Measuring Distribution Model Risk*

Thomas Breuer Imre Csiszár

17 May 2013

Abstract

We propose to interpret distribution model risk as sensitivity of expected loss to changes in the risk factor distribution, and to measure the distribution model risk of a portfolio by the maximum expected loss over a set of plausible distributions defined in terms of some divergence from an estimated distribution. The divergence may be relative entropy or another f-divergence or Bregman distance. We use the theory of minimising convex integral functionals under moment constraints to give formulas for the calculation of distribution model risk and to explicitly determine the worst case distribution from the set of plausible distributions. We also evaluate related risk measures describing divergence preferences.

Keywords: multiple priors, divergence preferences, relative entropy, f-divergence, Bregman distance, maximum entropy principle, convex integral functional, generalised exponential family

JEL classification: D81, C44, C60, G01, G32, M48
AMS classification: 62C20, 90B50, 91B30, 94A17

*Thomas Breuer, PPE Research Centre, FH Vorarlberg, thomas.breuer@fhv.at. Imre Csiszár, Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, csiszar.imre@mta.renyi.hu.

1 The problem of model risk

Financial risk measurement, pricing of financial instruments, and portfolio selection are all based on statistical models. If the model is wrong, risk numbers, prices, or optimal portfolios are wrong. Model risk quantifies the consequences of using the wrong models in risk measurement, pricing, or portfolio selection.

The two main elements of a statistical model in finance are a risk factor distribution and a payoff function. This work considers a one-stage set-up. Given a portfolio (or a financial instrument), the first question is: On which kind of random events does the payoff of the portfolio depend? The answer to this question determines the state space Ω. A point $r \in \Omega$ is specified by a collection of possible values of the risk factors, the uncertain events affecting the value of the given instrument or portfolio. The outcomes of these events are governed by a probability law $\mathbb{P}$ on Ω, called the risk factor distribution. The second central element is the payoff function $X : \Omega \to \mathbb{R}$ describing *how* risk factors impact the portfolio payoff at some given future time horizon. The risk factor distribution and the payoff function determine the expected payoff $E_{\mathbb{P}}(X)$.

Corresponding to the two central elements of a statistical model we distinguish two kinds of model risk, one due to uncertain knowledge of the risk factor distribution, the other to incorrect specification of the payoff function. This paper addresses risk of the first kind, called *distribution model risk.*[1]

While the risk factor distribution can be inferred by statistical modeling and a suitable estimation procedure applied to historical data, in general the so obtained "best guess" $\mathbb{P}_0$ may differ from the true data generating $\mathbb{P}$, due to model specification and estimation errors. Distribution model risk should quantify the consequences of working with $\mathbb{P}_0$ instead of the true but unknown $\mathbb{P}$. It is natural to measure this risk by

$$\mathrm{MR(X)} := -\inf_{\mathbb{P}\in\Gamma} E_{\mathbb{P}}(X) \tag{1}$$

where Γ is the set of plausible alternative risk factor distributions.[2] If desired one could normalise this risk measure by adding to (1) the constant $E_{\mathbb{P}_0}(X)$, which ensures that distribution model risk is zero if the Γ equals $\{\mathbb{P}_0\}$ and there is no distribution model uncertainty. The evaluation of the normalised measure of model risk is equivalent to that of (1).

[1]Gibson [2000] uses the term model risk for what we call distribution model risk. It encompasses both estimation risk and misspecification risk in the sense of Kerkhof et al. [2010].

[2]Cont [2006] measures model risk by $\sup_{\mathbb{P}\in\Gamma} E_{\mathbb{P}}(X) - \inf_{\mathbb{P}\in\Gamma} E_{\mathbb{P}}(X)$, where Γ is a set of arbitrage-free pricing models (i.e. risk-neutral distributions) which are consistent with market prices of liquid benchmark instruments. Our results about the solution of Problem 1 carry over directly to an evaluation of the two summands of Cont's measure of model risk, as long as Γ is of form (2) below.

So MR(X) is the negative of the worst expected payoff which could result from risk factor distributions in the set Γ. Recall that *any* risk measure satisfying the natural postulates of *coherence* can be represented by (1) for some convex set Γ of probabilities.[3]

For another interpretation of (1), note that the payoff function X may represent the utility of any act as a function of the risk factors, not just the monetary payoff of a portfolio. A widely used class of preferences allowing for ambiguity aversion are the multiple priors preferences, also known as maxmin expected utility preferences, axiomatised by Gilboa and Schmeidler [1989]. Agents with multiple priors preferences choose acts X whose worst expected utility is largest, the worst case taken over a convex set Γ of priors held by the agent; ambiguity is reflected by the multiplicity of the priors. Interpreting the choice of a portfolio as an act, the risk measure representation (1) and the multiple priors preference representation agree, see Föllmer and Schied [2002].

The axiomatic theory gives no hint how to choose the set Γ, i.e., which distributions $\mathbb{P}$ should be considered as plausible alternative risk factor distributions or priors. We propose to take balls of distributions, defined in terms of some divergence, centered at $\mathbb{P}_0$:

$$\Gamma = \{\mathbb{P} : D(\mathbb{P} \,||\, \mathbb{P}_0) \leq k\}, \tag{2}$$

where the divergence D could be relative entropy (synonyms: Kullback-Leibler distance, I-divergence), some Bregman distance, or some f-divergence, see Section 2 for definitions. This means that those risk factor distributions $\mathbb{P}$ are considered plausible whose divergence from $\mathbb{P}_0$ does not exceed some radius $k > 0$. The parameter k has to be chosen by hand, like $\alpha \in [0, 1]$ for Value at Risk or Expected Shortfall, and describes the degree of uncertainty about the risk factor distribution.

