500	13.250	19.264
3000	17.439	44.461
3500	17.202	43.343
4000	16.713	42.267
4500	16.611	41.507
5000	16.161	40.680
5500	16.306	40.219
6000	15.967	39.580

Table 4: SAA for robust $\text{CVaR}_{0.975}$ risk measure with radius $r = 0.05$, computed for a range of sample sizes drawn from the nominal distribution $f_0(x) = 2.2/|x|^{3.2}, x \leq -1$. The differences are compared for ϕ_2 in (31) chosen to be the polynomial divergence with $p = 3$ and the KL-divergence.

Sample Size	Robust $\text{CVaR}_{0.975}$ Polynomial	Robust $\text{CVaR}_{0.975}$ KL
500	21.870	122.423
1000	21.471	116.577
1500	21.911	175.000
2000	22.271	178.624
2500	21.377	186.414
3000	21.548	1400.889
3500	21.693	1384.170
4000	21.649	1369.866
4500	21.666	1356.501
5000	21.669	1346.487
5500	21.703	1336.832
6000	21.796	1328.153

Table 5: SAA for Robust $\text{CVaR}_{0.975}$ risk measure with radius $r = 0.05$, computed for samples drawn using importance sampling under the sampling distribution $\tilde{g}(x) = 1/x^2, x \leq -1$. The differences are compared for ϕ_2 being the polynomial divergence with $p = 3$ and the KL-divergence.

7.2 Discrete Delta Hedging of Black-Scholes Model

We consider a discrete delta hedging strategy for a call option under the Black-Scholes framework. In this model, the stock price S follows a geometric Brownian motion with drift μ_S and volatility $\sigma_S > 0$. The call option $C(S_T)$ with a strike price K and a maturity time T, has a final payoff given by $C(S_T) = \max\{S_T - K, 0\}$.

The discrete delta hedging strategy is implemented by constructing a self-financing portfolio that aims to replicate the payoff of a call option. At each time step $t_i = iT/n$, the value of the portfolio is the sum of the stock value $\text{Stock}(i)$ and the cash value $\text{Cash}(i)$ that grows with an interest rate r_f. The portfolio is self-financing, which means that any rebalancing of the shares in one asset is only achieved by buying or selling from the other asset. Let δ_{S,t_i} denote the number of shares of stocks held at time $[t_{i-1}, t_i)$, which we calculate using the Black-Scholes formula as in the continuous setting. The dynamics of the stock value and the cash value are determined by the following recursive relation:

$$\begin{aligned}\text{Stock}(t_i) &= S_{t_i} \cdot \delta_{S,t_i} \\ \text{Cash}(t_i) &= \exp\left(\frac{r_f}{n}\right) \text{Cash}(t_{i-1}) - S_{t_i} \cdot (\delta_{S,t_i} - \delta_{S,t_{i-1}}) \\ &\qquad - k_0 - k \cdot |\delta_{S,t_i} - \delta_{S,t_{i-1}}| \cdot S_{t_i},\end{aligned} \tag{36}$$

where we have also added the transaction cost, $k_0 + k \cdot |\delta_{S,t_i} - \delta_{S,t_{i-1}}| \cdot S_{t_i}$, which consists of a fixed amount k_0 and a cost that is proportional to the volumes of the trade executed. We define the initial investment of $\text{Stock}(0), \text{Cash}(0)$ to be the portfolio that determines the price of the call option by the Black-Scholes formula, without the transaction cost. The hedging error, which we aim to measure, is defined as the absolute difference between the call option payoff and the payoff of this hedging strategy, evaluated at the maturity date T:

$$X_{\text{error}} = |C(S_T) - \text{Stock}(T) - \text{Cash}(T)| .$$

In this experiment, we will measure the risk of the hedging error X_{error} both in the nominal (when $r = 0$) and the robust setting (when $r > 0$). Measurements of model risk of hedging error have been previously studied by Kruse et al. [2019] and Glasserman and Xu [2014]. In contrast to these studies, where they examined the expected hedging error, we measure the nominal and robust Conditional Value-at-Risk of X_{error} at $\alpha = 0.95$.

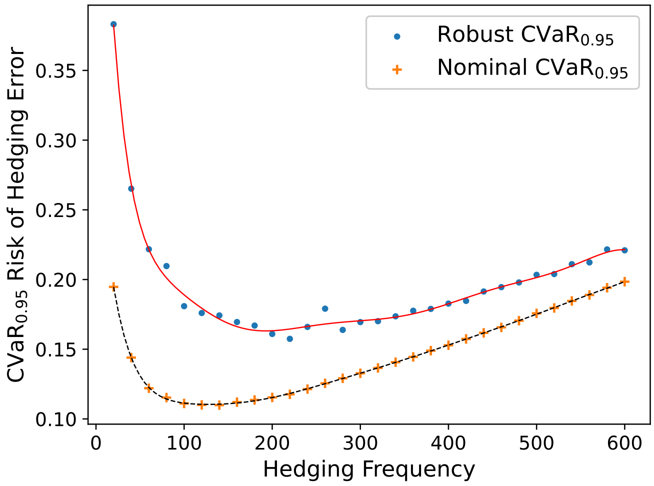
To choose a proper ϕ_2-divergence for measuring ambiguity, we first note that $0 < \delta_{S_{t_i}} < 1$ by the Black-Scholes formula. Hence, according to the dynamics given in (36), one can show that there exists a constant $C > 0$, such that X_{error} satisfies (11):

$$|X_{\text{error}}| \leq C \left(1 + \sum_{i=1}^{n} S_{t_i}\right).$$

Therefore, we can choose the divergence according to the nominal distributions of the risk factors S_{t_i}, which, for each i, follows the log-normal distribution with $\sigma_i = i\sigma_S/\sqrt{n}$. Hence, we can set the volatility parameter of the divergence in (15) to be $\sigma = \max_i \sigma_i = \sigma_S$. Furthermore, we choose $d = 2$, so that our ϕ_2-divergence set may contain distributions with tails as heavy as $|x|^{-2}$.

As a comparison, we choose the same parameters as in Kruse et al. [2019], where $\mu_S = 0.05, \sigma_S = 0.3, r_f = 0.01, T = 1, S_0 = 1, K = 1, k_0 = 0.0002$ and $k = 0.005$. We vary the hedging frequency n and simulate for each frequency 50000 hedging error data points. Then, we calculate the SAA approximation of the $\text{CVaR}_{0.95}$ value of the hedging error, as well as that of the robust $\text{CVaR}_{0.95}$ value, with the radius in (5) set to be $r = 0.1$. The results are given in Figure 1. As we can see, both the nominal and the robust tail risk of the hedging error follow a trend of a U-shape as the hedging frequency increases, suggesting a tradeoff between transaction cost and hedging error. We see that the optimal hedging frequency in the nominal case is around 100, slightly larger than the case investigated in Kruse et al. [2019], where the expected hedging error

Figure 1: Absolute hedging error measured by the nominal and robust Conditional Value-at-Risk at confidence level $\alpha = 0.95$, as a function of the hedging frequency.



was measured. In the robust case, the optimal frequency is increased to around 200, which is two times larger from the nominal case. Moreover, the difference between robust risk value and nominal risk value is much larger at low hedging frequencies, and starts decreasing at higher frequency until reaching an equilibrium for frequency higher than 400. This shows that for high hedging frequency, the hedging portfolio is more robust against model uncertainty.

7.3 Risk-Averse Newsvendor Minimization

In the newsvendor problem, the stock owner must decide a priori the amount of products that needs to be ordered, before the demand is realized. Let d be a realized demand value of the uncertain demand variable D, c be the cost of one unit of order, $v > c$ be the selling price, $s < c$ be the salvage value per unsold item returned to the factory, and l be the loss per unit of unmet demand. If the stock owner decides to order y number of items of the product, then the profit function is given by:

$$\begin{aligned} \pi(y, d) \triangleq\ & v \min\{d, y\} + s \max\{y - d, 0\} \\ & - l \max\{d - y, 0\} - cy. \end{aligned}$$

In the classical newsvendor problem, one minimizes the expected loss $\mathbb{E}[-\pi(y, D)]$ with respect to y. In the risk-averse setting, Conditional Value-at-Risk $\mathrm{CVaR}_\alpha(\pi(y, D))$ is minimized, where α controls the risk-aversion of the decision maker. Without ambiguity, the risk-averse newsvendor CVaR minimization problem has a closed-form solution (see Gotoh and Takano 2007), which is given by:

$$\begin{aligned} y^*_{\mathrm{nom}} = & \frac{E + V}{E + U} F^{-1}\left(\frac{U(1-\alpha)}{E + U}\right) \\ & + \frac{U - V}{E + U} F^{-1}\left(\frac{E\alpha + U}{E + U}\right) \end{aligned} \tag{37}$$

where F^{-1} is the quantile function of the demand, and $E = c - s, U = v + l - c, V = v - c$.

We investigate the effect of model uncertainty on the optimal number of orderings, by solving the following robust $\text{CVaR}_{0.95}(\pi(y, D))$ newsvendor minimization problem using the ellipsoid method, for $N = 50000$:

$$\begin{aligned}\min_{\substack{\lambda \geq 0 \\ y, \theta_1, \theta_2 \in \mathbb{R}}} \quad & -\theta_1 - \theta_2 + \lambda r \\ & + \frac{1}{N} \sum_{i=1}^{N} \lambda \phi_2^* \left(\frac{\phi_1^*(\theta_2 - \pi(y, D_i)) + \theta_1}{\lambda} \right),\end{aligned} \tag{38}$$

where $D_1, \ldots, D_{50000}$ are demand samples drawn from the heavy-tailed log-normal distribution with mean $\mu_D = 0$ and standard deviation $\sigma_D = 1$, and $\phi_1^*(s) = \max\{s/(1-\alpha), 0\}$. We solve (38) for a range of radii $r > 0$, where a larger r represents a higher degree of model uncertainty. To choose a proper ϕ_2-divergence, we note that $\pi(y, D)$ is a piecewise-linear concave function in D, and thus $|\pi(y, D)|$ also satisfies (11) for some constant $C > 0$.[6] Hence, we may choose the divergence (15) with $\sigma = \sigma_D = 1$, and $p = 2, d = 2$ similarly as in the Black-Scholes example. For the remaining parameters, we choose $v = 8, c = 4, s = 2, l = 4$. Then, one can calculate using (37) that the optimal decision without model uncertainty is given by $y^*_{\text{nom}} = 4.2$. The results are displayed in Table 6, which shows that a higher model uncertainty increases the optimal number of orderings. This is because for each unmet demand, the stock owners pays a price of $l(y - d)$. When the demand is heavy-tailed, a large loss in profit can occur when extreme high demand values are realized. Therefore, the optimal ordering in the robust newsvendor problem (38) is higher than the nominal solution (37), since it must hedges against the potential large loss due to unmet demand.

r	Optimal Ordering
0.001	4.514
0.005	5.093
0.01	5.488
0.02	6.168
0.05	7.557
0.1	8.894
0.2	11.506
0.4	16.244
0.5	19.115

Table 6: The effect of model uncertainty on the optimal ordering of robust newsvendor optimization problem (38). The larger the parameter $r > 0$, the greater the model uncertainty. The parameters of the newsvendor problem are given by $v = 8, c = 4, s = 2, l = 4$.

[6] for $C = \max\{l, |v - s|\} \cdot \max\left\{1, \frac{\max\{|v+l-c|, |s-c|\}}{\max\{l, |v-s|\}} |y|\right\}$.

8 Concluding Remarks

In this paper, we characterized a broad class of robust risk measures in terms of two ϕ-divergences (ϕ_1, ϕ_2), where ϕ_1 specifies the risk measures and ϕ_2 specifies the ambiguity set. We have shown that this framework is not only computationally tractable, but also offers a blueprint to systematically construct ϕ-divergence uncertainty sets that are tailored to a given risk measure and nominal model. In addition, this characterization extends naturally to include higher level of uncertainty using globalized robust optimization. For many robust risk measures, we can also elicit the divergence functions (ϕ_1, ϕ_2), which serve as an extra tool for calibrating the uncertainty set.

References

Daniel Bartl, Samuel Drapeau, and Ludovic Tangpi. Computational aspects of robust optimized certainty equivalents and option pricing. *Mathematical Finance*, 30(1):287–309, 2020.

Aharon Ben-Tal and Adi Ben-Israel. A recourse certainty equivalent for decisions under uncertainty. *Annals of Operations Research*, 30:1–44, 1991.

Aharon Ben-Tal and Marc Teboulle. Penalty functions and duality in stochastic programming via ϕ-divergence functionals. *Mathematics of Operations Research*, 12(2):224–240, 1987.

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

Aharon Ben-Tal, Adi Ben-Israel, and Marc Teboulle. Certainty equivalents and information measures: Duality and extremal Principles. *Journal of Mathematical Analysis And Applications*, 157(1):211–236, 1991.

Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Princeton University Press, Princeton, NJ, 2009.

Aharon Ben-Tal, Dick den Hertog, Anja de Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

Aharon Ben-Tal, Ruud Brekelmans, Dick den Hertog, and J. Vial. Globalized robust optimization for nonlinear uncertain inequalities. *INFORMS Journal on Computing*, 29(2):197–376, 2017.

Carole Bernard, Silvana M. Pesenti, and Steven Vanduffel. Robust distortion risk measures. *Mathematical Finance*, 00:1–45, 2023.

Dimitris Bertsimas and David B. Brown. Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6):1483–1495, 2009.

Robert G. Bland, Donald Goldfarb, and Michael J. Todd. The ellipsoid method: A survey. *Operations Research*, 29(6):1039–1091, 1981.

Thomas Breuer and Imre Csiszár. Measuring distibution model risk. *Mathematical Finance*, 26 (2):395–411, 2016.

Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3(1):146–158, 1975.

John C. Duchi, Peter W. Glynn, and Hongseok Namkoong. Statistics of robust optimization: a generalized empirical likelihood approach. *Mathematics of Operations Research*, 46(3):946–969, 2021.

Rick Durrett. *Probability: Theory and Examples*, volume 5. Cambridge Series in Statistical and Probabilistic Mathematics, 2019.

Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1):115–166, 2018.

Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.

Paul Glasserman and Xingbo Xu. Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58, 2014.

Jun-ya Gotoh and Yuichi Takano. Newsvendor solutions via conditional value-at-risk minimization. *European Journal of Operational Research*, 179(1):80–96, 2007.

Lars Peter Hansen and Thomas J. Sargent. *Robustness.* Princeton University Press, stu-student edition, 2008.

Thomas Kruse, Judith C. Schneider, and Nikolaus Schweizer. Technical note—the joint impact of F-divergences and reference models on the contents of uncertainty sets. *Operations Research*, 67(2):428–435, 2019.

Thomas Kruse, Judith C. Schneider, and Nikolaus Schweizer. A toolkit for robust risk assessment using F-divergences. *Management Science*, 67(10):6529–6552, 2021.

Henry Lam. Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. *Operations Research*, 67(4):1090–1105, 2019.

Henry Lam and Clementine Mottet. Tail analysis without parametric models: A worst-case perspective. *Operations Research*, 65(6):1696–1711, 2017.

David G. Luenberger. *Optimization by Vector Space Methods.* Wiley Professional Paperback Series, 1969.

Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.

Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory.* Oxford University Press, 1995.

Matthew Norton, Valentyn Khokhlov, and Stan Uryasev. Calculating CVaR and bPOE for common probability distributions with application to portfolio optimization and density estimation. *Annals of Operations Research*, 299:1281–1315, 2021.

Leandro Pardo. *Statistical Inference Based on Divergence Measures.* Chapman & Hall/ CRC Boca Raton, 2006.

Ralph Tyrrell Rockafellar. Integrals which are convex functionals. *Pacific Journal of Mathematics*, 24(3):525–539, 1968.

Ralph Tyrrell Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

Ralph Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis.* Springer Verlag, Berlin-Heidelberg-New York, 2004.

Judith C. Schneider and Nikolaus Schweizer. Robust measurement of (heavy-tailed) risks: Theory and implementation. *Journal of Economics Dynamics and Control*, 61(1-2):183–203, 2015.

Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*, volume 9. Society for Industrial and Applied Mathematics: Philadelphia, PA, 2009.

James E. Smith and Robert L. Winkler. The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322, 2006.

Bart P.G. van Parys, Peyman Mohajerin Esfahani, and Daniel Kuhn. From data to decisions: Distributionally robust optimization is optimal. *Management Science*, 67(6):3387–3402, 2021.

Appendix

EC.1 Proofs

Proof of Theorem 1. Due to the absolute continuity relations $\bar{\mathbb{Q}} \ll \mathbb{Q} \ll \mathbb{P}_0$, there exists probability density functions $g, \bar{g}$ with respect to the measure $\mathbb{P}_0$, such that the optimization problem:

$$\sup_{\substack{\{\mathbb{Q}:\mathbb{Q}\ll\mathbb{P}_0\} \\ \{\bar{\mathbb{Q}}:\bar{\mathbb{Q}}\ll\mathbb{Q}\}}} \mathbb{E}_{\bar{\mathbb{Q}}}[-X] - I_{\phi_1}(\bar{\mathbb{Q}}, \mathbb{Q}) - I_{\phi_2}(\mathbb{Q}, \mathbb{P}_0),$$

can be formulated as an optimization problem over density functions:

$$-\inf_{\substack{g,\bar{g}\geq 0, \\ \int_\Omega g \mathrm{d}\mathbb{P}_0 = 1, \int_\Omega \bar{g} \mathrm{d}\mathbb{P}_0 = 1}} \int_\Omega \beta(\omega, g(\omega), \bar{g}(\omega)) \mathrm{d}\mathbb{P}_0(\omega),$$

where

$$\beta(\omega, s, t) = X(\omega)t + s\phi_1\left(\frac{t}{s}\right) + \phi_2\left(s\right), \qquad s, t \geq 0.$$

Note that $\beta(\omega, s, t) = +\infty$ for $s < 0$ or $t < 0$ since both ϕ_1 and ϕ_2 take the value $+\infty$ on the negative real line. By assumption, ϕ_1 and ϕ_2 are convex lower-semicontinuous functions that have effective domain with non-empty interior. Therefore, $\beta(\omega, s, t)$ is a normal convex integrand (in the sense of Lemma 2 of Rockafellar 1968 and Proposition 14.39 of Rockafellar and Wets 2004). Furthermore, there exists a feasible point, namely the constant function $(g, \bar{g}) \equiv (1, 1)$: $\int_\Omega \beta(\omega, 1, 1) \mathrm{d}\mathbb{P}_0(\omega) = \mathbb{E}_{\mathbb{P}_0}[X] + \phi_1(1) + \phi_2(1) < \infty$. Therefore, applying Theorem EC.2.2 yields

$$\inf_{\substack{g,\bar{g}\geq 0, \\ \int_\Omega g \mathrm{d}\mathbb{P}_0 = 1, \int_\Omega \bar{g} \mathrm{d}\mathbb{P}_0 = 1}} \int_\Omega \beta(\omega, g(\omega), \bar{g}(\omega)) \mathrm{d}\mathbb{P}_0(\omega) = \sup_{\theta_1, \theta_2 \in \mathbb{R}} \theta_1 + \theta_2 - \int_\Omega \beta^*(\omega, \boldsymbol{\theta}) \mathrm{d}\mathbb{P}_0(\omega),$$

where

$$\begin{aligned}
\beta^*(\omega, \boldsymbol{\theta}) &= \sup_{s,t\geq 0} \theta_1 s + \theta_2 t - X(\omega)t - s\phi_1\left(\frac{t}{s}\right) - \phi_2\left(s\right) \\
&= \sup_{s\geq 0} \theta_1 s - \phi_2\left(s\right) + \sup_{t\geq 0} (\theta_2 - X(\omega))t - s\phi_1\left(\frac{t}{s}\right) \\
&= \sup_{s\geq 0} \theta_1 s - \phi_2\left(s\right) + s\phi_1^*(\theta_2 - X(\omega)) \\
&= \phi_2^*(\theta_1 + \phi_1^*(\theta_2 - X(\omega))).
\end{aligned}$$

Finally, to obtain the worst-case density, we have that $\mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*(\phi_1^*(\theta_2 - X) + \theta_1)\right] < \infty$ for some (θ_1, θ_2). Hence, the dual problem (7) is finite and there exists a dual solution (θ_1^*, θ_2^*) by Theorem EC.2.1. If furthermore that $(\theta_1^*, \theta_2^*) \in \mathrm{int}(\mathcal{F}(\boldsymbol{\theta}))$, then $\mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*(\phi_1^*(\theta_2 - X) + \theta_1)\right]$ is differentiable at (θ_1^*, θ_2^*) by Theorem EC.2.3. Therefore, the partial derivatives $(g^*(\omega), \bar{g}^*(\omega)) \triangleq (\partial_{\theta_1}\beta^*(\omega, \boldsymbol{\theta}), \partial_{\theta_2}\beta^*(\omega, \boldsymbol{\theta}))$ evaluated at $\boldsymbol{\theta}^*$, for each $\omega \in \Omega$, satisfies the first-order condition of the dual problem (7)

$$\int_\Omega g^*(\omega)\mathrm{d}\mathbb{P}_0(\omega) = 1, \quad \int_\Omega \bar{g}^*(\omega)\mathrm{d}\mathbb{P}_0(\omega) = 1.$$

Hence, it follows from Theorem EC.2.4 that the worst-case densities are the partial derivatives.

Similarly, the optimization problem

$$\sup_{\{\mathbb{Q}: I_{\phi_2}(\mathbb{Q},\mathbb{P}_0)\leq r\}} \sup_{\{\bar{\mathbb{Q}}:\bar{\mathbb{Q}}\ll\mathbb{Q}\}} \mathbb{E}_{\bar{\mathbb{Q}}}[-X] - I_{\phi_1}(\bar{\mathbb{Q}},\mathbb{Q})$$

can be rewritten to the following optimization problem over density functions with respect to the measure $\mathbb{P}_0$:

$$-\inf_{\substack{g,\bar{g}\geq 0,\\ \int_\Omega g\mathrm{d}\mathbb{P}_0=1, \int_\Omega \bar{g}\mathrm{d}\mathbb{P}_0=1\\ \int_\Omega \phi_2(g(\omega))\mathrm{d}\mathbb{P}_0(\omega)\leq r}} \int_\Omega X(\omega)\bar{g}(\omega) + g(\omega)\phi_1\left(\frac{\bar{g}(\omega)}{g(\omega)}\right)\mathrm{d}\mathbb{P}_0(\omega) =: -J_0. \tag{EC.39}$$

Assume without loss of generality that $J_0 > -\infty$. Then, we also have that $J_0 < \infty$, since $(g,\bar{g}) \equiv (1,1)$ is a feasible solution. Then, Theorem EC.2.1 implies that

$$J_0 = \sup_{\lambda\geq 0} -\lambda r + \inf_{\substack{g,\bar{g}\geq 0,\\ \int_\Omega g\mathrm{d}\mathbb{P}_0=1, \int_\Omega \bar{g}\mathrm{d}\mathbb{P}_0=1}} \int_\Omega X(\omega)\bar{g}(\omega) + g(\omega)\phi_1\left(\frac{\bar{g}(\omega)}{g(\omega)}\right) + \lambda\phi_2(g(\omega))\mathrm{d}\mathbb{P}_0(\omega).$$

We first examine the case when $\lambda = 0$, for which the above would become

$$\begin{aligned} J_0 &= \inf_{\substack{g,\bar{g}\geq 0,\\ \int_\Omega g\mathrm{d}\mathbb{P}_0=1, \int_\Omega \bar{g}\mathrm{d}\mathbb{P}_0=1}} \int_\Omega X(\omega)\bar{g}(\omega) + g(\omega)\phi_1\left(\frac{\bar{g}(\omega)}{g(\omega)}\right)\mathrm{d}\mathbb{P}_0(\omega) \\ &= \operatorname{ess\,inf}(X). \end{aligned}$$

On the other hand, since $0\phi_2^*(s/0) = 0$ for $s \leq 0$ and $+\infty$ otherwise, we have

$$\begin{aligned} &\sup_{\theta_1,\theta_2\in\mathbb{R}} \theta_1 + \theta_2 - \mathbb{E}_{\mathbb{P}_0}\left[0\phi_2^*\left(\frac{\phi_1^*(\theta_2 - X) + \theta_1}{0}\right)\right] \\ &= \sup_{\theta_1,\theta_2\in\mathbb{R}} \{\theta_1 + \theta_2 \mid \operatorname{ess\,sup} \phi_1^*(\theta_2 - X) + \theta_1 \leq 0\} \\ &= \sup_{\theta_2\in\mathbb{R}} \theta_2 - \phi_1^*(\theta_2 - \operatorname{ess\,inf}(X)) \\ &= \operatorname{ess\,inf}(X). \end{aligned}$$

Hence, the duality formula holds in the case of $\lambda = 0$. Assume now $\lambda > 0$. Define $\beta_\lambda(\omega, s, t) \triangleq X(\omega)t + s\phi_1\left(\frac{t}{s}\right) + \lambda\phi_2(s)$. Note that β_λ is only finite when $s, t \geq 0$. Then, applying Theorem EC.2.1 yields

$$J_0 = \sup_{\substack{\lambda\geq 0\\ \theta_1,\theta_2\in\mathbb{R}}} -\lambda r + \theta_1 + \theta_2 - \int_\Omega \beta_\lambda^*(\omega, \theta_1, \theta_2)\mathrm{d}\mathbb{P}_0(\omega).$$

It remains to compute $\beta^*_\lambda(\omega,\theta_1,\theta_2)$, which is equal to $\beta^*_\lambda(\omega,\theta_1,\theta_2)=\lambda\phi_2^*\left(\frac{\theta_1+\phi_1^*(\theta_2-X(\omega))}{\lambda}\right)$.

Now for the worst-case densities, we note that the domain set $\mathcal{F}(\boldsymbol{\theta},\lambda)$ is assumed to be non-empty, thus a dual solution $(\theta_1^*,\theta_2^*,\lambda^*)$ exists by Theorem EC.2.1. If the dual solution also lies in the interior of $\mathcal{F}(\boldsymbol{\theta},\lambda)$, then we have differentiability by Theorem EC.2.3, and thus the dual solution must satisfy the gradient conditions

$$\begin{aligned}
&\int_\Omega (\phi_2^*)'\left(\frac{\theta_1^*+\phi_1^*(\theta_2^*-X(\omega))}{\lambda^*}\right)\mathrm{d}\mathbb{P}_0(\omega)=1\\
&\int_\Omega (\phi_2^*)'\left(\frac{\theta_1^*+\phi_1^*(\theta_2^*-X(\omega))}{\lambda^*}\right)(\phi_1^*)'(\theta_2^*-X(\omega))\mathrm{d}\mathbb{P}_0(\omega)=1\\
&\int_\Omega \frac{\theta_1^*+\phi_1^*(\theta_2^*-X(\omega))}{\lambda^*}(\phi_2^*)'\left(\frac{\theta_1^*+\phi_1^*(\theta_2^*-X(\omega))}{\lambda^*}\right)-\phi_2^*\left(\frac{\theta_1^*+\phi_1^*(\theta_2^*-X(\omega))}{\lambda^*}\right)\mathrm{d}\mathbb{P}_0(\omega)=r,
\end{aligned}$$

where the latter condition is equivalent to $\int_\Omega \phi_2\left((\phi_2^*)'\left(\frac{\theta_1^*+\phi_1^*(\theta_2^*-X(\omega))}{\lambda^*}\right)\right)\mathrm{d}\mathbb{P}_0(\omega)=r$, due to the relation $\phi_2=\phi_2^{**}$. Therefore, the partial derivatives $(g^*(\omega),\bar{g}^*(\omega))\triangleq(\partial_{\theta_1}\beta^*_\lambda(\omega,\boldsymbol{\theta}),\partial_{\theta_2}\beta^*_\lambda(\omega,\boldsymbol{\theta}))$ evaluated at $(\boldsymbol{\theta}^*,\lambda^*)$, are indeed feasible density functions of the primal problem (EC.39).

Finally, the partial derivatives are indeed also the worst-case densities, since

$$\begin{aligned}
J_0&=-\lambda^*r+\theta_1^*+\theta_2^*-\int_\Omega \beta^*_{\lambda^*}(\omega,\theta_1^*,\theta_2^*)\mathrm{d}\mathbb{P}_0(\omega)\\
&\overset{(*)}{=}-\lambda^*r+\int_\Omega X(\omega)\bar{g}^*(\omega)+g^*(\omega)\phi_1\left(\frac{\bar{g}^*(\omega)}{g^*(\omega)}\right)+\lambda^*\phi_2(g^*(\omega))\mathrm{d}\mathbb{P}_0(\omega)\\
&=\int_\Omega X(\omega)\bar{g}^*(\omega)+g^*(\omega)\phi_1\left(\frac{\bar{g}^*(\omega)}{g^*(\omega)}\right)\mathrm{d}\mathbb{P}_0(\omega),
\end{aligned}$$

where $(*)$ follows from the worst-case densities of (6). □

Proof of Corollary 1. Recall that from Theorem 1 we have,

$$\sup_{\{\mathbb{Q}:\mathbb{Q}\ll\mathbb{P}_0\}}\mathbb{E}_{\bar{\mathbb{Q}}}[-X]-I_{\phi_1}(\bar{\mathbb{Q}},\mathbb{Q})-I_{\phi_2}(\mathbb{Q},\mathbb{P}_0)=\inf_{\theta_1,\theta_2\in\mathbb{R}}-\theta_1-\theta_2+\mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*(\phi_1^*(\theta_2-X)+\theta_1)\right].\tag{EC.40}$$

Hence, if $\mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*(\phi_1^*(\theta_2-X)+\theta_1)\right]=\infty$ holds for all θ_1,θ_2, then the infimum on the right-hand side of (EC.40) is indeed always ∞. If $\mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*(\phi_1^*(\theta_2-X)+\theta_1)\right]<\infty$ for some $\theta_1,\theta_2\in\mathbb{R}$, then this infimum is bounded away from $-\infty$. Similarly, (9) follows from (8). □

Proof of Proposition 1. Apply Corollary 1 and note that

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*\left(\theta_1+\phi_1^*(\theta_2-X)\right)\right] &\le \mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*\left(\theta_1+\phi_1^*\left(\theta_2+C(1+\sum_{i=1}^m |Z_i|)\right)\right)\right] \\
&= \mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*\left(\theta_1+\phi_1^*\left(\frac{1}{m}\sum_{i=1}^m \theta_2+C(1+m|Z_i|)\right)\right)\right] \\
&\le \mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*\left(\theta_1+\frac{1}{m}\sum_{i=1}^m \phi_1^*\left(\theta_2+C(1+m|Z_i|)\right)\right)\right] \\
&= \mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*\left(\frac{1}{m}\sum_{i=1}^m \theta_1+\phi_1^*\left(\theta_2+C(1+m|Z_i|)\right)\right)\right] \\
&\le \frac{1}{m}\sum_{i=1}^m \mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*\left(\theta_1+\phi_1^*\left(\theta_2+C(1+m|Z_i|)\right)\right)\right],
\end{aligned}
$$

where we used Jensen's inequality and the monotonicity of the conjugate functions ϕ_1^*, ϕ_2^*. $\square$

Proof of Proposition 2. By the assumptions on ψ, we have that $\phi^*(s)$ as defined in (13) is an increasing, continuous convex function on $\mathbb{R}$ with $\phi^*(0) = 0$ and $(\phi^*)'(0) = 1$. By definition, we have that $\phi(t) \triangleq \sup_{s\in\mathbb{R}} st - \phi^*(s)$. Therefore, ϕ is convex. We also have that $\phi(t) = +\infty$ for $t < 0$, since $\phi^*(s) < 0$ for $s < 0$ and thus $\phi(t) \ge \sup_{s\le 0} st = \infty$, if $t < 0$. Moreover, ϕ is non-negative on $[0, \infty)$ with $\phi(1) = 0$. Indeed, $\phi^*(0) = 0$ ensures that $\phi(t) \ge 0 \cdot t - \phi^*(0) = 0$, and $(\phi^*)'(0) = 1$ implies that $\phi(1) = 0$ (Theorem 23.5, Rockafellar 1970). Therefore, $\phi \in \Phi_0$ and is thus a ϕ-divergence function.

If ψ is strictly convex and differentiable, then ϕ^* is by construction strictly convex and differentiable on $\mathbb{R}$. Theorem 26.3 of Rockafellar [1970] implies that its conjugate $\phi \triangleq (\phi^*)^*$ is essentially smooth. Therefore, ϕ is differentiable on the interior of its effective domain, denoted as $\mathrm{dom}(\phi)$. By Corollary 23.5.1 and Theorem 23.4 of Rockafellar [1970], we have that $\mathrm{dom}(\phi)$ contains the image of the derivative $(\phi^*)'$, which is the set $(0, \infty)$, since by assumption ψ' is continuous and strictly increasing with $\lim_{s\to\infty} \psi'(s) = \infty$, and that $(\phi^*)'(s) = \exp(s)$, for $s \le 0$. Furthermore, $\phi(0) = 1 < \infty$, since $\phi(t) = t\log t - t + 1$ on $t \le 1$. Hence, $\mathrm{dom}(\phi) = [0, \infty)$. Moreover, Corollary 23.5.1 of Rockafellar [1970] implies that $\phi' = ((\phi^*)')^{-1}$. $\square$

The proof of Proposition 3 relies on the following Lemma.