More generally, we will consider the case when Γ is given by a level set of a convex integral functional (10), containing the mentioned choices as special cases. As main mathematical result, we provide in Section 4 the solution to Problem (1) in that generality, including the characterization of the minimiser when it exists. The special cases of relative entropy, Bregman, and f-divergence balls are treated in Section 5. Finally, in Section 6 we provide an explicit solution to the related, mathematically simpler, problem

$$W = W(X) := \inf_{\mathbb{P}} \left[E_{\mathbb{P}}(X) + \lambda D(\mathbb{P} \,||\, \mathbb{P}_0)\right], \quad \lambda > 0. \tag{3}$$

Note that $-W$ is a convex risk measure (non-coherent in general), and decision makers with "divergence preferences" rank alternatives X by $W(X)$.

[3]The representation theorem is due to Artzner et al. [1999] for finite sample spaces. For general probability spaces see Delbaen [2002] or Föllmer and Schied [2002]. Its formal statement is not needed for our purposes.

In the present context, the choice of Γ by (2) with $D(\mathbb{P}||\mathbb{P}_0)$ equal to relative entropy has been proposed by Hansen and Sargent [2001] and Friedman [2002], see also Ahmadi-Javid [2011], Breuer and Csiszár [2013]. Maccheroni et al. [2006] presented a unified framework encompassing both the multiple priors preference (1) and the divergence preferences (3). They proposed to use weighted f-divergences, which are also covered in our framework. Ben-Tal and Teboulle [2007, Theorem 4.2] showed that their optimised certainty equivalent for a utility function u can be represented by (3) with D an f-divergence, the function f satisfying $u(x) = -f^*(-x)$. For both, the worst case solution is a member of the same generalised exponential family. This paper makes clear the reasons. Finally but importantly, the work of Ahmadi-Javid [2011] has to be cited for solutions of (1) and (3), in case of relative entropy and of f-divergences, in the form of convex optimization formulas involving two real variables (one in the case of relative entropy). The relationship of these results to ours will not be discussed here.

An objection against the choice of the set Γ in (2) with D equal to relative entropy or a related divergence should also be mentioned. It is that all distributions in this set are absolutely continuous with respect to μ. In the literature of the subject, even if not working with divergences, it is a rather common assumption that the set of feasible distributions is dominated; one notable exception is Cont [2006]. Sometimes the assumption of dominatedness, i.e., that each $\mathbb{P} \in \Gamma$ assigns probability 0 to the sets of scenarios $A \subset \Omega$ with $\mu(A) = 0$, is hard to justify. For example, in a multiperiod setting where Ω is the canonical space of continuous paths and Γ is a set of martingale laws for the canonical process, corresponding to different scenarios of volatilities, this Γ is typically not dominated (see Nutz and Soner [2012]). Note that for some choices of f the f-divergence balls are not dominated, see (8). But also in that case, f-divergences are unsuitable to describe approximation of continuous distributions by discrete ones, for they have a constant value if $\mathbb{P}$ and $\mathbb{P}_0$ are mutually singular.

To overcome objections to a dominated set Γ of alternative risk factor distributions, one might regard another (non-dominated) set $\tilde{\Gamma}$ as genuine, for which the risk (1) is equal to or negligibly differs from that for Γ. Then it is justified to consider Problem (1) for Γ. Apparently, such $\tilde{\Gamma}$ not depending on X (for X in a given class of payoff functions) typically exists, consisting of distributions "close" in a suitable sense to some $\mathbb{P} \in \Gamma$.

The mathematical approach of this paper is to exploit the relationship of Problem (1) to that of minimizing convex integral functionals (and specifically relative entropy) under moment constraints. The tools we need do not go beyond convex duality for $\mathbb{R}$ and $\mathbb{R}^2$, and many results directly follow from known ones about the moment problem. The assumption frequently made in the literature that X is essentially bounded, will not be needed. An explicit necessary and sufficient condition for our results to be meaningful is given in Theorem 2.

2 Divergences

We define divergences between non-negative functions on the state space Ω, which may be any set equipped with a σ-algebra not mentioned in the sequel, and with a finite or σ-finite measure μ on that σ-algebra. Then the divergence between distributions (probability measures on Ω) absolutely continuous with respect to μ is taken to be the divergence between the corresponding density functions. In our terminology, a divergence is non-negative and vanishes only for identical functions or distributions. (Functions which are equal μ-a.e. are regarded as identical.) A divergence need not be a metric, may be non-symmetric, and the divergence balls need not form a basis for a topology in the space of probability distributions.