Lemma EC.1.1. *Let f, g be two strictly increasing functions. If for some $y_0 \in \mathbb{R}$ that $f^{-1}(y) \le g^{-1}(y)$ for all $y \ge y_0$, then $f(x) \ge g(x)$, for all $x \ge f^{-1}(y_0)$.*

Proof of Lemma EC.1.1. We prove by contradiction. Suppose there is a $x_0 \ge f^{-1}(y_0)$ such that $f(x_0) < g(x_0)$. Then, $f(x_0) \ge y_0$. Since f, g are increasing, we have that f^{-1}, g^{-1} are also increasing. Therefore,

$$x_0 = f^{-1}(f(x_0)) \le g^{-1}(f(x_0)) < g^{-1}(g(x_0)) = x_0,$$

which is a contradiction. $\square$

Proof of Proposition 3. Since ϕ_2' is strictly increasing and non-negative for $t \geq 1$, we have that for $t \geq 1$,

$$\phi_2(t) = \int_1^t \phi_2'(s)\mathrm{d}s \leq (t-1)\phi_2'(t) \leq t\phi_2'(t).$$

Since $(\psi_1')^{-1}(y) \leq (\phi_2')^{-1}(y)$ for all $y \geq y_0$, Lemma EC.1.1 implies that for all $t \geq t_0$, where $t_0 := \max\{(\psi_1')^{-1}(y_0), 1\}$, we have that $\psi_1'(t) \geq \phi_2'(t) \geq 0$. Therefore, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} \phi_2\left(\frac{g(x)}{f_0(x)}\right) f_0(x)\mathrm{d}x \\
&= \int_{x:\frac{g(x)}{f_0(x)}\leq 1} \phi_2\left(\frac{g(x)}{f_0(x)}\right) f_0(x)\mathrm{d}x + \int_{x:\frac{g(x)}{f_0(x)}>1} \phi_2\left(\frac{g(x)}{f_0(x)}\right) f_0(x)\mathrm{d}x \\
&\leq \int_{x:\frac{g(x)}{f_0(x)}\leq 1} \phi_2\left(\frac{g(x)}{f_0(x)}\right) f_0(x)\mathrm{d}x + \int_{x:\frac{g(x)}{f_0(x)}>1} g(x)\phi_2'\left(\frac{g(x)}{f_0(x)}\right) \mathrm{d}x
\end{aligned}$$

Since ϕ_2 is bounded on $[0,1]$, we only need to bound the second integral. By the limit assumption (14), there exists a u_0, such that for all $x \in \mathbb{R}$ with $|x| \geq u_0$, we have $\psi_2'\left(\frac{1}{f_0(x)}\right) \leq M|x|^d$, for some constant $M > 0$. We have that $f_0(x) \to 0$ as $|x| \to \infty$. Hence, there exists a $x_0 \geq u_0$ such that for all $x \in \mathbb{R}$ with $|x| \geq x_0$, we have that $\frac{1}{f_0(x)} \geq t_0$. We then split the following integral by x_0:

$$\begin{aligned}
&\int_{x:\frac{g(x)}{f_0(x)}>1} g(x)\phi_2'\left(\frac{g(x)}{f_0(x)}\right) \mathrm{d}x \\
&= \int_{x:\frac{g(x)}{f_0(x)}>1\cap\{x:|x|\leq x_0\}} g(x)\phi_2'\left(\frac{g(x)}{f_0(x)}\right) \mathrm{d}x + \int_{x:\frac{g(x)}{f_0(x)}>1\cap\{x:|x|\geq x_0\}} g(x)\phi_2'\left(\frac{g(x)}{f_0(x)}\right) \mathrm{d}x.
\end{aligned}$$

The first term is bounded by the assumption that $\frac{g}{f}$ is bounded on any compact interval, and that ϕ_2' is bounded on any compact interval as well, due to its continuity. Therefore, we only need to examine the second term. Due to the non-negativity and monotonicity of ϕ_2', ψ_2' on $[1,\infty)$, we have,

$$\begin{aligned}
\int_{x:\frac{g(x)}{f_0(x)}>1\cap\{x:|x|\geq x_0\}} g(x)\phi_2'\left(\frac{g(x)}{f_0(x)}\right) \mathrm{d}x &\leq \int_{x:|x|\geq x_0} g(x)\phi_2'\left(\frac{g(x)}{f_0(x)}\right) \mathrm{d}x \\
&\leq \int_{x:|x|\geq x_0} g(x)\phi_2'\left(\frac{1}{f_0(x)}\right) \mathrm{d}x \\
&\leq \int_{x:|x|\geq x_0} g(x)\psi_1'\left(\frac{1}{f_0(x)}\right) \mathrm{d}x \\
&\leq \int_{x:|x|\geq x_0} g(x)|x|^d\mathrm{d}x < \infty.
\end{aligned}$$

Conversely, if there exists a divergence ψ_2 such that $(\phi_2^*)'(y) \leq (\psi_2')^{-1}(y)$ for all $y \geq 0$, then we have that $\phi_2'(t) \geq \psi_2'(t)$ for all $t \geq 1$. Since $\phi_2(1) = \psi_2(1) = 0$, this also means that

$$\phi_2(t) = \int_1^t \phi_2'(s)\mathrm{d}s \geq \int_1^t \psi_2'(s)\mathrm{d}s = \psi_2(t),\ \forall t \geq 1.$$

Therefore, we have

$$\int_{x:\frac{g(x)}{f_0(x)}>1} \phi_2\left(\frac{g(x)}{f_0(x)}\right) f_0(x)\mathrm{d}x \geq \int_{x:\frac{g(x)}{f_0(x)}>1} \psi_2\left(\frac{g(x)}{f_0(x)}\right) f_0(x)\mathrm{d}x.$$

Since ψ_2 is bounded on $[0,1]$, we have that $I_{\psi_2}(g,f)=\infty$ if and only if $\int_{x:\frac{g(x)}{f_0(x)}>1} \psi_2\left(\frac{g(x)}{f_0(x)}\right) f_0(x)\mathrm{d}x = \infty$, which implies that $I_{\phi_2}(g,f)=\infty$. □

Proof of Lemma 1. Following the construction in (13), we choose $\psi(s) = (s+e)\exp(a\log^p(s+e))$, where $a = 1/(p\,(\sigma d)^p)$. Clearly, $\psi(s)$ is increasing and twice continuously differentiable on $s \geq 0$. Moreover, it is strictly convex, since its second derivative is given by

$$\psi''(s) = \frac{ap\log^{p-2}(s+e)\exp(a\log^p(s+e))(ap\log^p(s+e)+p+\log(s+e)-1)}{s+e},$$

which is strictly positive for all $s \geq 0$. Hence, ϕ_2^* as defined in (15) satisfies all properties given in Proposition 2.

We now show that the conjugate function ϕ_2^* satisfies the condition (10), which also implies (9). For any $\theta_1 \geq 0, \theta_2 \in \mathbb{R}$, we have that $\phi_1^*(\theta_2 - x) = \max\{(\theta_2 - x)/\alpha, 0\}$. Since $f_0(x)$ has support $(-\infty, 0]$, we have that $(\theta_2 - x)/\alpha \geq 0$, if $x \leq x_0$ for some x_0. Therefore, for all $x \leq x_0$, we have

$$\phi_2^*\left(\phi_1^*(\theta_2 - x) + \theta_1\right) f_0(x) = \phi_2^*\left(\frac{1}{\alpha}(\theta_2 - x) + \theta_1\right) \frac{1}{C\sigma|x|} \exp\left\{-\frac{1}{p\sigma^p}(\log(|x|) - \mu)^p\right\}.$$

We analyze the asymptotic behavior of the above function. We have that

$$\begin{aligned}
&\phi_2^*\left(\frac{1}{\alpha}(\theta_2 - x) + \theta_1\right) \frac{1}{C\sigma|x|} \exp\left\{-\frac{1}{p\sigma^p}(\log|x| - \mu)^p\right\} \\
&= \frac{c_1\left(\frac{1}{\alpha}(\theta_2 + |x|) + \theta_1 + e\right)}{C\sigma|x|} \cdot \exp\left(\frac{1}{p(\sigma d)^p}\log^p\left(\frac{1}{\alpha}(\theta_2 + |x|) + \theta_1 + e\right) - \frac{1}{p\sigma^p}(\log|x| - \mu)^p\right) \\
&\quad + \left(\frac{c_2\left(\frac{1}{\alpha}(\theta_2 + |x|) + \theta_1\right) + c_3}{C\sigma|x|}\right) \cdot \exp\left(-\frac{1}{p\sigma^p}(\log|x| - \mu)^p\right) \\
&= O(1) \cdot \exp\left(\frac{1}{p(\sigma d)^p}\log^p\left(\frac{1}{\alpha}(\theta_2 + |x|) + \theta_1 + e\right) - \frac{1}{p\sigma^p}(\log|x| - \mu)^p\right) + \\
&\quad O(1) \cdot \exp\left(-\frac{1}{p\sigma^p}(\log(|x|) - \mu)^p\right).
\end{aligned}$$

We only need to examine that,

$$
\begin{aligned}
&\exp\left(\frac{1}{p(\sigma d)^p}\log^p\left(\frac{1}{\alpha}(\theta_2+|x|)+\theta_1+e\right)-\frac{1}{p\sigma^p}(\log|x|-\mu)^p\right)\\
&=\exp\left(\frac{1}{p(\sigma d)^p}\log^p\left(|x|\left(\frac{1}{\alpha}+\frac{\frac{\theta_2}{\alpha}+\theta_1+e}{|x|}\right)\right)-\frac{1}{p\sigma^p}\log^p|x|(1+o(1))^p\right)\\
&=\exp\left(\frac{1}{p(\sigma d)^p}\log^p|x|\,(1+o(1))^p-\frac{1}{p\sigma^p}\log^p|x|(1+o(1))^p\right)\\
&=\exp\left(-\frac{1}{p\sigma^p}\log^p|x|\left(\frac{1}{d^p}(1+o(1))^p-(1+o(1))^p\right)\right).
\end{aligned}
$$

Therefore, the above expression has the same asymptotic as $\exp\left(-\frac{1}{p\sigma^p}\log^p|x|\left(\frac{1}{d^p}-1\right)\right)$. Since $d>1, p\geq 2 \Rightarrow 1/d^p-1<0$, we have that (10) is satisfied.

Finally, to show that the divergence ball of ϕ_2 contains distributions that have finite d-th moment. First, we examine the derivative of ϕ_2^*, which is also equal to $(\phi_2')^{-1}$:

$$
(\phi_2^*)'(s)=\begin{cases} c_1\exp\left\{\frac{1}{p(\sigma d)^p}\log^p(s+e)\right\}\left(\frac{1}{(\sigma d)^p}\log^{p-1}(s+e)+1\right)+c_2 & s>0\\ \exp(s) & s\leq 0.\end{cases}
$$

We define a function

$$
(\psi_2')^{-1}(s)=\begin{cases} c_1\exp\left\{\frac{1}{p(\sigma d)^p}\log^p(s+e)\right\}\left(\frac{1}{(\sigma d)^p}+1\right)+c_2 & s>0\\ \exp(s) & s\leq 0.\end{cases}
$$

We note that $(\phi_2^*)'(s)\geq(\psi_2')^{-1}(s)$, for all $s\in\mathbb{R}$. Moreover, ψ_2' is given by

$$
\psi_2'(t)=\begin{cases} \exp\left\{\left(p\,(\sigma d)^p\log\left(\frac{t-c_2}{c_1\left(\frac{1}{(\sigma d)^p}+1\right)}\right)\right)^{\frac{1}{p}}\right\}-e & t\geq 1\\ \log(t) & 0<t\leq 1.\end{cases}
$$

By Proposition 3, we only have to show that the limit (14) holds. We have that,

$$
\begin{aligned}
\psi_2'\left(\frac{1}{f(x)}\right)&=\psi_2'\left(C\sigma|x|\exp\left\{\frac{1}{p\sigma^p}(\log|x|-\mu)^p\right\}\right)\\
&=\exp\left\{\left(p(\sigma d)^p\cdot\log\left(|x|\exp\left\{\frac{1}{p\sigma^p}(\log|x|-\mu)^p\right\}\cdot O(1)\right)\right)^{\frac{1}{p}}\right\}\\
&=\exp\left\{\left(p(\sigma d)^p\cdot\left(\frac{1}{p\sigma^p}(\log|x|-\mu)^p+\log|x|+O(1)\right)\right)^{\frac{1}{p}}\right\}\\
&=\exp\left(\left(d^p\,(\log|x|-\mu)^p\cdot O(1)\right)^{\frac{1}{p}}\right)\\
&=O(|x|^d).
\end{aligned}
$$

Hence, ψ_2' satisfies the limit (14)

Similarly, we can also construct a function that bounds $(\phi_2^*)'$ from above. For any $\gamma>d$, we define

$$(\hat{\psi}_2^*)'(s) = \begin{cases} \max\left\{\exp\left\{\frac{1}{p(\gamma\sigma)^p}(\log^p(s+e)-1)\right\}, (\phi_2^*)'(s)\right\} & s > 0 \\ \exp(s) & s \le 0. \end{cases}$$

Then, we have that $(\hat{\psi}_2^*)'(s) \ge (\phi_2^*)'(s)$ for all $s \in \mathbb{R}$. Moreover, $\hat{\psi}_2^*(s) := \int_0^s (\hat{\psi}_2^*)'(y)\mathrm{d}y$ is an increasing convex function with $\hat{\psi}_2^*(0) = 0, (\hat{\psi}_2^*)'(0) = 1$. Hence, $\hat{\psi}_2 := \hat{\psi}_2^{**}$ is also a divergence function, and $(\hat{\psi}_2^*)' = (\hat{\psi}_2')^{-1}$ by Proposition 2.

To show that the divergence $\hat{\psi}_2$ excludes all densities g that satisfy $\liminf_{x\to-\infty} |x|^{t+1} g(x) > 0$, for any $t \in (1, d)$. We let $\gamma \in (t, d)$ and use Proposition 3 of Kruse et al. [2019], for which we have to verify the limit conditions

$$\limsup_{u\to\infty} \frac{u\hat{\psi}_2''(u)}{\hat{\psi}_2'(u)} < \infty,$$

and

$$\liminf_{x\to-\infty} \frac{\hat{\psi}_2'\left(\frac{1}{f_0(x)} \cdot \frac{1}{|x|^{t+1}} \cdot c\right)}{|x|^t} > 0, \tag{EC.41}$$

for all constants $c > 0$. We note that since $\frac{1}{p(\gamma\sigma)^p} > \frac{1}{p(d\sigma)^p}$ due to $\gamma < d$, we have that $(\hat{\psi}_2')^{-1}(s) = (\hat{\psi}_2^*)'(s) = \exp\left\{\frac{1}{p(\gamma\sigma)^p}(\log^p(s+e)-1)\right\}$, if s is sufficiently large. This means that for u sufficiently large, we have $\hat{\psi}_2'(u) = \exp\left\{(p(\gamma\sigma)^p \log(u)+1)^{\frac{1}{p}}\right\} - e$. Hence, we have that

$$\begin{aligned}\limsup_{u\to\infty} \frac{u\hat{\psi}_2''(u)}{\hat{\psi}_2'(u)} &= \limsup_{u\to\infty} \frac{(\hat{\psi}_2'(u)+e)(\gamma\sigma)^p(p(\gamma\sigma)^p\log(u)+1)^{\frac{1}{p}-1}}{\hat{\psi}_2'(u)} \\ &= \lim_{u\to\infty} (\gamma\sigma)^p(p(\gamma\sigma)^p\log(u)+1)^{\frac{1}{p}-1} = 0 < \infty,\end{aligned}$$

since $1/p - 1 < 0$ due to $p \ge 2$. We also have that, for any constant $c > 0$:

$$\begin{aligned}\hat{\psi}_2'\left(\frac{1}{f_0(x)} \cdot \frac{1}{|x|^{t+1}} \cdot c\right) &= \exp\left((\gamma^p(\log|x|-\mu)^p + O(1)\cdot\log|x| + O(1))^{\frac{1}{p}}\right) \\ &= |x|^\gamma \cdot \exp(O(1)).\end{aligned}$$

Therefore, we have that $\lim_{x\to-\infty} |x|^{-t}\hat{\psi}_2'\left(\frac{1}{f_0(x)} \cdot \frac{1}{|x|^{t+1}} \cdot c\right) = \lim_{x\to-\infty} |x|^{\gamma-t}\exp(O(1)) > 0$, since $\gamma > t$. Therefore, we have that $I_{\hat{\psi}_2}(g, f_0) = \infty$, which implies that $I_{\phi_2}(g, f_0) = \infty$, by Proposition 3. $\square$

Proof of Lemma 2. Following the construction in (13), we choose $\psi(s) = (s+1)\exp((s+1)^p)$, where $p = \frac{k}{d}$. The second derivative of ψ is given by:

$$\psi''(s) = p(s+1)^{p-1}\exp((s+1)^p)(p + p(s+1)^p + 1),$$

which is positive for all $s \ge 0$. Hence, ψ is an increasing, strictly convex and twice continuously differentiable function on $s \ge 0$. By Proposition 2, ϕ_2^* defined in (16) induces a divergence function. We examine condition (10). Choose an arbitrary $\theta_2 \in \mathbb{R}, \theta_1 \geq 0$. For $x \to -\infty$, we have that

$$\begin{aligned}
\phi_2^*(\phi_1^*(\theta_2 - x) + \theta_1) f_0(x) &= \phi_2^*\left(\frac{1}{\alpha}(\theta_2 - x) + \theta_1\right) \frac{k}{\lambda}\left(\frac{|x|}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{|x|}{\lambda}\right)^k\right) \\
&= O(|x|^k) \cdot \exp\left(\frac{1}{\alpha}|x|^{\frac{k}{d}} \cdot O(1) - \left(\frac{|x|}{\lambda}^k\right)\right) + O\left(|x|^k \exp\left(-\left(\frac{|x|}{\lambda}\right)^k\right)\right).
\end{aligned}$$

Since $\frac{k}{d} < k$, we have that the term $\exp\left(\frac{1}{\alpha}|x|^{\frac{k}{d}} \cdot O(1) - \left(\frac{|x|}{\lambda}^k\right)\right)$ is integrable. Hence, condition (10) is satisfied.

To examine the content of the divergence ball, we first examine the derivative of ϕ_2^*, which is equal to

$$(\phi_2^*)'(s) = \begin{cases} c_1 \exp((s+1)^{\frac{k}{d}}) \left(1 + \frac{k}{d}(s+1)^{\frac{k}{d}}\right) + c_2 & s > 0 \\ \exp(s) & s \leq 0. \end{cases}$$

We then define the auxiliary function

$$(\psi_2')^{-1}(s) = \begin{cases} c_1 \exp((s+1)^{\frac{k}{d}}) \left(1 + \frac{k}{d}\right) + c_2 & s > 0 \\ \exp(s) & s \leq 0, \end{cases}$$

which has inverse

$$\psi_2'(t) = \begin{cases} \log^{\frac{d}{k}}\left(\frac{t - c_2}{c_1(1 + k/d)}\right) - 1 & t > 1 \\ \log(t) & 0 < t \leq 1. \end{cases}$$

We examine the asymptotic behavior:

$$\begin{aligned}
\psi_2'\left(\frac{1}{f_0(x)}\right) &= \log^{\frac{d}{k}}\left(|x|^{1-k} \exp\left(\left(\frac{|x|}{\lambda}\right)^k\right) \cdot O(1)\right) - 1 \\
&= \left(\left(\frac{|x|}{\lambda}\right)^k + (1-k)\log|x| + O(1)\right)^{\frac{d}{k}} - 1 \\
&= O(|x|^d).
\end{aligned}$$

Hence, we have $\limsup_{x \to -\infty} |x|^{-d} \psi_2'\left(\frac{1}{f_0(x)}\right) < \infty$. Proposition 3 implies that $I_{\phi_2}(g, f_0) < \infty$ for any continuous density g with finite d-th moment.

On the other hand, for any $\gamma > 1$, we can define the function

$$(\hat{\psi}_2^*)'(s) = \begin{cases} \max\left\{\exp\left(\gamma(s+1)^{\frac{k}{d}} - \gamma\right), (\phi_2^*)'(s)\right\} & s > 0 \\ \exp(s) & s \leq 0. \end{cases}$$

Then, we have that $(\hat{\psi}_2^*)'(s) \geq (\phi_2^*)'(s)$ for all $s \in \mathbb{R}$, and that $\hat{\psi}_2^*(s) := \int_0^s (\hat{\psi}_2^*)'(y) \mathrm{d}y$ is an increasing convex function with $\hat{\psi}_2^*(0) = 0, (\hat{\psi}_2^*)'(0) = 1$. Hence, $\hat{\psi}_2 := \hat{\psi}_2^{**}$ is a divergence function by Proposition 2. Since the term $\exp\left(\gamma(s+1)^{\frac{k}{d}}\right)$ dominates the term $\exp\left((s+1)^{\frac{k}{d}}\right)$ in $(\phi_2^*)'$ as $s \to \infty$, we have that for s sufficiently large, that $(\hat{\psi}_2^*)'(s) = \exp\left(\gamma(s+1)^{\frac{k}{d}} - \gamma\right)$. Hence, for u sufficiently large, we have that

$$\hat{\psi}_2'(u) = \left(\frac{\log(u)+\gamma}{\gamma}\right)^{\frac{d}{k}} - 1.$$

Therefore, to apply Proposition 3 of Kruse et al. [2019], we examine the limit conditions

$$\limsup_{u\to\infty} \frac{u(\hat{\psi}_2)''(u)}{(\hat{\psi}_2)'(u)} = \limsup_{u\to\infty} \frac{d}{\gamma k}\left(\frac{\log(u)+\gamma}{\gamma}\right)^{-1} < \infty,$$

and for any $t \in (1,d)$, any constants $c > 0$, we have

$$\begin{aligned}\liminf_{x\to-\infty} \hat{\psi}_2'\left(\frac{1}{f_0(x)}\cdot\frac{1}{|x|^{t+1}}\cdot c\right)|x|^{-t} &= \liminf_{x\to-\infty}\left(\left(\left(\frac{|x|}{\lambda}\right)^k + O(\log|x|)\right)\cdot O(1)\right)^{\frac{d}{k}}|x|^{-t}\\ &= O(|x|^{d-t}) > 0.\end{aligned}$$

Hence, by Proposition (3), we have that $I_{\phi_2}(g,f_0)=\infty$, for any g such that $\liminf_{x\to-\infty}|x|^{t+1}g(x) > 0$. □

Proof of Lemma 3. It follows from the proof of Lemma 1 that ϕ_2^* as in (17) is convex and twice differentiable and indeed defines a ϕ-divergence. It remains to show that ϕ_2^* satisfies condition (10) and (9). To verify this, we choose $\theta_1 = \theta_2 = 0$. Then, we examine

$$\begin{aligned}&\phi_2^*\left(\gamma(\exp\{\gamma|x|\}-1)\right)\frac{k}{\lambda}\left(\frac{|x|}{\lambda}\right)^{k-1}\exp\left\{-\left(\frac{|x|}{\lambda}\right)^k\right\}\\ &= c_1(\gamma(\exp\{\gamma|x|\}-1)+e)\exp\left\{\frac{1}{(2\gamma\lambda)^k}\log^k\left(\gamma(\exp\{\gamma|x|\}-1)\right)\right\}\\ &\quad\cdot\frac{k}{\lambda}\left(\frac{|x|}{\lambda}\right)^{k-1}\exp\left\{-\left(\frac{|x|}{\lambda}\right)^k\right\} + O\left(\exp\left(\gamma|x| - \left(\frac{|x|}{\lambda}\right)^k\right)\right).\end{aligned}$$

Since $k>1$, $O\left(\exp\left(\gamma|x| - \left(\frac{|x|}{\lambda}\right)^k\right)\right)$ is integrable. Therefore, we only examine the dominant terms that determine the asymptotics:

$$\begin{aligned}&\exp\left\{\gamma|x| + \frac{1}{(2\gamma\lambda)^k}\log^k\left(\gamma(\exp\{\gamma|x|\}-1)\right) - \left(\frac{|x|}{\lambda}\right)^k\right\}\\ &= \exp\left\{\gamma|x| + \left(\frac{|x|}{2\lambda}\right)^k(1+o(1)) - \left(\frac{|x|}{\lambda}\right)^k\right\}\\ &= O\left(\exp\left\{\gamma|x| + \left(\frac{1}{(2\lambda)^k} - \frac{1}{\lambda^k}\right)|x|^k\right\}\right).\end{aligned}$$

Hence, this term is integrable and therefore (10) is satisfied.

We have that the derivative of ϕ_2^* is given by

$$(\phi_2^*)'(s) = \begin{cases} c_1 \exp\left\{\frac{1}{(2\gamma\lambda)^k} \log^k(s+e)\right\} \left(\frac{k}{(2\gamma\lambda)^k} \log^{k-1}(s+e) + 1\right) + c_2 & s > 0 \\ \exp(s) & s \leq 0. \end{cases}$$

We define a function

$$(\psi_2')^{-1}(s) = \begin{cases} c_1 \exp\left\{\frac{1}{(2\gamma\lambda)^k} \log^k(s+e)\right\} \left(\frac{k}{(2\gamma\lambda)^k} + 1\right) + c_2 & s > 0 \\ \exp(s) & s \leq 0. \end{cases}$$

We note that $(\phi_2^*)'(s) \geq (\psi_2')^{-1}(s)$, for all $s \in \mathbb{R}$. Moreover, ψ_2' is given by

$$\psi_2'(t) = \begin{cases} \exp\left\{\left((2\gamma\lambda)^k \log\left(\frac{t-c_2}{c_1\left(\frac{k}{(2\gamma\lambda)^k}+1\right)}\right)\right)^{\frac{1}{k}}\right\} - e & t \geq 1 \\ \log(t) & 0 < t \leq 1. \end{cases}$$

We consider a Weibull distribution g with slightly heavier tail: $g(x) = \frac{l}{\lambda}\left(\frac{|x|}{\lambda}\right)^{l-1} \exp\left(-\left(\frac{|x|}{\lambda}\right)^l\right)$, where $1 < l < k$. Then, we show that $I_{\phi_2}(g, f_0) < \infty$. Following the proof of Proposition 3, it is sufficient to show that the following function is integrable on $\{x : |x| > x_0\}$ for some $x_0 \in \mathbb{R}$:

$$\begin{aligned} &g(x)\psi_2'\left(\frac{1}{f(x)}\right) \\ &= O\left(|x|^{l-1} \exp\left\{-\left(\frac{|x|}{\lambda}\right)^l + (2\gamma\lambda)\left(\left(\frac{|x|}{\lambda}\right)^k + (1-k)\log\left(\frac{|x|}{\lambda}\right) + O(1)\right)^{\frac{1}{k}}\right\}\right) \\ &= O\left(|x|^{l-1} \exp\left\{-\left(\frac{|x|}{\lambda}\right)^l + 2\gamma|x|\right\}\right). \end{aligned}$$

Since $l > 1$, we have that $g(x)\psi_2'\left(\frac{1}{f_0(x)}\right)$ is indeed integrable. Therefore, $I_{\phi_2}(g, f_0) < \infty$. $\square$

Proof of Theorem 2. We express the following optimization problem in terms of density functions:

$$\begin{aligned} &\sup_{\lambda>0} \sup_{\{\mathbb{Q}: I_{\phi_2}(\mathbb{Q},\mathbb{P}_0)\leq r\}} \sup_{\{\bar{\mathbb{Q}}:\bar{\mathbb{Q}}\ll\mathbb{Q}\}} \mathbb{E}_{\bar{\mathbb{Q}}}[-X] - I_{\phi_1}(\bar{\mathbb{Q}}, \lambda\mathbb{Q}). \\ &= - \inf_{\substack{\lambda>0, g, \bar{g}\geq 0, \\ \int_\Omega g(\omega)\mathrm{d}\mathbb{P}_0(\omega)=1, \int_\Omega \bar{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)=1 \\ \int_\Omega \phi_2(g(\omega))\mathrm{d}\mathbb{P}_0(\omega)\leq r}} \int_\Omega X(\omega)\bar{g}(\omega) + \lambda g(\omega)\phi_1\left(\frac{\bar{g}(\omega)}{\lambda g(\omega)}\right) \mathrm{d}\mathbb{P}_0(\omega). \end{aligned}$$

Due to the product $\lambda g(\omega)$, this is a non-convex problem. However, using the substitution

$\tilde{g}(\omega) = \lambda g(\omega)$, this is equivalent to the following convex problem

$$\mu_0 := \inf_{\substack{\lambda>0,\tilde{g},\bar{g}\geq 0, \\ \int_\Omega \tilde{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)=\lambda, \int_\Omega \bar{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)=1 \\ \int_\Omega \lambda\phi_2\left(\frac{\tilde{g}(\omega)}{\lambda}\right)\mathrm{d}\mathbb{P}_0(\omega)\leq \lambda r}} \int_\Omega X(\omega)\bar{g}(\omega) + \tilde{g}(\omega)\phi_1\left(\frac{\bar{g}(\omega)}{\tilde{g}(\omega)}\right)\mathrm{d}\mathbb{P}_0(\omega).$$

Assume without loss of generality that μ_0 is finite. To apply Theorem EC.2.1, we choose $\mathcal{X} = (0,\infty)\times\mathcal{L}^1(\Omega)\times\mathcal{L}^1(\Omega)$, $\mathcal{C} = (0,\infty)\times\mathcal{L}^1_+(\Omega)\times\mathcal{L}^1_+(\Omega)$, $H(\lambda,\tilde{g},\bar{g}) = \left(\int_\Omega \tilde{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega), \int_\Omega \bar{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)\right) - (\lambda, 1)$, $G(\lambda,\tilde{g},\bar{g}) = \int_\Omega \lambda\phi_2\left(\frac{\tilde{g}(\omega)}{\lambda}\right)\mathrm{d}\mathbb{P}_0(\omega) - \lambda r$. Then, we have that for $f_1 \equiv 1$ the constant function, $(1, f_1, f_1)$ is a point such that $H(1, f_1, f_1) = (0,0)$ and $G(1, f_1, f_1) < 0$. Furthermore, $(0,0)$ is an interior point of the image set of H. Hence, strong duality implies that

$$\mu_0 = \sup_{\eta\geq 0} -\eta r\lambda + \inf_{\substack{\lambda>0,\tilde{g},\bar{g}\geq 0, \\ \int_\Omega \tilde{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)=\lambda \\ \int_\Omega \bar{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)=1}} \int_\Omega X(\omega)\bar{g}(\omega) + \tilde{g}(\omega)\phi_1\left(\frac{\bar{g}(\omega)}{\tilde{g}(\omega)}\right) + \eta\lambda\phi_2\left(\frac{\tilde{g}(\omega)}{\lambda}\right)\mathrm{d}\mathbb{P}_0(\omega).$$

We first study the case when the supremum is attained at $\eta = 0$, in which case we have that

$$\begin{aligned}
\mu_0 &= \inf_{\substack{\lambda>0,\tilde{g},\bar{g}\geq 0, \\ \int_\Omega \tilde{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)=\lambda \\ \int_\Omega \bar{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)=1}} \int_\Omega X(\omega)\bar{g}(\omega) + \tilde{g}(\omega)\phi_1\left(\frac{\bar{g}(\omega)}{\tilde{g}(\omega)}\right)\mathrm{d}\mathbb{P}_0(\omega) \\
&\geq \inf_{\substack{\lambda>0,\tilde{g},\bar{g}\geq 0, \\ \int_\Omega \tilde{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)=\lambda \\ \int_\Omega \bar{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)=1}} \int_\Omega X(\omega)\bar{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega) = \operatorname{ess\,inf}(X).
\end{aligned}$$

On the other hand, taking $\lambda = 1, \tilde{g} = \bar{g}$ shows that $\operatorname{ess\,inf}(X)$ is also an upper bound of μ_0. Hence, $\mu_0 = \operatorname{ess\,inf}(X)$. The dual formula (20) states that when $\eta = 0$, $\operatorname{ess\,inf}(X)$ is equal to

$$\sup_{\theta_1,\theta_2\in\mathbb{R}} \left\{ -\theta_2 \;\middle|\; \mathbb{E}_{\mathbb{P}_0}\left[0\phi_2^*\left(\frac{\phi_1^*(-\theta_2 - X) + \theta_1}{0}\right)\right] \leq \theta_1 \right\},$$

which implies that $0 \leq \theta_1 \leq -\phi_1^*(-\theta_2 - \operatorname{ess\,inf}(X))$. This inequality is satisfied if we take $-\theta_2 = \operatorname{ess\,inf}(X)$ and $\theta_1 = 0$, but violated if $-\theta_2 > \operatorname{ess\,inf}(X)$. Hence, the duality formula (20) holds for $\eta = 0$, and we only have to examine the case where the supremum is attained at $\eta > 0$, where strong duality implies

$$\mu_0 = \sup_{\substack{\eta>0 \\ \theta_1,\theta_2\in\mathbb{R}}} \inf_{\substack{\lambda>0 \\ \tilde{g},\bar{g}\geq 0}} (-\eta r + \theta_1)\lambda + \theta_2 + \int_\Omega \beta_{\eta,\lambda}(\omega, \tilde{g}(\omega), \bar{g}(\omega)) - \theta_1\tilde{g}(\omega) - \theta_2\bar{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega),$$

where

$$\beta_{\eta,\lambda}(\omega, s, t) = X(\omega)t + s\phi_1\left(\frac{t}{s}\right) + \lambda\eta\phi_2\left(\frac{s}{\lambda}\right).$$

Note that when $\eta > 0$, $\beta_{\eta,\lambda}(\omega, s, t)$ is only finite when $s, t \geq 0$. Analogous to the proof of and