The relative entropy of two non-negative functions p, p_0 is defined as

$$I(p||\, p_0) := \int_\Omega [p(r)\log\frac{p(r)}{p_0(r)} - p(r) + p_0(r)]d\mu(r).$$

If p, p_0 are μ-densities of probability distributions $\mathbb{P}, \mathbb{P}_0$ this reduces to the original definition of Kullback and Leibler [1951],

$$I(\mathbb{P}\,||\,\mathbb{P}_0) = \int \log\frac{d\mathbb{P}}{d\mathbb{P}_0}(r)d\mathbb{P}(r) \quad \text{if } \mathbb{P} \ll \mathbb{P}_0.$$

If a distribution $\mathbb{P}$ is not absolutely continuous with respect to $\mathbb{P}_0$, take $I(\mathbb{P}\,||\,\mathbb{P}_0) = +\infty$.[4]

Bregman distances, introduced by Bregman [1967], and f-divergences, introduced by Csiszár [1963] and Ali and Silvey [1966], are classes of divergences parametrised by convex functions $f : (0, \infty) \to \mathbb{R}$, extended to $[0, \infty)$ by setting $f(0) := \lim_{t\to 0} f(t)$. Below, f is assumed strictly convex but not necessarily differentiable. The Bregman distance of non-negative (measurable) functions p, p_0 on Ω, with respect to μ, is defined by

$$B_{f,\mu}(p, p_0) := \int_\Omega \Delta_f(p(r), p_0(r))\mu(dr), \tag{4}$$

where, for s, t in $[0, +\infty)$,

$$\Delta_f(s,t) := \begin{cases} f(s) - f(t) - f'(t)(s-t) & \text{if } t > 0 \text{ or } t = 0,\ f(0) < +\infty \\ s \cdot (+\infty) & \text{if } t = 0 \text{ and } f(0) = +\infty. \end{cases} \tag{5}$$

If the convex function f is not differentiable at t, the right or left derivative is taken for $f'(t)$ according as $s > t$ or $s < t$.

The Bregman distance of distributions $\mathbb{P} \ll \mu, \mathbb{P}_0 \ll \mu$ is defined by

$$B_{f,\mu}(\mathbb{P}, \mathbb{P}_0) := B_{f,\mu}\left(\frac{d\mathbb{P}}{d\mu}, \frac{d\mathbb{P}_0}{d\mu}\right). \tag{6}$$

[4]Note that $I(\mathbb{P}\,||\,\mathbb{P}_0)$ is a less frequent notation for relative entropy than $D(\mathbb{P}\,||\,\mathbb{P}_0)$, it has been chosen here because we use the latter to denote any divergence.

Clearly, $B_{f,\mu}$ is a bona fide divergence whenever f is strictly convex in $(0, +\infty)$. For $f(s) = s \log s - s + 1$, B_f is the relative entropy I. For $f(s) = -\log s$, B_f is the Itakura-Saito distance. For $f(s) = s^2$, B_f is the squared L^2-distance.

The f-divergence between non-negative (measurable) functions p and p_0 is defined, when f additionally satisfies $f(s) \geq f(1) = 0$,[5] by

$$D_f(p||p_0) := \int_\Omega f\left(\frac{p(r)}{p_0(r)}\right) p_0(r)\, \mu(dr). \tag{7}$$

At places where $p_0(r) = 0$, the integrand by convention is taken to be $p(r) \lim_{s\to\infty} f(s)/s$. The f-divergence of distributions $\mathbb{P} \ll \mu, \mathbb{P}_0 \ll \mu$, defined as the f-divergence of the corresponding densities, does not depend on μ and is equal to

$$D_f(\mathbb{P}||\mathbb{P}_0) := \int_\Omega f\left(\frac{d\mathbb{P}_a}{d\mathbb{P}_0}\right) d\mathbb{P}_0 + \mathbb{P}_s(\Omega) \lim_{s\to\infty} \frac{f(s)}{s}, \tag{8}$$

where $\mathbb{P}_a$ and $\mathbb{P}_s$ are the absolutely continuous and singular components of $\mathbb{P}$ with respect to $\mathbb{P}_0$. Note that if f is cofinite, i.e. if the limit in (8) is $+\infty$, then $\mathbb{P} \ll \mathbb{P}_0$ is a necessary condition for the finiteness of $D_f(\mathbb{P}||\mathbb{P}_0)$, while otherwise not.

For $f(s) = s \log s - s + 1$, D_f is the relative entropy I. For $f(s) = -\log s + s + 1$, D_f is the reversed relative entropy. For $f(s) = (\sqrt{s} - 1)^2$, D_f is the squared Hellinger distance. For $f(s) = (s-1)^2/2$, D_f is the relative Gini concentration index. For more details about f-divergences see Liese and Vajda [1987].

Remark 1. Variation distance is also an f-divergence but it corresponds to $f(s) = |s - 1|$ that does not meet the strict convexity assumption. The familiar risk measure Expected Shortfall can be represented by (1) taking $\Gamma = \{\mathbb{P}\colon D_f(\mathbb{P}||\mathbb{P}_0) \leq k\}$ (for any $k \geq 0$), with D_f as in (8) for the pathological convex function f that equals 0 in the interval $[0, 1/\alpha]$ and $+\infty$ otherwise, see [Föllmer and Schied, 2004, Theorem 4.47]; this D_f takes only values 0 and $+\infty$ and is not a divergence in our sense.

Relative entropy appears the most versatile divergence measure for probability distributions or non-negative functions, extensively used in diverse fields including statistics, information theory, statistical physics, etc. To our knowledge, in the context of this paper first Hansen and Sargent [2001] have used expected value minimization over relative entropy balls. Arguments for (2) with any f-divergence in the role of D, or more generally

[5]This makes sure that (7) indeed defines a divergence between any non-negative functions; if attention is restricted to probability densities resp. probability distributions, it suffices to assume that $f(1) = 0$.

with a weighted f-divergence involving a (positive) weight function $w(r)$ in the integral in (7), have been put forward by Maccheroni et al. [2006]. Results of Ahmadi-Javid [2011] indicate advantages of relative entropy over other f-divergences also in this context. In another context, Grunwald and Dawid [2004] argue that distances between distributions might be chosen in a utility dependent way, and relative entropy is natural only for decision makers with logarithmic utility. Picking up this idea, decision makers with non-logarithmic utility might define the ball (2) in terms of some utility dependent Bregman distance. We are, however, unaware of references employing Bregman distances at all in our context.