Theorem EC.2.2, we have

$$\begin{aligned}
&\inf_{\tilde{g},\bar{g}\geq 0}\int_{\Omega}\beta_{\eta,\lambda}(\omega,\tilde{g}(\omega),\bar{g}(\omega))-\theta_1\tilde{g}(\omega)-\theta_2\bar{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)\\
&=-\int_{\Omega}\beta^*_{\eta,\lambda}(\omega,\theta_1,\theta_2)\mathrm{d}\mathbb{P}_0(\omega),
\end{aligned}$$

where $\int_{\Omega}\beta^*_{\eta,\lambda}(\omega,\theta_1,\theta_2)\mathrm{d}\mathbb{P}_0(\omega)=+\infty$ if $\eta=0$ and for $\eta>0$, we have

$$\int_{\Omega}\beta^*_{\eta,\lambda}(\omega,\theta_1,\theta_2)\mathrm{d}\mathbb{P}_0(\omega)=\int_{\Omega}\lambda\eta\phi_2^*\left(\frac{\theta_1+\phi_1^*(\theta_2-X(\omega))}{\eta}\right)\mathrm{d}\mathbb{P}_0(\omega).$$

Hence, we have

$$\begin{aligned}
\mu_0&=\sup_{\substack{\eta>0\\ \theta_1,\theta_2\in\mathbb{R}}}\ \theta_2+\inf_{\lambda>0}\left(\theta_1-\eta r-\int_{\Omega}\eta\phi_2^*\left(\frac{\theta_1+\phi_1^*(\theta_2-X(\omega))}{\eta}\right)\mathrm{d}\mathbb{P}_0(\omega)\right)\lambda\\
&=\sup_{\substack{\eta>0\\ \theta_1,\theta_2\in\mathbb{R}}}\ \left\{\theta_2\ \middle|\ \theta_1-\eta r-\int_{\Omega}\eta\phi_2^*\left(\frac{\theta_1+\phi_1^*(\theta_2-X(\omega))}{\eta}\right)\mathrm{d}\mathbb{P}_0(\omega)\geq 0\right\}.
\end{aligned}$$

□

Proof of Corollary 2. Clearly, (19) can only hold if there actually exists $\theta_1,\theta_2\in\mathbb{R}$, $\eta\geq 0$, such that (21) is finite. Hence, suppose there exists some $\tilde{\theta}_1,\tilde{\theta}_2\in\mathbb{R}$, $\tilde{\eta}\geq 0$ such that (21) is finite. If there also exists a $\hat{\theta}_2\in\mathbb{R}$, such that

$$\int_{\mathbb{R}}\tilde{\eta}\phi_2^*\left(\frac{\phi_1^*(-\hat{\theta}_2-x)+\tilde{\theta}_1}{\tilde{\eta}}\right)f_0(x)\mathrm{d}x\leq\tilde{\theta}_1-\tilde{\eta}r,\tag{EC.42}$$

then, we must have

$$\inf_{\theta_2\in\mathbb{R}}\int_{\mathbb{R}}\tilde{\eta}\phi_2^*\left(\frac{\phi_1^*(-\theta_2-x)+\tilde{\theta}_1}{\tilde{\eta}}\right)f_0(x)\mathrm{d}x\leq\tilde{\theta}_1-\tilde{\eta}r.$$

since both ϕ_2^* and ϕ_1^* are non-decreasing, monotonce convergence theorem implies that

$$\begin{aligned}
&\inf_{\theta_2\in\mathbb{R}}\int_{\mathbb{R}}\tilde{\eta}\phi_2^*\left(\frac{\phi_1^*(-\theta_2-x)+\tilde{\theta}_1}{\tilde{\eta}}\right)f_0(x)\mathrm{d}x\\
&=\int_{\mathbb{R}}\lim_{\theta_2\to-\infty}\tilde{\eta}\phi_2^*\left(\frac{\phi_1^*(-\theta_2-x)+\tilde{\theta}_1}{\tilde{\eta}}\right)f_0(x)\mathrm{d}x\\
&=\int_{\mathbb{R}}\tilde{\eta}\phi_2^*\left(\frac{\phi_1^*(-\infty)+\tilde{\theta}_1}{\tilde{\eta}}\right)f_0(x)\mathrm{d}x=\tilde{\eta}\phi_2^*\left(\frac{\phi_1^*(-\infty)+\tilde{\theta}_1}{\tilde{\eta}}\right).
\end{aligned}$$

Therefore, if $\hat{\theta}_2$ satisfies (EC.42), then one cannot have

$$\tilde{\eta}\phi_2^*\left(\frac{\phi_1^*(-\infty)+\tilde{\theta}_1}{\tilde{\eta}}\right)>\tilde{\theta}_1-\tilde{\eta}r.$$

Hence, if (22) holds for all θ_1, θ_2, η, then (19) cannot hold.

On the other hand, if there exits some $\tilde{\theta}_1, \tilde{\theta}_2 \in \mathbb{R}$, $\tilde{\eta} \geq 0$ such that (21) is finite and

$$\tilde{\eta}\phi_2^*\left(\frac{\phi_1^*(-\infty)+\tilde{\theta}_1}{\tilde{\eta}}\right) < \tilde{\theta}_1 - \tilde{\eta}r,$$

then it follows from the monotone convergence theorem that there exists a $\hat{\theta}_2 \in \mathbb{R}$ such that (EC.42) is satisfied, and thus (19). $\square$

Proof of Proposition 4. Let $x \in \mathbb{R}$, and denote $S(x,p) = \rho^l_{\text{oce},\mathbb{P}}(X_p)$. We note that

$$\lim_{p\downarrow 0} \rho^l_{\text{oce},\mathbb{P}}(X_p)/p = \frac{\partial S(x,p)}{\partial p}|_{p=0}.$$

We have that,

$$S(x,p) = \inf_{\theta_1,\theta_2\in\mathbb{R}} -\theta_1 - \theta_2 + p(\phi_2^*(\phi_1^*(\theta_2 - x) + \theta_1) + (1-p)\phi_2^*(\phi_1^*(\theta_2) + \theta_1),$$

and we denote

$$F(\theta_1,\theta_2,p) := -\theta_1 - \theta_2 + p(\phi_2^*(\phi_1^*(\theta_2 - x) + \theta_1) + (1-p)\phi_2^*(\phi_1^*(\theta_2) + \theta_1).$$

We note that for each p, the optimal value $\theta_1(x,p), \theta_2(x,p)$ satisfy the first-order conditions

$$\begin{aligned} p(\phi_2^*)'(\phi_1^*(\theta_2 - x) + \theta_1) + (1-p)(\phi_2^*)'(\phi_1^*(\theta_2) + \theta_1) = 1 \\ p(\phi_2^*)'(\phi_1^*(\theta_2 - x) + \theta_1)(\phi_1^*)'(\theta_2 - x) + (1-p)(\phi_2^*)'(\phi_1^*(\theta_2) + \theta_1)(\phi_1^*)'(\theta_2) = 1. \end{aligned}$$

Since $\phi_i^*(0) = 0, (\phi_i^*)'(0) = 1$ for $i = 1,2$, we have that $(0,0)$ satisfies the above condition for $p = 0$. Hence, $\theta_1(x,0) = \theta_2(x,0) = 0$. Now, by the Envelope theorem (see e.g., Mas-Colell et al., 1995), we have that

$$\begin{aligned} \frac{\partial S(x,p)}{\partial p} &= \frac{\partial F(\theta_1,\theta_2,p)}{\partial p}\bigg|_{(\theta_1(x,p),\theta_2(x,p),p)} \\ &= \phi_2^*(\phi_1^*(\theta_2(x,p) - x) + \theta_1(x,p)) - \phi_2^*(\phi_1^*(\theta_2(x,p)) + \theta_1(x,p)). \end{aligned}$$

Hence,

$$\begin{aligned} \frac{\partial S(x,p)}{\partial p}\bigg|_{p=0} &= \phi_2^*(\phi_1^*(\theta_2(x,0) - x) + \theta_1(x,0)) - \phi_2^*(\phi_1^*(\theta_2(x,0)) + \theta_1(x,0)) \\ &= \phi_2^*(\phi_1^*(-x)). \end{aligned}$$

To show (29), we first note that by weak duality, we have $\rho^l_{\text{oce},\mathbb{P}}(X_p) \leq \rho^l_{\text{sf},\mathbb{P}}(X_p)$ for all $p > 0$. Therefore, we already have

$$\liminf_{p\downarrow 0} \frac{\rho^l_{\text{sf},\mathbb{P}}(X_p)}{p} \geq \phi_2^*(\phi_1^*(-x)).$$

We will now show the reverse. By definition, we have that

$$\begin{aligned}\frac{1}{p}\rho^l_{\mathrm{sf},\mathbb{P}}(X_p) &= \inf_{\theta_1,\theta_2\in\mathbb{R}}\left\{\frac{1}{p}\theta_2 \;\middle|\; p\phi_2^*(\phi_1^*(-\theta_2-x)+\theta_1)+(1-p)\phi_2^*(\phi_1^*(-\theta_2)+\theta_1)\le\theta_1\right\}\\ &= \inf_{\theta_1,\theta_2\in\mathbb{R}}\left\{\theta_2 \mid p\phi_2^*(\phi_1^*(-p\theta_2-x)+\theta_1)+(1-p)\phi_2^*(\phi_1^*(-p\theta_2)+\theta_1)\le\theta_1\right\}.\end{aligned}$$

We bound the above infimum by consider only cases when $\theta_1=0$. Then, we show that as $p\downarrow 0$, the following inequality holds for any $\theta_2>\phi_2^*(\phi_1^*(-x))$:

$$p\phi_2^*(\phi_1^*(-p\theta_2-x))+(1-p)\phi_2^*(\phi_1^*(-p\theta_2))\le 0.$$

The above inequality holds if and only if

$$\phi_2^*(\phi_1^*(-p\theta_2-x))-\phi_2^*(\phi_1^*(-p\theta_2))\le\theta_2\cdot\frac{\phi_2^*(\phi_1^*(-p\theta_2))}{-p\theta_2}.$$

As $p\downarrow 0$, the left-hand side converges to $\phi_2^*(\phi_1^*(-x))$, and the right-hand side converges to $\theta_2\cdot\lim_{t\to 0}\frac{\phi_2^*(\phi_1^*(t))}{t}=\theta_2\cdot 1=\theta_2$. Since $\theta_2>\phi_2^*(\phi_1^*(-x))$, the above inequality is feasible for p sufficiently small. Therefore, we also have $\limsup_{p\downarrow 0}\frac{\rho^l_{\mathrm{sf},\mathbb{P}}(X_p)}{p}\le\phi_2^*(\phi_1^*(-x))$. □

Proof of Theorem 4. We have that (EC.43) is equivalent (up to a sign convention) to the following optimization problem over density functions:

$$\begin{aligned}&\text{Minimize } \int_\Omega X(\omega)\bar{g}(\omega)+\tilde{g}(\omega)\phi_1\left(\frac{\bar{g}(\omega)}{\tilde{g}(\omega)}\right)+g(\omega)\phi_2\left(\frac{\tilde{g}(\omega)}{g(\omega)}\right)\mathrm{d}\mathbb{P}_0(\omega)\\ &\text{subject to } \int_\Omega g(\omega)\mathrm{d}\mathbb{P}_0(\omega)=\int_\Omega\tilde{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)=\int_\Omega\bar{g}(\omega)\mathrm{d}\mathbb{P}_0(\omega)=1\\ &\qquad\int_\Omega\phi_3(g(\omega))\mathrm{d}\mathbb{P}_0(\omega)\le r\\ &\qquad g(\omega),\tilde{g}(\omega),\bar{g}(\omega)\ge 0.\end{aligned}\tag{EC.43}$$

We denote the optimal value of the primal problem (EC.43) as J_0. Note that $J_0<\infty$ since the constant unity function $g,\tilde{g},\bar{g}\equiv 1$ is a feasible solution and we assumed $\int_\Omega X(\omega)\mathrm{d}\mathbb{P}_0(\omega)<\infty$. Strong duality then implies that

$$\begin{aligned}J_0=\sup_{\substack{\theta_1,\theta_2,\theta_3\in\mathbb{R}\\ \lambda\ge 0}} -\lambda r+\inf_{\substack{g,\tilde{g},\bar{g}\ge 0\\ \int g=\int\tilde{g}=\int\bar{g}=1}}\int_\Omega X(\omega)\bar{g}(\omega)+\tilde{g}(\omega)\phi_1\left(\frac{\bar{g}(\omega)}{\tilde{g}(\omega)}\right)+g(\omega)\phi_2\left(\frac{\tilde{g}(\omega)}{g(\omega)}\right)\mathrm{d}\\ +\lambda\phi_3(g(\omega))\mathbb{P}_0(\omega).\end{aligned}$$

Suppose this supremum is attained at $\lambda=0$, then we would have $J_0=\operatorname{ess\,inf}(X)$. The dual formula in (34) would also yield

$$\begin{aligned}&\sup_{\theta_1,\theta_2,\theta_3\in\mathbb{R}}\theta_1+\theta_2+\theta_3-\mathbb{E}_{\mathbb{P}_0}\left[0\phi_3^*\left(\frac{\phi_2^*(\phi_1^*(\theta_3-X(\omega))+\theta_2)+\theta_1}{0}\right)\right]\\ &=\sup_{\theta_2,\theta_3\in\mathbb{R}}\theta_2+\theta_3-\phi_2^*(\phi_1^*(\theta_2+\theta_3-\operatorname{ess\,inf}(X)))\\ &=\operatorname{ess\,inf}(X).\end{aligned}$$

Hence, the duality formula holds if the supremum is attained at $\lambda = 0$. We may now assume that the supremum is at $\lambda > 0$. Applying Theorem EC.2.1 and Theorem 14.60 of Rockafellar and Wets [2004], gives the duality

$$J_0 = \sup_{\substack{\theta_1,\theta_2,\theta_3\in\mathbb{R}\\ \lambda>0}} -\lambda r + \theta_1 + \theta_2 + \theta_3 - \int_\Omega \beta_\lambda^*(\omega,\theta_1,\theta_2,\theta_3)\mathrm{d}\mathbb{P}_0(\omega),$$

where

$$\beta_\lambda(\omega,t_1,t_2,t_3) = X(\omega)t_3 + t_2\phi_1\left(\frac{t_3}{t_2}\right) + t_1\phi_2\left(\frac{t_2}{t_1}\right) + \lambda\phi_3(t_1).$$

We note that $\beta_\lambda(\omega,t_1,t_2,t_3)$ is only finite if $t_1,t_2,t_3 \geq 0$, since $\lambda > 0$. Therefore, we have

$$\begin{aligned}
\beta_0^*(\omega,\theta_1,\theta_2,\theta_3) &= \sup_{t_1,t_2,t_3\geq 0} \theta_1 t_1 + \theta_2 t_2 + \theta_3 t_3 - X(\omega)t_3 - t_2\phi_1\left(\frac{t_3}{t_2}\right) - t_1\phi_2\left(\frac{t_2}{t_1}\right) - \lambda\phi_3(t_1)\\
&= \sup_{t_1\geq 0,t_2\geq 0} \theta_1 t_1 + \theta_2 t_2 + t_2\phi_1^*(\theta_3 - X(\omega)) - t_1\phi_2\left(\frac{t_2}{t_1}\right) - \lambda\phi_3(t_1)\\
&= \sup_{t_1\geq 0} \theta_1 t_1 + t_1\phi_2^*(\phi_1^*(\theta_3 - X(\omega)) + \theta_2) - \lambda\phi_3(t_1)\\
&= \lambda\phi_3^*\left(\frac{\phi_2^*(\phi_1^*(\theta_3 - X(\omega)) + \theta_2) + \theta_1}{\lambda}\right).
\end{aligned}$$

□

Proof of Proposition 5. Denote

$$K(\theta_1,\theta_2) := -\theta_1 - \theta_2 + \mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*(\phi_1^*(\theta_2 - X) + \theta_1)\right]. \tag{EC.44}$$

Since $\phi_2(1) = 0$, we have that $\phi_2^*(s) \geq s, \forall s \in \mathbb{R}$. Therefore, for any $\theta_1,\theta_2 \in \mathbb{R}$:

$$K(\theta_1,\theta_2) \geq -\theta_2 + \mathbb{E}_{\mathbb{P}_0}\left[\phi_1^*(\theta_2 - X)\right].$$

Since $\phi_2 \in \Phi_0$, we have that $\mathrm{dom}(\phi_2)$ has a non-empty interior around 1. Therefore, let $x_0 < 1 < y_0$ be such that $\phi_2(x_0), \phi_2(y_0) < \infty$. Then, we have that

$$\begin{aligned}
\phi_1^*(\theta_2 - X) \geq \max\{&x_0(\theta_2 - X) - \phi_1(x_0),\\
&y_0(\theta_2 - X) - \phi_1(y_0)\}.
\end{aligned}$$

Hence,

$$\begin{aligned}
K(\theta_1,\theta_2) \geq \max\{&(x_0 - 1)\theta_2 - \phi_1(x_0) - x_0\mathbb{E}_{\mathbb{P}_0}[X],\\
&(y_0 - 1)\theta_2 - \phi_1(y_0) - x_0\mathbb{E}_{\mathbb{P}_0}[X]\}.
\end{aligned}$$

By assumption, there exists a (θ_1^*,θ_2^*) such that $K(\theta_1^*,\theta_2^*) < \infty$. We now claim that θ_2 can be restricted on the set

$$\left[\frac{1}{x_0-1}(\phi_1(x_0)+x_0\mathbb{E}_{\mathbb{P}_0}[X]+K(\theta_1^*,\theta_2^*)),\ \frac{1}{y_0-1}(\phi_1(y_0)+y_0\mathbb{E}_{\mathbb{P}_0}[X]+K(\theta_1^*,\theta_2^*))\right]. \quad \text{(EC.45)}$$

Indeed, if $\theta_2 < \frac{1}{x_0-1}(\phi_1(x_0)+x_0\mathbb{E}_{\mathbb{P}_0}[X]+K(\theta_1^*,\theta_2^*))$, then since $(x_0-1)<0$, we have that

$$\begin{aligned} K(\theta_1,\theta_2) &\geq (x_0-1)\theta_2-\phi_1(x_0)-x_0\mathbb{E}_{\mathbb{P}_0}[X] \\ &> K(\theta_1^*,\theta_2^*). \end{aligned}$$

Similarly, if $\theta_2 > \frac{1}{y_0-1}(\phi_1(y_0)+y_0\mathbb{E}_{\mathbb{P}_0}[X]+K(\theta_1^*,\theta_2^*))$, then $K(\theta_1,\theta_2)>K(\theta_1^*,\theta_2^*)$.

Now, let $L_2>0$ be a number such that $[-L_2,L_2]$ contains (EC.45). We may thus restrict θ_2 on $[-L_2,L_2]$. Then, by the non-decreasing property of ϕ_1^*,ϕ_2^*, we have that

$$\begin{aligned} K(\theta_1,\theta_2) &\geq -\theta_1-L_2+\mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*(\phi_1^*(-L_2-X)+\theta_1)\right] \\ &\geq -\theta_1-L_2+\mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*(-L_2-X+\theta_1)\right] \end{aligned} \quad \text{(EC.46)}$$

Again, let $\tilde{x}_0<1<\tilde{y}_0$ be such that $\phi_2^*(\tilde{x}_0),\phi_2^*(\tilde{y}_0)<\infty$. Then, with exactly the same argument, we may restrict θ_1 on the interval

$$\begin{aligned} &\left[\frac{1}{\tilde{x}_0-1}\left((\tilde{x}_0+1)L_2+\tilde{x}_0\mathbb{E}_{\mathbb{P}_0}[X]+\phi_2(\tilde{x}_0)+K(\theta_1^*,\theta_2^*)\right)\right. \\ &\qquad \left., \ \frac{1}{\tilde{y}_0-1}\left((\tilde{y}_0+1)L_2+\tilde{y}_0\mathbb{E}_{\mathbb{P}_0}[X]+\phi_2(\tilde{y}_0)+K(\theta_1^*,\theta_2^*)\right)\right] \end{aligned}$$

Hence, θ_1,θ_2 can be both restricted on compact sets, without changing the optimum. We now examine

$$V(\theta_1,\theta_2,\lambda) := -\theta_1-\theta_2+\lambda r+\mathbb{E}_{\mathbb{P}_0}\left[\lambda\phi_2^*\left(\frac{\phi_1^*(\theta_2-X)+\theta_1}{\lambda}\right)\right].$$

We note that since $\lambda\phi_2^*\left(\frac{s}{\lambda}\right)=\sup_{t\geq 0}\{st-\lambda\phi_2(s)\}=(\lambda\phi_2)^*(s)$ and $\phi_2\geq 0$, we have that $\lambda\phi_2^*\left(\frac{s}{\lambda}\right)$ is non-increasing in λ, and that $(\lambda\phi_2)^*(s)\geq s,\ \forall s\in\mathbb{R}$. Therefore, for any $\theta_1,\theta_2\in\mathbb{R}$ and $\lambda>0$, we have

$$\begin{aligned} V(\theta_1,\theta_2,\lambda) &= -\theta_1-\theta_2+\lambda r+\mathbb{E}_{\mathbb{P}_0}\left[(\lambda\phi_2)^*(\phi_1^*(\theta_2-X)+\theta_1)\right] \\ &\geq -\theta_1-\theta_2+\lambda r+\mathbb{E}_{\mathbb{P}_0}\left[\phi_1^*(\theta_2-X)+\theta_1\right] \\ &\geq \lambda r-\mathbb{E}_{\mathbb{P}_0}[X]. \end{aligned}$$

Since $V(\theta_1^*,\theta_2^*,1)=K(\theta_1^*,\theta_2^*)<\infty$, we have that λ may be restricted on the set

$$\left[0,\frac{V(\theta_1^*,\theta_2^*,1)+\mathbb{E}_{\mathbb{P}_0}[X]}{r}\right].$$

Let L_λ denotes this upper bound of λ. Then, we have that for all $\lambda \in [0, L_\lambda]$,

$$\begin{aligned} V(\theta_1, \theta_2, \lambda) &\geq -\theta_1 - \theta_2 + \mathbb{E}_{\mathbb{P}_0}\left[L_\lambda \phi_2^*\left(\frac{\phi_1^*(\theta_2 - X) + \theta_1}{L_\lambda}\right)\right] \\ &\geq -\theta_2 + \mathbb{E}_{\mathbb{P}_0}\left[\phi_1^*(\theta_2 - X)\right]. \end{aligned}$$

Hence, θ_2 can be bounded by the same interval as in (EC.45), thus we may assume that $\theta_2 \in [-L_2, L_2]$ for some $L_2 > 0$. Then, we also have

$$\begin{aligned} V(\theta_1, \theta_2, \lambda) &\geq -\theta_1 - L_2 + \mathbb{E}_{\mathbb{P}_0}\left[L_\lambda \phi_2^*\left(\frac{\phi_1^*(-L_2 - X) + \theta_1}{L_\lambda}\right)\right] \\ &\geq -\theta_1 - L_2 + \mathbb{E}_{\mathbb{P}_0}\left[L_\lambda \phi_2^*\left(\frac{-L_2 - X + \theta_1}{L_\lambda}\right)\right] \\ &= -\theta_1 - L_2 + \mathbb{E}_{\mathbb{P}_0}\left[(L_\lambda \phi_2)^*\left(-L_2 - X + \theta_1\right)\right]. \end{aligned}$$

Since $L_\lambda \phi_2$ is also a divergence function belonging to Φ_0. It follows from the analysis of (EC.46) that θ_1 can also be bounded in the interval

$$\left[\frac{1}{\tilde{x}_0 - 1}\left((\tilde{x}_0 + 1)L_2 + \tilde{x}_0 \mathbb{E}_{\mathbb{P}_0}[X] + L_\lambda \phi_2(\tilde{x}_0) + K(\theta_1^*, \theta_2^*)\right)\right.$$
$$\left., \ \frac{1}{\tilde{y}_0 - 1}\left((\tilde{y}_0 + 1)L_2 + \tilde{y}_0 \mathbb{E}_{\mathbb{P}_0}[X] + L_\lambda \phi_2(\tilde{y}_0) + K(\theta_1^*, \theta_2^*)\right)\right],$$

where $\tilde{x}_0 < 1 < \tilde{y}_0$ are such that $\tilde{x}_0, \tilde{y}_0 \in \mathrm{dom}(\phi_2)$. □

Proof of Theorem 3. We verify the six conditions of Theorem 5.4 in Shapiro et al. [2009]. We have that (7) is equal to

$$\inf_{\theta_1, \theta_2 \in \mathbb{R}} \mathbb{E}_{\mathbb{P}_0}\left[\phi_2^*(\phi_1^*(\theta_2 - X) + \theta_1) - \theta_1 - \theta_2\right] =: \inf_{\boldsymbol{\theta} \in \mathbb{R}^2} \mathbb{E}_{\mathbb{P}_0}[F(\boldsymbol{\theta}, X)],$$

and (8) is equal to

$$\inf_{\lambda \geq 0, \theta_1, \theta_2 \in \mathbb{R}} \mathbb{E}_{\mathbb{P}_0}\left[\lambda \phi_2^*\left(\frac{\phi_1^*(\theta_2 - X) + \theta_1}{\lambda}\right) - \theta_1 - \theta_2 + \lambda r\right] =: \inf_{\lambda \geq 0, \boldsymbol{\theta} \in \mathbb{R}^2} \mathbb{E}_{\mathbb{P}_0}[G(\boldsymbol{\theta}, \lambda, X)],$$

where we replaced $\lambda > 0$ with $\lambda \geq 0$, which does not change the optimization problem.

By our assumptions $\mathrm{dom}(\phi_2^*) = \mathrm{dom}(\phi_1^*) = \mathbb{R}$, it follows from convexity that ϕ_1^*, ϕ_2^* are continuous on $\mathbb{R}$. Therefore, the functions $F(\boldsymbol{\theta}, X(\omega))$ and $G((\boldsymbol{\theta}, \lambda, X(\omega))$ are convex, lower-semicontinuous functions in $\boldsymbol{\theta}$ or $(\boldsymbol{\theta}, \lambda)$ with non-empty interior effective domain, for each $\omega \in \Omega$. Clearly, for each $\boldsymbol{\theta}$ or $(\boldsymbol{\theta}, \lambda)$, they are also measurable functions in ω. Hence, it follows from Proposition 14.39 of Rockafellar and Wets [2004] that they are normal convex integrand. Therefore, conditions (i) and (ii) of Theorem 5.4 are satisfied. The sets $\mathbb{R}^2$ and $\mathbb{R}^2 \times [0, \infty)$ are also closed and convex. Thus, (iii) of Theorem 5.4 is satisfied. We also have that for any $\boldsymbol{\theta} \in \mathbb{R}^2$ and $\omega \in \Omega$, that $F(\boldsymbol{\theta}, X(\omega)) \geq -X(\omega)$ and for any $\boldsymbol{\theta} \in \mathbb{R}^2, \lambda \geq 0$ and $\omega \in \Omega$, that $G(\boldsymbol{\theta}, \lambda, X(\omega)) \geq -X(\omega) + \lambda r \geq -X(\omega)$. Since X is assumed to be $\mathbb{P}_0$-integrable, it follows from Theorem 7.42 of Shapiro et al. [2009] that condition (iv) of Theorem 5.4 holds. Condition (v), which requires boundedness and non-emptiness of the set of solutions follows from Proposition 5 and the assumption. Finally, condition (vi), requires that the law of large number must hold pointwise for $F(\boldsymbol{\theta}, X)$ and $G(\boldsymbol{\theta}, \lambda, X)$. For cases where $\mathbb{E}_{\mathbb{P}_0}[F(\boldsymbol{\theta}, X)] < \infty$ or $\mathbb{E}_{\mathbb{P}_0}[G(\boldsymbol{\theta}, \lambda, X)] < \infty$, this is trivial. For other cases, this holds by Theorem 2.4.5 of Durrett [2019], due to the fact that $F(\boldsymbol{\theta}, X) \geq -X$ and $G(\boldsymbol{\theta}, \lambda, X) \geq -X$, with $\mathbb{E}_{\mathbb{P}_0}[-X] < \infty$. $\square$

EC.2 Minimizing Convex Integral

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Consider the convex integral minimization problem over integrable functions.

$$J_\beta(\mathbf{a}) := \inf_{\mathbf{g} \in \mathcal{L}^1_+(\mathbf{a})} \int_\Omega \beta(\omega, \mathbf{g}(\omega)) \mathrm{d}\mathbb{P}(\omega), \tag{EC.47}$$

where $\mathbf{a} \in \mathbb{R}^d$ and the set of d-dimensional non-negative $\mathbb{P}$-integrable functions with value $\mathbf{a}$, denoted as:

$$\mathcal{L}^1_+(\mathbf{a}) = \left\{ \mathbf{g} = (g_1, \ldots, g_d) \;\middle|\; g_i(\omega) \geq 0, \; \int_\Omega g_i(\omega) \mathrm{d}\mathbb{P}(\omega) = a_i, \forall i = 1, \ldots, d \right\}.$$

Here, $\beta(\omega, \mathbf{y}) : \mathbb{R}^d \to \mathbb{R}$ is a convex function in $\mathbf{y}$ for $\mathbb{P}$-almost every ω and a normal convex integrand in the sense of Rockafellar and Wets [2004]. Furthermore, we assume that $\mathrm{dom}(\beta(\omega, \mathbf{y})) \subset [0, \infty)^d$ for all $\omega \in \Omega$. Therefore, (EC.47) is also equivalent to

$$J_\beta(\mathbf{a}) = \inf_{\mathbf{g} \in \mathcal{L}^1(\mathbf{a})} \int_\Omega \beta(\omega, \mathbf{g}(\omega)) \mathrm{d}\mathbb{P}(\omega), \tag{EC.48}$$

where $\mathcal{L}^1(\mathbf{a})$ is the set of all $\mathbb{P}$-integrable functions with integral equals to $\mathbf{a}$:

$$\mathcal{L}^1(\mathbf{a}) = \left\{ \mathbf{g} = (g_1, \ldots, g_d) \;\middle|\; \int_\Omega g_i(\omega) \mathrm{d}\mathbb{P}(\omega) = a_i, \forall i = 1, \ldots, d \right\}.$$

In the sequel, we let $\mathcal{L}^1 := \bigcup_{\mathbf{a} \in \mathbb{R}^d} \mathcal{L}^1(\mathbf{a})$ denote the set of all $\mathbb{P}$-integrable functions. Problem (EC.48) can be reformulated using Lagrangian duality. Since we are optimizing over density functions, which belong to an infinite-dimensional vector space of integrable functions, we invoke the following Lagrangian duality theorem given in Luenberger [1969], which we state here for completeness.

Theorem EC.2.1 (Luenberger 1969, exercise 8.8.7)**.** *Let $\mathcal{X}$ be a real vector space, $\mathcal{C}$ a convex subset of $\mathcal{X}$. Let f be a real-valued convex functional on $\mathcal{C}$ and $G : \mathcal{C} \to \mathbb{R}^d$ a convex mapping, i.e., $G(\lambda x_1 + (1-\lambda) x_2) \leq_{\mathbb{R}^d} \lambda G(x_1) + (1-\lambda) G(x_2), \forall x_1, x_2 \in \mathcal{C}, \lambda \in (0,1)$ ($x \leq_{\mathbb{R}^d} y \Leftrightarrow x_i \leq y_i, \forall i$). Let $H : \mathcal{X} \to \mathbb{R}^m$ be an affine mapping, i.e, $H(x) = Ax + b$ for some linear mapping A. Consider the optimization problem,*

$$\mu_0 = \inf_{x \in \mathcal{X}} \{ f(x) \mid x \in \mathcal{C}, G(x) \leq \mathbf{0}, H(x) = \mathbf{0} \},$$

where $\mathbf{0}$ denotes the zero-vector. Assume the existence of a $x_1 \in \mathcal{C}$ such that $G(x_1) <_{\mathbb{R}^d} \mathbf{0}$ and $H(x_1) = \mathbf{0}$. Suppose also that $\mathbf{0}$ is an interior point of the image set $\{y \in \mathbb{R}^m \mid H(x) =$

y for some $x \in \mathcal{C}\}$. If μ_0 is finite, then there exists a $\boldsymbol{\mu}^* \in \mathbb{R}^d$ with $\mu \geq_{\mathbb{R}^d} \mathbf{0}$ and $\boldsymbol{\nu}^* \in \mathbb{R}^m$, such that

$$\begin{aligned}
\mu_0 &= \inf_{x \in \mathcal{C}} \{f(x) + \boldsymbol{\mu}^{*T} G(x) + \boldsymbol{\nu}^{*T} H(x)\} \\
&= \sup_{\substack{\boldsymbol{\mu} \geq_{\mathbb{R}^d} \mathbf{0} \\ \boldsymbol{\nu} \in \mathbb{R}^m}} \inf_{x \in \mathcal{C}} \{f(x) + \boldsymbol{\mu}^T G(x) + \boldsymbol{\nu}^T H(x)\}.
\end{aligned}$$

To apply Theorem EC.2.1, we set $\mathcal{X} = (\mathcal{L}^1)^d$ and $H(\mathbf{g}) = \int_\Omega \mathbf{g}(\omega) \mathrm{d}\mathbb{P}(\omega) - \mathbf{a}$. Note that the zero vector $\mathbf{0} \in \mathbb{R}^d$ is an interior point of the image set of H, since any vector $\boldsymbol{\epsilon} \in \mathbb{R}^d$ sufficiently close to $\mathbf{0}$ is the image of $H(\mathbf{a} + \boldsymbol{\epsilon})$ of the constant function $\mathbf{g}(\omega) = \mathbf{a} + \boldsymbol{\epsilon}$. Hence, Theorem EC.2.1 can be applied to $J_\beta(\mathbf{a})$ if $J_\beta(\mathbf{a}) < +\infty$.