In the context of inference, the method of maximum entropy (or relative entropy minimization) is distinguished by axiomatic considerations. Shore and Johnson [1980], Paris and Vencovská [1990], and Csiszár [1991] showed that it is the only method that satisfies certain intuitively desirable postulates. Still, relative entropy cannot be singled out as providing the only reasonable method of inference. Csiszár [1991] determined what alternatives (specifically, Bregman distances and f-divergences) come into account if some postulates are relaxed. In the context of measuring risk or evaluating preferences under ambiguity aversion, axiomatic results distinguishing relative entropy or some other divergence are not available.

3 General framework

Now we construct a unified framework that covers the choices of Γ in (2) when D is an f-divergence or a Bregman distance, as well as others. In this framework, Γ is chosen as a set of probability measures $P \ll \mu$ of the form

$$\Gamma = \{\mathbb{P} \ll \mu : \; p = d\mathbb{P}/d\mu \;\text{ satisfies }\; H(p) \leq k\}, \tag{9}$$

where H is a convex integral functional defined as

$$H(p) = H_\beta(p) := \int_\Omega \beta(r, p(r))\mu(dr), \tag{10}$$

for measurable, non-negative functions p on Ω. Here $\beta : \Omega \times (0, +\infty) \to \mathbb{R}$ is a mapping such that $\beta(r, s)$ is a measurable function of r for each $s \in (0, +\infty)$ and a strictly convex function of s for each $r \in \Omega$. The definition of β is extended to $s \leq 0$ by

$$\beta(r, 0) := \lim_{s \downarrow 0} \beta(r, s), \quad \beta(r, s) := +\infty \text{ if } s < 0. \tag{11}$$

No differentiability assumptions are made about β but the convenient notations $\beta'(r, 0)$ and $\beta'(r, +\infty)$ will be used for the common limits of the left and right derivatives of $\beta(r, s)$ by s as $s \downarrow 0$ resp. $s \uparrow +\infty$. Note that

$$\beta'(r, +\infty) = \lim_{s \uparrow +\infty} \frac{\beta(r, s)}{s}. \tag{12}$$

With the understandings (11), the mapping $\beta : \Omega \times \mathbb{R} \to (-\infty, +\infty]$ is a convex normal integrand in the sense of Rockafellar and Wets [1997], which ensures the measurability[6] of the function $\beta(r, p(r))$ in (10) and of similar functions later on, as in (23) and (27).

Depending on the choice of β, $H_\beta(p)$ will be relative entropy to $\mathbb{P}_0$, some Bregman distance, some f-divergence, or some other divergence, as in Section 5 below. One motivation for admitting non-autonomous integrands ($\beta(r, s)$ actually depending on r) has been to cover Bregman distances. A general assumption about the relation of β and the best guess distribution $\mathbb{P}_0$, always satisfied in the above cases, will be that the minimum of $H(p)$ among probability densities p is attained for p_0, the density of $\mathbb{P}_0$; without any loss of generality, this minimum is supposed to be 0, thus

$$H(p) \geq H(p_0) = 0 \text{ whenever } \int p d\mu = 1. \tag{13}$$

Another standing assumption will be that $E_{\mathbb{P}_0}(X) = \int X p_0 d\mu$ exists and

$$m := \mu\text{-ess}\inf(X) \ < b_0 := E_{\mathbb{P}_0}(X) < \ M := \mu\text{-ess}\sup(X). \tag{14}$$

The distribution model risk (1) with Γ as in (9) is evaluated by solving the worst case problem

$$\inf_{p: \int p d\mu = 1, H(p) \leq k} \int X p d\mu =: V(k) \tag{15}$$

and then taking $\mathrm{MR(X)} = -V(k)$. Our goal is to determine $V(k)$, and also the minimiser (the density of the worst case distribution in Γ), if $V(k)$ is finite and the minimum in (15) is attained. If this minimiser exists, it is unique, by strict convexity of β.

Our approach to Problem (15) will be based on its relationship to the *moment problem*

$$\inf_{p: \int p d\mu = 1, \int X p d\mu = b} H(p) =: F(b), \tag{16}$$

described by the next Proposition.

Proposition 1. *Supposing*

$$0 < k < k_{\max} := \lim_{b \downarrow m} F(b), \tag{17}$$

there exists a unique b with

$$F(b) = k, \quad m < b < b_0, \tag{18}$$

[6] Measurability issues will not be entered below. For the measurability of functions we deal with, see references in Csiszár and Matúš [2012] to the book of Rockafellar and Wets [1997].

and then

$$V(k) = b. \tag{19}$$

The minimum in (15) *is attained if and only if that in* (16) *is attained (for this* b*), in which case the same* p *attains both minima.*