Theorem EC.2.2. *Assume $J_\beta(\mathbf{a}) < +\infty$. Then, we have the duality*

$$J_\beta(\mathbf{a}) = \sup_{\boldsymbol{\theta} \in \mathbb{R}^d} \boldsymbol{\theta}^T \mathbf{a} - K_\beta(\boldsymbol{\theta}), \tag{EC.49}$$

where $K_\beta(\boldsymbol{\theta}) = \int_\Omega \beta^*(\omega, \boldsymbol{\theta}) \mathrm{d}\mathbb{P}(\omega)$ and $\beta^*(\omega, \boldsymbol{\theta})$ is the convex conjugate of $\beta(\omega, \mathbf{y})$. Moreover, if $J_\beta(\mathbf{a}) > -\infty$, then there exists $\boldsymbol{\theta}^* \in \mathbb{R}^d$ such that the supremum in (EC.49) *is attained.*

Proof. Applying Theorem EC.2.1 to (EC.48) yields that

$$J_\beta(\mathbf{a}) = \sup_{\boldsymbol{\theta} \in \mathbb{R}^d} \boldsymbol{\theta}^T \mathbf{a} + \inf_{\mathbf{g} \in \mathcal{L}^1} \int_\Omega \beta(\omega, \mathbf{g}(\omega)) - \boldsymbol{\theta}^T \mathbf{g}(\omega) \mathrm{d}\mathbb{P}(\omega).$$

We have,

$$\begin{aligned}
&\inf_{\mathbf{g} \in \mathcal{L}^1} \int_\Omega \beta(\omega, \mathbf{g}(\omega)) - \boldsymbol{\theta}^T \mathbf{g}(\omega) \mathrm{d}\mathbb{P}(\omega) \\
&\overset{(*)}{=} \int_\Omega \inf_{\mathbf{y} \in \mathbb{R}^d} (\beta(\omega, \mathbf{y}) - \boldsymbol{\theta}^T \mathbf{y}) \mathrm{d}\mathbb{P}(\omega) \\
&\overset{(**)}{=} \int_\Omega \inf_{\mathbf{y} \geq \mathbf{0}} (\beta(\omega, \mathbf{y}) - \boldsymbol{\theta}^T \mathbf{y}) \mathrm{d}\mathbb{P}(\omega) \\
&= -\int_\Omega \beta^*(\omega, \boldsymbol{\theta}) \mathrm{d}\mathbb{P}(\omega),
\end{aligned}$$

where the interchange of infimum and integral in $(*)$ follows from [Rockafellar and Wets [2004], Theorem 14.60], for which we may apply the theorem since $J_\beta(\mathbf{a}) < \infty$ and $\mathcal{L}^1$ is decomposable (see Rockafellar and Wets [2004], Definition 14.59). In $(**)$ we used that $\beta(\omega, \mathbf{y}) = +\infty$ if $\mathbf{y}$ has negative component. Finally, the existence of $\boldsymbol{\theta}^*$ is guaranteed by Theorem EC.2.1. □

Hence, the *primal problem* $J_\beta(\mathbf{a})$ can be reformulated into a finite-dimensional *dual problem* as in (EC.49). We call the solutions of the corresponding problems the *primal* and *dual solution* respectively. The following Theorem establishes the condition in which a convex function of the form $\boldsymbol{\theta} \mapsto K_\beta(\boldsymbol{\theta}) = \int_\Omega \beta^*(\omega, \boldsymbol{\theta}) \mathrm{d}\mathbb{P}(\omega)$ is differentiable.

Theorem EC.2.3. *Let $\boldsymbol{\theta} \in \mathbb{R}^d$. If $\operatorname{int}(\operatorname{dom}(K_\beta)) \neq \emptyset$ and $\beta^*(\omega, \cdot)$ is differentiable on its interior* *effective domain for $\mathbb{P}$-a.e. $\omega \in \Omega$. Then, K_β is differentiable on* $\mathrm{int}(\mathrm{dom}(K_\beta))$ *and we have*

$$\nabla K_\beta(\boldsymbol{\theta}) = \int_\Omega \nabla \beta^*(\omega, \boldsymbol{\theta}) \mathrm{d}\mathbb{P}(\omega).$$

Proof. Since K_β is a convex function, we can show its differentiability by showing its directional derivative function is linear (Rockafellar [1970], Theoreom 25.2). Let $\boldsymbol{\theta}_0 \in \mathrm{int}(\mathrm{dom}(K_\beta))$. For any other $\boldsymbol{\theta} \in \mathrm{int}(\mathrm{dom}(K_\beta))$, we examine its directional derivative

$$\lim_{t \downarrow 0} \frac{1}{t} \left(K_\beta(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)) - K_\beta(\boldsymbol{\theta}_0) \right) = \lim_{t \downarrow 0} \int_\Omega \frac{1}{t} \left(\beta^*(\omega, \boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)) - \beta^*(\omega, \boldsymbol{\theta}_0) \right) \mathrm{d}\mathbb{P}(\omega).$$

Since, $\boldsymbol{\theta}_0 \in \mathrm{int}(\mathrm{dom}(K_\beta))$, we have that $\boldsymbol{\theta}_0 \in \mathrm{int}(\mathrm{dom}(\beta^*(\omega, \cdot)))$ for $\mathbb{P}$-a.e. $\omega \in \Omega$. For these ω, the difference quotient

$$q_\omega(t) := \frac{1}{t}(\beta^*(\omega, \boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)) - \beta^*(\omega, \boldsymbol{\theta}_0))$$

is non-increasing as $t \downarrow 0$ (Theorem 23.1, Rockafellar [1970]). Therefore, the quotient $q_\omega(t) - q_\omega(\frac{1}{2})$ is negative and non-increasing for $t \leq 1/2$. Hence, applying the monotone convergence theorem and removing the term $q_\omega(\frac{1}{2})$ afterward (note that $q_\omega(\frac{1}{2})$ is a finite number since $\boldsymbol{\theta}, \boldsymbol{\theta}_0 \in \mathrm{dom}(\beta^*(\omega, \cdot))$, for $\mathbb{P}$-a.e. $\omega \in \Omega$) shows that K_β is differentiable at $\boldsymbol{\theta}_0$. Indeed,

$$\begin{aligned}
\lim_{t \downarrow 0} \frac{1}{t} \left(K_\beta(\boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)) - K_\beta(\boldsymbol{\theta}_0) \right) &= \int_\Omega \lim_{t \downarrow 0} \frac{1}{t}(\beta^*(\omega, \boldsymbol{\theta}_0 + t(\boldsymbol{\theta} - \boldsymbol{\theta}_0)) - \beta^*(\omega, \boldsymbol{\theta}_0)) \mathrm{d}\mathbb{P}(\omega) \\
&\overset{(*)}{=} \int_\Omega \langle \boldsymbol{\theta} - \boldsymbol{\theta}_0, \nabla \beta^*(\omega, \boldsymbol{\theta}_0) \rangle \mathrm{d}\mathbb{P}(\omega) \\
&\overset{(**)}{=} \left\langle \boldsymbol{\theta} - \boldsymbol{\theta}_0, \int_\Omega \nabla \beta^*(\omega, \boldsymbol{\theta}_0) \mathrm{d}\mathbb{P}(\omega) \right\rangle,
\end{aligned}$$

where in $(*)$ we used the differentiability of $\beta^*(\omega, \cdot)$ and in $(**)$ we used the linearity of inner product. $\square$

Theorem EC.2.4. *Assume $J_\beta(\mathbf{a}) < +\infty$. Let $\boldsymbol{\theta}^*$ be such that $K_\beta(\boldsymbol{\theta}^*) < \infty$. If the partial derivatives $\mathbf{g}_{\boldsymbol{\theta}^*}(\omega) = \nabla \beta^*(\omega, \boldsymbol{\theta}^*)$ exist for $\mathbb{P}$-almost every ω and satisfy:*

$$\int_\Omega \mathbf{g}_{\boldsymbol{\theta}^*}(\omega) \mathrm{d}\mathbb{P}(\omega) = \mathbf{a}. \tag{EC.50}$$

Then, $\boldsymbol{\theta}^*$ is a dual solution and $\mathbf{g}_{\boldsymbol{\theta}^*}$ is a primal solution. On the other hand, if a primal solution $\mathbf{g}^*$ exists, then $\mathbf{g}^* = \mathbf{g}_{\boldsymbol{\theta}^*}$ for a particular dual solution $\boldsymbol{\theta}^*$, if the partial derivatives $\nabla \beta^*(\omega, \boldsymbol{\theta}^*)$ at this particular $\boldsymbol{\theta}^*$ exists, for $\mathbb{P}$-almost every $\omega \in \Omega$.

Proof. By the definition of the conjugate, we have for any $\boldsymbol{\theta} \in \mathbb{R}^d$ and functions $\mathbf{g}$, the inequality

$$\beta(\omega, \mathbf{g}(\omega)) + \beta^*(\omega, \boldsymbol{\theta}) \geq \boldsymbol{\theta}^T \mathbf{g}(\omega). \tag{EC.51}$$

Inequality (EC.51) is tight if and only if the gradient of $\beta^*(\omega, \cdot)$ at $\boldsymbol{\theta}$ exists and is equal to $\mathbf{g}(\omega)$ ([Rockafellar, 1970], Theorem 23.5). Integrating both sides of (EC.51) with respect to ω, for any function $\mathbf{g}$ such that $\int_\Omega \mathbf{g}(\omega)\mathrm{d}\mathbb{P}(\omega) = \mathbf{a}$ yields

$$\int_\Omega \beta(\omega, \mathbf{g}(\omega))\mathrm{d}\mathbb{P}(\omega) + \int_\Omega \beta^*(\omega, \boldsymbol{\theta})\mathrm{d}\mathbb{P}(\omega) \geq \boldsymbol{\theta}^T \mathbf{a}. \tag{EC.52}$$

In particular, inequality (EC.51) is an equality for $\mathbf{g}_{\boldsymbol{\theta}^*}(\omega) = \nabla\beta^*(\omega, \boldsymbol{\theta}^*)$ with $\boldsymbol{\theta}^*$ that satisfies condition (EC.50). Hence, we have that (EC.52) is also an equality and thus

$$\int_\Omega \beta(\omega, \mathbf{g}_{\boldsymbol{\theta}^*}(\omega))\mathrm{d}\mathbb{P}(\omega) = \mathbf{a}^T\boldsymbol{\theta}^* - \int_\Omega \beta^*(\omega, \boldsymbol{\theta}^*)\mathrm{d}\mathbb{P}(\omega).$$

Therefore, by Theorem EC.2.2 we conclude that $\boldsymbol{\theta}^*$ is a dual solution and $\mathbf{g}_{\boldsymbol{\theta}^*}$ is a primal solution.

Conversely, let $\mathbf{g}^*$ be a primal solution. Then, $J_\beta(\mathbf{a})$ is finite, and thus a dual solution $\boldsymbol{\theta}^*$ exists and we must have $K_\beta(\boldsymbol{\theta}^*) < \infty$. Furthermore, (EC.52) holds for $\mathbf{g}^*$ and $\boldsymbol{\theta}^*$ and is an equality. Therefore, we have that the following integral with positive integrand is zero:

$$\int_\Omega \beta(\omega, \mathbf{g}^*(\omega)) + \beta^*(\omega, \boldsymbol{\theta}^*) - (\mathbf{g}^*(\omega))^T\boldsymbol{\theta}^* \ \mathrm{d}\mathbb{P}(\omega) = 0.$$

Hence, (EC.51) is an equality for $\mathbf{g}^*$ and $\boldsymbol{\theta}^*$ for $\mathbb{P}$-a.e. $\omega \in \Omega$. Therefore, $\mathbf{g}^* = \mathbf{g}_{\boldsymbol{\theta}^*}$, by Rockafellar [1970], Theorem 23.5. □

EC.3 Characterization of Risk Aversion of Robust OCE and its Preservation of Convex Order

The formulation of the robust OCE and robust shortfall risk measures in terms of (ϕ_1^*, ϕ_2^*) also allows us to characterize risk aversion by simple convexity conditions imposed on ϕ_1^*, ϕ_2^*. We say that a robust risk measure $\rho \in \{\rho^l_{\mathrm{oce},\mathbb{P}}, \rho^l_{\mathrm{sf},\mathbb{P}}\}$ is risk-averse, if and only if

$$\rho(X) \geq \mathbb{E}_\mathbb{P}[-X], \forall X. \tag{EC.53}$$

We show that for the robust OCE and shortfall risk measures, the risk aversion property can be easily characterized by some mild conditions on ϕ_1^*, ϕ_2^*.

Proposition EC.3. *For any non-decreasing convex function ϕ_1^*, ϕ_2^* with $\phi_1^*(0) = \phi_2^*(0) = 0$, we have that $\rho^l_{\mathrm{oce},\mathbb{P}}$ satisfies* (EC.53)*, if and only if*

$$\phi_2^*(x) \geq x, \phi_1^*(x) \geq x, \forall x \in \mathbb{R}.$$

If we further assume that $\phi_1^*(x) > 0, \forall x > 0$, then we have that $\rho^l_{\mathrm{sf},\mathbb{P}}$ satisfies (EC.53)*, if and only if*

$$\phi_2^*(x) \geq x, \forall x \in \mathbb{R}.$$

Proof of Proposition EC.3. For $\rho^l_{\mathrm{oce},\mathbb{P}}$, we have that (EC.53) implies that for all $x \in \mathbb{R}$, that

$$\inf_{\theta_1,\theta_2\in\mathbb{R}} \{-\theta_1 - \theta_2 + \phi_2^*(\phi_1^*(\theta_2 - x) + \theta_1)\} \geq -x,$$

which implies in particular that for $\theta_2 = x$:

$$\phi_2^*(\theta_1) \geq \theta_1, \forall \theta_1 \in \mathbb{R}.$$

It also implies that for $\theta_1 = -\phi_1^*(x), \theta_2 = 2x$, that

$$\phi_1^*(x) \geq x, \forall x \in \mathbb{R}.$$

Conversely, if $\phi_2^*(x) \geq x, \phi_1^*(x) \geq x, \forall x \in \mathbb{R}$, then by the monotonicity of ϕ_1^*, ϕ_2^*, we have that for any $\theta_1, \theta_2 \in \mathbb{R}$ and any X:

$$-\theta_1 - \theta_2 + \mathbb{E}_{\mathbb{P}}[\phi_2^*(\phi_1^*(\theta_2 - X) + \theta_1)] \geq -\theta_1 - \theta_2 + \mathbb{E}_{\mathbb{P}}[\theta_2 - X + \theta_1] \geq \mathbb{E}_{\mathbb{P}}[-X].$$

For $\rho^l_{\mathrm{sf},\mathbb{P}}$, we have that (EC.53) implies that for all $x \in \mathbb{R}$,

$$\inf_{\theta_1,\theta_2 \in \mathbb{R}} \{\theta_2 \mid \phi_2^*(\phi_1^*(-\theta_2 - x) + \theta_1) \leq \theta_1\} \geq -x$$

This implies that for all $x, \theta_1 \in \mathbb{R}$,

$$\phi_2^*(\phi_1^*(-(-x) - x) + \theta_1) = \phi_2^*(\theta_1) \geq \theta_1.$$

On the other hand, if $\phi_2^*(x) \geq x$ for all $x \in \mathbb{R}$, then for any $\theta_1 \in \mathbb{R}$ and any random variable X,

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[\phi_2^*(\phi_1^*(-\theta_2 - X) + \theta_1)] \leq \theta_1 &\Rightarrow \mathbb{E}_{\mathbb{P}}[\phi_1^*(-\theta_2 - X)] \leq 0 \\ &\Rightarrow \phi_1^*(-\theta_2 + \mathbb{E}[-X]) \leq 0\end{aligned}$$

By the non-decreasing property of ϕ_1^* and the assumption that $\phi_1^*(x) > 0, \forall x > 0$, we can then conclude that the above inequality implies that $\theta_2 \geq \mathbb{E}[-X]$. Therefore, $\rho^l_{\mathrm{sf},\mathbb{P}}(X) \geq \mathbb{E}_{\mathbb{P}}[-X]$. $\square$

Furthermore, it is also easy to show that the robust OCE and shortfall risk measure preserves convex order. We say that X is convex less order than Y, i.e. $X \leq_{\mathrm{cv}} Y$ if for all convex function f, we have $\mathbb{E}_{\mathbb{P}}[f(X)] \leq \mathbb{E}_{\mathbb{P}}[f(Y)]$.