Proof. Due to (13) the convex function F attains its minimum 0 at b_0, hence the assumption (17) trivially implies the existence of a unique b satisfying (18). Moreover, then each $t \in (b, b_0)$ satisfies $F(t) < k$, thus there exist functions p with $\int p d\mu = 1$, $\int X p d\mu = t$ such that $F(t) > k$. This proves that $V(k) \leq b$. On the other hand, $F(t) > k$ if $t \in (m, b)$ (hence also $F(m) > k$ if m is finite), which means that the conditions $\int p d\mu = 1$ and $\int X p d\mu = t$ imply $H(p) \geq F(t) > k$ for each $t \in (-\infty, b)$. Since $\int X p d\mu > -\infty$ if $H(p) < \infty$, as verified later (Corollary 3 of Theorem 2), this proves that $V(k) \geq t$. The last assertion of the Proposition follows obviously. □

Remark 2. The condition (17) in Proposition 1 covers all interesting values of k. Indeed, one easily sees that if $k > k_{\max}$ or $k \geq k_{\max} > 0$ then $V(k) = m$, while clearly $V(0) = b_0$. This also means that the functional H can be suitable for assigning model risk only if $k_{\max} > 0$. A necessary and sufficient condition for $k_{\max} > 0$ will be given in Theorem 2. Note that if $m = -\infty$ then $k_{\max} > 0$ implies $k_{\max} = \infty$, in which case each $k > 0$ meets condition (17).

For technical reasons, it will be convenient to regard $F(b)$ as the instance $a = 1$ of the function

$$J(a, b) := \inf_{p: \int p d\mu = a, \int X p d\mu = b} H(p), \quad (a, b) \in \mathbb{R}^2. \tag{20}$$

Problem (20) is a special case of minimising convex integral functionals under moment constraints, which has an extensive literature. For references, see the recent work of Csiszár and Matúš [2012], relied upon here also for results that date back much earlier, perhaps under less general conditions. The results in Csiszár and Matúš [2012] will be used (without further mentioning this) with the choice $\phi : r \to (1, X(r))$ of the moment mapping when the "value function" there reduces to the function J here.

Remark 3. Due to (13), the effective domain

$$\operatorname{dom} J := \{(a, b) : J(a, b) < +\infty\} \tag{21}$$

of J is nonempty, and by Csiszár and Matúš [2012, Lemma 6.6] its interior is

$$\operatorname{int}\operatorname{dom} J = \{(a, b) : a > 0, am < b < aM\}. \tag{22}$$

In particular, the convex function J is proper, i.e., it never equals $-\infty$ and is not identically $+\infty$.

Many results in Csiszár and Matúš [2012] need a condition called dual constraint qualification which, however, always holds in the current setting, namely, the set Θ defined in (26) is non-empty (see the passage following (26)).

Denote

$$K(\theta_1, \theta_2) := \int \beta^*(r, \theta_1 + \theta_2 X(r))\mu(dr), \quad (\theta_1, \theta_2) \in \mathbb{R}^2, \tag{23}$$

where β^* is the convex conjugate of β with respect to the second variable,

$$\beta^*(r, x) := \sup_{s\in\mathbb{R}} \left(xs - \beta(r, s)\right), \quad x \in \mathbb{R}. \tag{24}$$

The function K is equal to the convex conjugate of J in (20):

$$K(\theta_1, \theta_2) = J^*(\theta_1, \theta_2) := \sup_{(a,b)\in\mathbb{R}^2} \left(\theta_1 a + \theta_2 b - J(a, b)\right), \tag{25}$$

see [Csiszár and Matúš, 2012, Theorem 1.1]. In particular, K is a lower semicontinuous proper convex function.

The properties of β imply that $\beta^*(r, x)$ is a convex function of x which is finite, non-decreasing, and differentiable in the interval $(-\infty, \beta'(r, +\infty))$, see (12). At $x = \beta'(r, +\infty)$, if finite, $\beta^*(r, x)$ may be finite or $+\infty$. The derivative $(\beta^*)'(r, x)$ equals zero for $x \leq \beta'(r, 0)$, is positive for $\beta'(r, 0) < x < \beta'(r, +\infty)$, and grows to $+\infty$ as $x \uparrow \beta'(r, +\infty)$.

Let Θ be the following subset of $\operatorname{dom} K := \{(\theta_1, \theta_2) : K(\theta_1, \theta_2) < +\infty\}$:

$$\Theta := \left\{\theta : K(\theta_1, \theta_2) < +\infty, \;\; \theta_1 + \theta_2 X(r) < \beta'(r, +\infty)\; \mu\text{-a.e.}\right\}. \tag{26}$$

This set is nonempty, for if $(\theta_1, \theta_2) \in \operatorname{dom} K$ and $\overline{\theta_1} < \theta_1$ then $(\overline{\theta_1}, \theta_2) \in \Theta$. Also, Θ contains the interior of $\operatorname{dom} K$. If $\beta'(r, +\infty) = +\infty$ μ-a.e. then $\Theta = \operatorname{dom} K$.

The following functions on Ω will play a role like those in an exponential family do for relative entropy balls, see Breuer and Csiszár [2013] or Section 5, though they need not integrate to 1:

$$p_\theta(r) := (\beta^*)'(r, \theta_1 + \theta_2 X(r)), \quad \theta = (\theta_1, \theta_2) \in \Theta. \tag{27}$$

The definition (26) makes sure that in case $(\theta_1, \theta_2) \in \Theta$ the derivative in (27) exists for μ-a.e. $r \in \Omega$. For all other r, if any, one may set $p_{\theta_1,\theta_2}(r) = 0$ by definition. The family $\{p_\theta \colon \theta \in \Theta\}$ will be referred to as *generalised exponential family*. As verified later, see Remark 3, it always contains the default density p_0, equal to $p_{(\theta_0,0)}$ for some θ_0 with $(\theta_0, 0) \in \Theta$. Another useful fact is that K is differentiable in the interior of $\operatorname{dom} K$, and

$$\nabla K(\theta) = \left(\int p_\theta d\mu, \int X p_\theta d\mu\right), \quad \theta = (\theta_1, \theta_2) \in \operatorname{int}\operatorname{dom} K, \tag{28}$$

see [Csiszár and Matúš, 2012, Corollary 3.8].

The identity $J^* = K$ implies [Rockafellar, 1970, Theorem 12.2] that

$$J(a,b) = K^*(a,b), \quad (a,b) \in \text{int dom } J, \tag{29}$$

where K^* is the convex conjugate of K,

$$K^*(a,b) := \sup_{(\theta_1,\theta_2)\in\mathbb{R}^2} \left(\theta_1 a + \theta_2 b - K(\theta_1,\theta_2)\right), \quad (a,b) \in \mathbb{R}^2. \tag{30}$$

Using (22), for $F(b) = J(1,b)$ it follows, supplementing Proposition 1:

Lemma 1. *$F(b) = K^*(1,b)$ for each $b \in (m,M)$.*

4 Main results

Proposition 1 and Lemma 1 already provide a recipe for computing $V(k)$. In regular cases, a more explicit solution is available, based on the following key result about Problem (20), see Csiszár and Matúš [2012, Lemma 4.4, Lemma 4.10]:

Lemma 2. *If $\theta = (\theta_1,\theta_2) \in \Theta$ satisfies*

$$\int p_\theta \, d\mu = a, \quad \int X p_\theta \, d\mu = b \tag{31}$$

then it attains the maximum in (30). Moreover, in case $(a,b) \in$ int dom J, the existence of $\theta \in \Theta$ satisfying (31) is necessary and sufficient for the attainment of the minimum in (20), and then $p = p_\theta$ is the (unique) minimiser.

Theorem 1. *Assuming (13), (14), (17), if*

$$\overline{\theta}_2 < 0, \quad \int p_{\overline{\theta}} \, d\mu = 1, \quad \overline{\theta}_1 + \overline{\theta}_2 \int X p_{\overline{\theta}} \, d\mu - K(\overline{\theta}) = k \tag{32}$$

for some $\overline{\theta} = (\overline{\theta}_1, \overline{\theta}_2) \in \Theta$ then the value of the inf *in (15) is*

$$V(k) = \int X p_{\overline{\theta}} \, d\mu. \tag{33}$$

Essential smoothness[7] *of K is a sufficient condition for the existence of such $\overline{\theta}$. Further, a necessary and sufficient condition for p to attain the minimum in (15) is $p = p_{\overline{\theta}}$ for a $\overline{\theta} \in \Theta$ satisfying (32).*

[7] A lower semicontinuous proper convex function is essentially smooth if its effective domain has nonempty interior, the function is differentiable there, and at non-interior points of the effective domain the directional derivatives in directions towards the interior are $-\infty$. The latter trivially holds if the effective domain is open.

Corollary 1. *If the equations*

$$\frac{\partial}{\partial \theta_1} K(\theta) = 1, \quad \theta_1 + \theta_2 \frac{\partial}{\partial \theta_2} K(\theta) - K(\theta) = k \tag{34}$$

have a solution $\overline{\theta} = (\overline{\theta}_1, \overline{\theta}_2) \in \mathit{int\,dom}\, K$ *with* $\overline{\theta}_2 < 0$ *then* $\overline{\theta}$ *satisfies* (32) *and the solution to Problem* (15) *equals*

$$V(k) = \left. \frac{\partial K(\theta)}{\partial \theta_2} \right|_{\theta = \overline{\theta}}. \tag{35}$$

The Corollary follows from the Theorem because, for $\overline{\theta} \in \mathrm{int\,dom}\, K$, the equations in (32) are equivalent to those in (34), by (28). However, if K is not essentially smooth, $\overline{\theta} \in \mathrm{int\ dom}\, K$ is not a necessary condition for (32).

Proof. By Lemma 2, if $\theta = (\theta_1, \theta_2) \in \Theta$ satisfies (31) then it attains the maximum in (30), thus $K^*(a, b) = \theta_1 a + \theta_2 b - K(\theta_1, \theta_2)$. Hence by Lemma 1, if(32) holds and $b := \int X p_{\overline{\theta}}\, d\mu$ satisfies $m < b < M$ then

$$F(b) = K^*(1, b) = \overline{\theta}_1 + \overline{\theta}_2 b - K(\overline{\theta}) = k. \tag{36}$$

Due to Proposition 1, to prove (33) it remains to show that $m < b < b_0$. Clearly, $k < k_{\max}$ implies $m < b$. Further, (30) and (36) imply

$$F(t) = K^*(1, t) \geq \overline{\theta}_1 + \overline{\theta}_2 t - K(\overline{\theta}_1, \overline{\theta}_2) = F(b) + \overline{\theta}_2 (t - b), \quad t \in (m, M). \tag{37}$$

Since $\overline{\theta}_2 < 0$, this shows that $F(t) > F(b)$ if $t \in (m, b_0)$, completing the proof of (33).