Proposition EC.3. *Let ϕ_1^*, ϕ_2^* be a non-decreasing convex function. Then*

$$\begin{aligned}X \leq_{\mathrm{cv}} Y \Rightarrow\ & \rho^l_{\mathrm{oce},\mathbb{P}}(X) \leq \rho^l_{\mathrm{oce},\mathbb{P}}(Y) \\ & \text{and } \rho^l_{\mathrm{sf},\mathbb{P}}(X) \leq \rho^l_{\mathrm{sf},\mathbb{P}}(Y).\end{aligned}$$

Proof of Proposition EC.3. Suppose $X \leq_{\mathrm{cv}} Y$, then, we have that for all $\theta_1, \theta_2 \in \mathbb{R}$,

$$\mathbb{E}_{\mathbb{P}}[\phi_2^*(\phi_1^*(\theta_2 - X) + \theta_1)] \geq \mathbb{E}_{\mathbb{P}}[\phi_2^*(\phi_1^*(\theta_2 - Y + \theta_1)],$$

since the function $\phi_2^*(\phi_1^*(\theta_2 - x) + \theta_1)$ is convex in x. Hence, $\rho^l_{\mathrm{oce},\mathbb{P}}(X) \leq \rho^l_{\mathrm{oce},\mathbb{P}}(Y)$.

Similarly, the function $\phi_2^*(\phi_1^*(-\theta_2 - x) + \theta_1)$ is also convex in x. Therefore, for any $\theta_1, \theta_2 \in \mathbb{R}$, if $X \leq_{\text{cv}} Y$, then

$$\mathbb{E}[\phi_2^*(\phi_1^*(-\theta_2 - Y) + \theta_1)] \leq \theta_1 \Rightarrow \mathbb{E}[\phi_2^*(\phi_1^*(-\theta_2 - X) + \theta_1)] \leq \theta_1.$$

Hence, $\rho^l_{\text{sf},\mathbb{P}}(X) \leq \rho^l_{\text{sf},\mathbb{P}}(Y)$. $\square$

EC.4 Additional Details on Tail Analysis

In this section, we examine conditions (9) and (10) for the combinations of ϕ_2-divergences and nominal distributions f_0 (distribution of X under $\mathbb{P}_0$) that are displayed in Table 1 and 2, for respectively the Conditional-Value-at-Risk and the entropic risk measure. We examine the following nominal distributions $f_0(x)$:

- Gaussian $N(\mu, \sigma)$: $f_0(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}$
- Lognormal $\ln(\mu, \sigma^2)$: $f_0(x) = \frac{1}{|x|\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log|x|-\mu)^2}{2\sigma^2}\right), \quad x < 0$
- Weibull $W(\lambda_0, k)$: $f_0(x) = \frac{k}{\lambda_0}\left(\frac{|x|}{\lambda_0}\right)^{k-1} \exp\left(-\left(\frac{|x|}{\lambda_0}\right)^k\right), \quad x < 0$
- Pareto $P(x_m, p_0)$: $f_0(x) = \frac{p|x_m|^{p_0}}{|x|^{p_0+1}}, \quad x \leq x_m$
- Student-t $t(\nu)$: $f_0(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}$

For the divergences, we have the following conjugate functions (see Ben-Tal et al. 2013):

- Kullback-Leibler: $\phi_2^*(s) = \exp(s) - 1$
- Polynomial $p > 1$: $\phi_2^*(s) = \frac{1}{p}\max\{1 + s(p-1), 0\}^{\frac{p}{p-1}} - \frac{1}{p}$
- Modified χ^2-divergence: $\phi_2^*(s) = \begin{cases} -1 & s < -2 \\ s + s^2/4 & s \geq -2. \end{cases}$

EC.4.1 Details in Table 1

For CVaR_α, we have that $\phi_1^*(s) = \max\{s/(1-\alpha), 0\}$. We also note that all nominal distributions considered in Table 1 have $f_0(x) > 0$ for all $x \in (-\infty, c)$ for some $c \in \mathbb{R}$. Hence, for all divergences such that $\phi_2^*(s) = \infty$ if s sufficiently large (such as Burg entropy, variation distance), the corresponding robust CVaR_α risk measure is not finite.

- Gaussian vs KL-divergence: as $x \to -\infty$, we have

$$\begin{aligned}
&O\left(\lambda\phi_2^*\left(\frac{\phi_1^*(\theta_2 - x) + \theta_1}{\lambda}\right) f_0(x)\right) \\
&= O\left(\lambda\left(\exp\left(\frac{\frac{\theta_2+|x|}{1-\alpha} + \theta_1}{\lambda}\right) - 1\right) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right) \\
&= O\left(\lambda\exp\left(\frac{|x|}{\lambda(1-\alpha)} - \frac{(x-\mu)^2}{2\sigma^2}\right)\right),
\end{aligned}$$

which is integrable for any $\theta_1, \theta_2 \in \mathbb{R}, \lambda > 0$. As $x \to \infty$, we have $\phi_1^*(\theta_2 - x) = 0$ for all $x > \theta_2$. Then, the above integrand is also integrable.

- Pareto vs KL-divergence: as $x \to -\infty$, we have

$$
\begin{aligned}
&O\left(\lambda\phi_2^*\left(\frac{\phi_1^*(\theta_2 - x) + \theta_1}{\lambda}\right) f_0(x)\right) \\
&= O\left(\lambda\left(\exp\left(\frac{\frac{\theta_2+|x|}{1-\alpha} + \theta_1}{\lambda}\right) - 1\right)\frac{p_0|x_m|^{p_0}}{|x|^{p_0+1}}\right) \\
&= O\left(\lambda\exp\left(\frac{|x|}{\lambda(1-\alpha)}\right)|x|^{-p_0-1}\right),
\end{aligned}
$$

  which is not integrable for any $\theta_1, \theta_2 \in \mathbb{R}$, $\lambda \geq 0$.

- Log-normal vs KL-divergence: for $x \to -\infty$,

$$
\begin{aligned}
&O\left(\lambda\left(\exp\left(\frac{\frac{\theta_2+|x|}{1-\alpha} + \theta_1}{\lambda}\right) - 1\right)\frac{1}{|x|\sigma\sqrt{2\pi}}\exp\left(-\frac{(\log|x| - \mu)^2}{2\sigma^2}\right)\right) \\
&= O\left(\lambda\exp\left(\frac{|x|}{\lambda(1-\alpha)}\right) - \frac{(\log|x| - \mu)^2}{2\sigma^2} - \log|x|\right),
\end{aligned}
$$

  which is not integrable for any $\theta_1, \theta_2 \in \mathbb{R}$, $\lambda \geq 0$.

- Weibull vs KL-divergence: we have the tail behavior

$$
O\left(\lambda\left(\exp\left(\frac{\frac{\theta_2+|x|}{1-\alpha} + \theta_1}{\lambda}\right) - 1\right)\frac{k}{\lambda_0}\left(\frac{|x|}{\lambda_0}\right)^{k-1}\exp\left(-\left(\frac{|x|}{\lambda_0}\right)^k\right)\right)
$$

  For $\lambda > 0$, this is integrable if $k > 1$, and not integrable if $k < 1$. For $k = 1$, we note that if we choose any $\lambda > \frac{\lambda_0}{1-\alpha}$, then $\frac{1}{\lambda(1-\alpha)} - \frac{1}{\lambda_0} < 0$. Hence, condition (9) is satisfied.

- Student's t vs KL-divergence: As $x \to -\infty$, we have

$$
O\left(\lambda\left(\exp\left(\frac{\frac{\theta_2+|x|}{1-\alpha} + \theta_1}{\lambda}\right) - 1\right)\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}\right)
$$

  Clearly, the exponential term $\exp\left(\frac{|x|}{\lambda(1-\alpha)}\right)$ dominates the polynomial term $\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$. Hence, we have no integrability for any $\theta_1, \theta_2 \in \mathbb{R},\ \lambda \geq 0$.

- Gaussian vs Polynomial $p > 1$, we examine the tail behavior for $x \to -\infty$ (the case for $x \to \infty$ follows similarly, where $\phi_1^*(\theta_2 - x) = 0$ for all $x > \theta_2$), we have

$$
O\left(\lambda\left(1 + \left(\frac{\theta_2 + |x|}{\lambda(1-\alpha)} + \frac{\theta_1}{\lambda}\right)(p-1)\right)^{\frac{p}{p-1}}\exp\left(-(x-\mu)^2/(2\sigma^2)\right)\right),
$$

  which is integrable for any $\theta_1, \theta_2 \in \mathbb{R}$, $\lambda > 0$.

- Log-normal vs Polynomial $p > 1$, we study the tail behavior

$$
\begin{aligned}
&O\left(\lambda\left(1+\left(\frac{\theta_2+|x|}{\lambda(1-\alpha)}+\frac{\theta_1}{\lambda}\right)(p-1)\right)^{\frac{p}{p-1}}\frac{1}{|x|\sigma\sqrt{2\pi}}\exp\left(-\frac{(\log|x|-\mu)^2}{2\sigma^2}\right)\right)\\
&=O\left(|x|^{p/(p-1)-1}\exp\left(-\frac{(\log|x|-\mu)^2}{2\sigma^2}\right)\right)\\
&=O\left(\exp\left(-\frac{\log^2|x|}{2\sigma^2}+\frac{1}{p-1}\log(x)\right)\right)
\end{aligned}
$$

which is integrable for any $\theta_1, \theta_2 \in \mathbb{R}$, $\lambda > 0$.

- Weibull vs Polynomial $p > 1$, the tail behavior is

$$
O\left(\lambda\left(1+\left(\frac{\theta_2+|x|}{\lambda(1-\alpha)}+\frac{\theta_1}{\lambda}\right)(p-1)\right)^{\frac{p}{p-1}}\frac{k}{\lambda_0}\left(\frac{|x|}{\lambda_0}\right)^{k-1}\exp\left(-\left(\frac{|x|}{\lambda_0}\right)^k\right)\right)
$$

which is integrable for any $\theta_1, \theta_2 \in \mathbb{R}$, $\lambda > 0$, since $k > 0$.

- Pareto vs Polynomial $p > 1$, the tail behavior is

$$
\begin{aligned}
&O\left(\lambda\left(1+\left(\frac{\theta_2+|x|}{\lambda(1-\alpha)}+\frac{\theta_1}{\lambda}\right)(p-1)\right)^{\frac{p}{p-1}}\frac{p_0|x_m|^{p_0}}{|x|^{p_0+1}}\right)\\
&=O\left(|x|^{p/(p-1)-p_0-1}\right),
\end{aligned}
$$

which is integrable for for any $\theta_1, \theta_2 \in \mathbb{R}$, $\lambda > 0$, if and only if $\frac{p}{p-1} < p_0$.

- Student's t vs Polynomial $p > 1$: As $x \to -\infty$, we have

$$
\begin{aligned}
&O\left(\lambda\left(1+\left(\frac{\theta_2+|x|}{\lambda(1-\alpha)}+\frac{\theta_1}{\lambda}\right)(p-1)\right)^{\frac{p}{p-1}}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}\right)\\
&=O\left(|x|^{p/(p-1)-\nu-1}\right),
\end{aligned}
$$

which is integrable if and only if $p/(p-1) < \nu$.

- Modified χ^2 vs Gaussian: we have

$$
O\left(\lambda\left(\frac{\frac{\theta_2+|x|}{1-\alpha}+\theta_1}{\lambda}\right)^2\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right),
$$

which is integrable for any $\theta_1, \theta_2 \in \mathbb{R}, \lambda > 0$.

- Log-normal vs Modified χ^2: we have

$$
O\left(\lambda\left(\frac{\frac{\theta_2+|x|}{1-\alpha}+\theta_1}{\lambda}\right)^2\frac{1}{|x|}\exp\left(-\frac{(\log|x|-\mu)^2}{2\sigma^2}\right)\right),
$$

which is integrable for any $\theta_1, \theta_2 \in \mathbb{R}, \lambda > 0$.

- Pareto vs Modified χ^2: as $x \to -\infty$, we have

$$O\left(\lambda\left(\frac{\frac{\theta_2+|x|}{1-\alpha}+\theta_1}{\lambda}\right)^2 |x|^{-p_0-1}\right),$$

which is integrable if and only if $2-1-p_0 < -1$, which is equivalent to $p_0 > 2$.

- Modified χ^2 vs Weibull: as $x \to -\infty$, we have

$$O\left(\lambda\left(\frac{\frac{\theta_2+|x|}{1-\alpha}+\theta_1}{\lambda}\right)^2 \left(\frac{|x|}{\lambda_0}\right)^{k-1} \exp\left(-\left(\frac{|x|}{\lambda_0}\right)^k\right)\right),$$

which is integrable for $k > 0$ and any $\theta_1, \theta_2 \in \mathbb{R}, \lambda > 0$.

- Student's t vs Modified χ^2: as $x \to -\infty$, we have

$$O\left(\lambda\left(\frac{\frac{\theta_2+|x|}{1-\alpha}+\theta_1}{\lambda}\right)^2 \left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}\right),$$

which is integrable for any $\theta_1, \theta_2 \in \mathbb{R}, \lambda > 0$ if and only if $\nu > 2$.

EC.4.2 Details in Table 2

For the entropic risk measure $\rho_{e,\gamma}(X) = \log\left(\mathbb{E}[\exp(-\gamma X)]\right)/\gamma$, $\gamma > 0$, we have $\phi_2^*(s) = (\exp(s/\gamma)-1)/\gamma$. We note that Log-normal, Pareto, Weibull with $k < 1$ and Student's t distribution do not have a finite moment generating function. Hence, the robust entropic risk measure for those nominal distribution is not finite for any divergences. We also note that for all divergences such that $\phi_2^*(s) = \infty$ if s sufficiently large (such as Burg entropy, variation distance), the corresponding robust entropic risk measure is not finite, for the nominal distributions that we consider in Table 2.

- Weibull vs KL-divergence: we have

$$\begin{aligned}
&O\left(\lambda\phi_2^*\left(\frac{\phi_1^*(\theta_2-x)+\theta_1}{\lambda}\right) f_0(x)\right) \\
&= O\left(\lambda\exp\left(\frac{\frac{1}{\gamma}(\exp((\theta_2-x)/\gamma-1))+\theta_1}{\lambda}\right)\left(\frac{|x|}{\lambda_0}\right)^{k-1}\exp\left(-\left(\frac{|x|}{\lambda_0}\right)^k\right)\right),
\end{aligned}$$

which is not integrable for any $k > 0$, and $\theta_1, \theta_2 \in \mathbb{R}, \lambda \geq 0$.

- Gaussian vs KL-divergence: we have

$$O\left(\lambda\exp\left(\frac{\frac{1}{\gamma}(\exp((\theta_2-x)/\gamma-1))+\theta_1}{\lambda}\right)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right),$$

which is not integrable for any $\theta_1, \theta_2 \in \mathbb{R}, \lambda \geq 0$.

- Weibull vs Polynomial $p > 1$: we have

$$O\left(\left(\frac{1}{\gamma}(\exp((\theta_2 - x)/\gamma - 1))\right)^{\frac{p}{p-1}} \left(\frac{|x|}{\lambda_0}\right)^{k-1} \exp\left(-\left(\frac{|x|}{\lambda_0}\right)^k\right)\right)$$

We note that the integrability of the integrand above is mainly determined by the growth of

$$\exp\left(\frac{|x|}{\gamma}\frac{p}{p-1} - \left(\frac{|x|}{\lambda_0}\right)^k\right)$$

as $x \to -\infty$.

If $k > 1$, then we clearly have integrability. If $k = 1$, we have integrability if and only if $\frac{p}{p-1} < \frac{\gamma}{\lambda_0}$.

- Gaussian vs Polynomial $p > 1$: we have

$$O\left(\left(\frac{\frac{1}{\gamma}\left(\exp\left(\frac{\theta_2 - x}{\gamma}\right) - 1\right) + \theta_1}{\lambda}(p-1)\right)^{\frac{p}{p-1}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right),$$

the tail behaviour of this integrand is determined by the growth of

$$\exp\left(\frac{|x|}{\gamma}\frac{p}{p-1} - \frac{(x-\mu)^2}{2\sigma^2}\right),$$

which is integrable.

- Weibull vs Modified χ^2: we have

$$O\left(\lambda\left(\frac{\frac{1}{\gamma}(\exp((\theta_2 - x)/\gamma - 1)) + \theta_1}{\lambda}\right)^2 \left(\frac{|x|}{\lambda_0}\right)^{k-1} \exp\left(-\left(\frac{|x|}{\lambda_0}\right)^k\right)\right),$$

for which the integrability depends on

$$\exp\left(\frac{|x|^2}{\gamma^2} - \frac{|x|^k}{\lambda_0^k}\right)\left(\frac{|x|}{\lambda_0}\right)^{k-1}.$$

For $k > 2$, we have integrability, whereas for $k < 2$ we do not. If $k = 2$, then this is integrable if and only if $\gamma > \lambda_0$.

- Gaussian vs Modified χ^2: we have

$$O\left(\lambda\left(\frac{\frac{1}{\gamma}(\exp((\theta_2 - x)/\gamma - 1)) + \theta_1}{\lambda}\right)^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right),$$

for which the integrability depends on

$$\exp\left(\frac{x^2}{\gamma^2} - \frac{x^2}{2\sigma^2}\right),$$

which is integrable if and only if $\gamma > 2\sigma^2$.