Suppose next that K is essentially smooth. Then to b in (18) there exists $\overline{\theta} \in \mathrm{int\ dom}\ K$ with

$$(1, b) = \nabla K(\overline{\theta}), \tag{38}$$

because $(1, b) \in \mathrm{int\ dom}\ J$ and the gradient vectors of the essentially smooth K cover int dom $K^* =$ int dom J, see Rockafellar [1970, Corollary 26.4.1]. Clearly, (38) implies that $\overline{\theta}$ attains the maximum in (30) (for $a = 1$), hence it satisfies (36). This means by (38) that $\overline{\theta}$ satisfies the equations in (34), equivalent to those in (32). It remains to show that $\overline{\theta}_2 < 0$, but this follows from (37) applied to $t = b_0$.

Finally, the last assertion of Theorem 1 follows from Proposition 1 and Lemma 2. □

In Proposition 1 and Theorem 1 the condition $k_{\max} > 0$ has been assumed. Next we give a necessary and sufficient condition for this to hold.

Remark 4. A simpler instance of [Csiszár and Matúš, 2012, Lemma 4.10] than Lemma 2, obtained by taking the constant mapping $r \to 1$ for the moment mapping ϕ, gives the following: the necessary and sufficient condition for p to minimise $H(p)$ subject to $\int p\,d\mu = a$ ($a > 0$) is that $p(r) = (\beta^*)'(r, \theta)$ for some $\theta \in \mathbb{R}$ with $\beta^*(r, \theta)$ μ-integrable, and then the minimum is equal to $a\theta - \int \beta^*(r, \theta)\,d\mu(r)$. This establishes the claim that the default density p_0, minimising $H(p)$ subject to $\int p\,d\mu = 1$, equals $p_{(\theta_0,0)}$ for some θ_0 with $(\theta_0, 0) \in \Theta$; this θ_0 also satisfies $\theta_0 - \int \beta^*(r, \theta_0)\,d\mu(r) = H(p_0) = 0$.

Theorem 2. *Assuming* (13), (14), *for* $b < b_0$ *we have* $F(b) > 0$ *if and only if*

$$\text{there exists } \theta = (\theta_1, \theta_2) \in \operatorname{dom} K \text{ with } \theta_2 < 0. \tag{39}$$

In particular, condition (39) *is necessary and sufficient for* $k_{\max} > 0$.

Proof. To prove the necessity of (39), we may assume $m < b < b_0$. Then $(1, b) \in \operatorname{int} \operatorname{dom} J$, see (21), hence the convex function J has nonempty subgradient at $(1, b)$ [Rockafellar, 1970, Theorem 23.4]. As $J^* = K$, if $\overline{\theta} = (\overline{\theta}_1, \overline{\theta}_2)$ belongs to that subgradient then

$$F(b) = J(1, b) = \overline{\theta}_1 + \overline{\theta}_2 b - K(\overline{\theta}) \tag{40}$$

by [Rockafellar, 1970, Theorem 23.5], which implies as in the proof of Theorem 1 that this $\overline{\theta}$ also satisfies (37). In turn, (37) with $t = b_0$ implies that $\overline{\theta}_2 \le 0$, with the strict inequality if $F(b) > 0$. This proves the necessity of (39).

For sufficiency, suppose that $F(b) = 0$ for some $b \ne b_0$, $m < b < M$. By Remark 4, then $F(b) = 0 = \theta_0 - K(\theta_0, 0)$, hence $\theta^* := (\theta_0, 0)$ is a maximiser of $g(\theta) := \theta_1 + \theta_2 b - K(\theta)$, see (29), (30). It follows that for no $\bar{\theta} \in \operatorname{dom} K$ can the directional derivative $g'(\theta^*; \bar{\theta} - \theta^*)$ be positive. By [Csiszár and Matúš, 2012, Lemma 3.6, Remark 3.7], this directional derivative is equal to

$$(\bar{\theta}_1 - \theta_0) + \bar{\theta}_2 b - \int (\bar{\theta}_1 - \theta_0 + \bar{\theta}_2 X) p_{\theta^*}\,d\mu = \bar{\theta}_2(b - b_0).$$

Thus, the existence of $\bar{\theta} \in \operatorname{dom} K$ with $\bar{\theta}_2 < 0$ rules out $b < b_0$, proving the sufficiency part of the Theorem. □

Corollary 2. *Sufficient conditions for* $k_{\max} > 0$ *are* $m > -\infty$ *or the essential smoothness of* K. *If* β *is an autonomous integrand and* $\beta(r, s) = f(s)$ *is not cofinite then* $m > -\infty$ *is also necessary for* $k_{\max} > 0$.

Proof. If m is finite then each $\theta_2 < 0$ satisfies condition (39) with some θ_1. Indeed, since $\theta_1 + \theta_2 X \le \theta_1 + \theta_2 m$ μ-a.e., if the right hand side is less than θ_0 in Remark 4 then $(\theta_1, \theta_2) \in \operatorname{dom} K$. If K is essentially smooth then condition (39) holds because int dom K contains $\theta^* = (\theta_0, 0)$. Indeed, otherwise the directional derivatives of K at θ^* in directions towards interior points were equal to $-\infty$, and θ^* could not maximize $\theta_1 + \theta_2 b_0 - K(\theta)$.

Finally, suppose $\beta(r, s) = f(s)$ with $\lim_{s\to+\infty} \frac{f(s)}{s}$ finite. Then $f^*(x)$ is infinite for x above this limit, hence $\int f^*(\theta_1 + \theta_2 X)d\mu = +\infty$ for all $\theta_1 \in \mathbb{R}$ and $\theta_2 < 0$ if $m = -\infty$. $\square$

Corollary 3. *If* $k_{\max} > 0$ *then* $\int p d\mu = 1, H(p) < +\infty$ *imply* $\int X p d\mu > -\infty$.

Proof. Substitute in the Fenchel inequality $xs \leq \beta(r, s) + \beta^*(r, x)$ (a consequence of (24)) $x := \theta_1 + \theta_2 X(r)$, $s := p(r)$ and integrate. It follows that if $(\theta_1, \theta_2) \in \operatorname{dom} K$ and p satisfies the hypotheses then

$$\theta_1 + \theta_2 \int X p d\mu \leq H(p) + K(\theta_1, \theta_2) < +\infty.$$

Taking (θ_1, θ_2) as in (39), the assertion follows. $\square$

5 Application to relative entropy, f-divergence, and Bregman balls

We now come back to the specific choices (2) where Γ is a ball of distributions in terms of some divergence D, centered at $\mathbb{P}_0$.

Relative entropy balls For this case we briefly check how the unified framework leads to an earlier result Breuer and Csiszár [2013, Theorem 1], and then show that it provides the solution also when the hypothesis there fails.

To let the set of distributions (9) be the ball (2) with D equal to relative entropy, set $\mu = \mathbb{P}_0$ and $\beta(r, s) := f(s) := s \log s - s + 1$. Then $\beta^*(r, x) = f^*(x) = \exp(x) - 1$. Define

$$\Lambda(\tau) := \log \int \exp(\tau X) d\mathbb{P}_0, \quad \tau \in \mathbb{R}. \tag{41}$$

The function (23) and the set (26) will be

$$K(\theta_1, \theta_2) = \int (\exp(\theta_1 + \theta_2 X) - 1)\, d\mathbb{P}_0 = \exp(\theta_1 + \Lambda(\theta_2)) - 1,$$

and $\Theta = \operatorname{dom} K = \mathbb{R} \times \operatorname{dom} \Lambda$. The necessary and sufficient condition for $k_{\max} > 0$ in Theorem 2 becomes that $\operatorname{dom} \Lambda$ contains some $\tau < 0$, i.e.,

$$\tau_{\min} := \inf\{\tau : \Lambda(\tau) < +\infty\} \ < \ 0. \tag{42}$$

The functions p_θ, $\theta \in \Theta$ of (27) are of form $\exp(\theta_1 + \theta_2 X(r))$, which means a positive constant times a member of the exponential family of densities

$$\exp(\tau X - \Lambda(\tau)), \quad \tau \in \operatorname{dom} \Lambda. \tag{43}$$

The first equation in (32) requires $p_{\overline{\theta}}$ to be a density, thus member of the above exponential family, equivalently $\overline{\theta}_1 = -\Lambda(\overline{\theta}_2)$ for $\overline{\theta} = (\overline{\theta}_1, \overline{\theta}_2)$. Then $\int X p_{\overline{\theta}} \, d\mathbb{P}_0 = \Lambda'(\overline{\theta}_2)$, and the second equation in (32) reads $-\Lambda(\overline{\theta}_2) + \overline{\theta}_2 \Lambda'(\overline{\theta}_2) = k$. Thus Theorem 1 gives that if the equation

$$-\Lambda(\tau) + \tau\Lambda'(\tau) = k \tag{44}$$

has a negative solution $\bar{\tau} < 0$ then $V(k) = \Lambda'(\bar{\tau}))$, a worst case distribution exists, and its density is $\exp(\bar{\tau}X - \Lambda(\bar{\tau}))$, recovering Breuer and Csiszár [2013, Theorem 1].

As the left hand side of (44) is equal to $\Lambda^*(\Lambda'(\tau))$, the above result follows also directly from Proposition 1. Indeed, Lemma 1 gives

$$\begin{aligned} F(b) = K^*(1, b) &= \sup_{\theta_1, \theta_2} [\theta_1 + \theta_2 b - \exp(\theta_1 + \Lambda(\theta_2)) + 1] \\ &= \sup_{\theta_2} [-\Lambda(\theta_2) + \theta_2 b] = \Lambda^*(b), \quad b \in (m, M), \end{aligned}$$

as the supremum for θ_1 is attained for $\theta_1 = -\Lambda(\theta_2)$. Moreover, assuming (42), the solution b of equation (18) always equals $\Lambda'(\tau)$ for some $\tau < 0$ if $0 < k < k_{\max} = \lim_{b \downarrow m} \Lambda^*(b)$, provided that $\tau_{\min}$ equals $-\infty$ or it is finite but

$$k \leq -\Lambda(\tau_{\min}) + \tau_{\min}\Lambda'(\tau_{min}). \tag{45}$$

On the other hand, if $\tau_{\min}$ and $\Lambda'(\tau_{\min})$ are finite and (45) does noy hold then (44) has no solution $\tau < 0$. Still, equation (18) does have a solution also in that case. As $b < \Lambda(\Lambda'(\tau_{\min}))$ implies $\Lambda^*(b) = -\Lambda(\tau_{\min}) + \tau_{\min} b$, it follows that if (45) does not hold (when the hypothesis of [Breuer and Csiszár, 2013, Theorem 1] fails), we have $V(k) = (k + \Lambda(\tau_{\min}))/\tau_{\min}$. In this case the infimum in (15) is not a minimum, thus a worst case distribution does not exist.

f-divergence balls Setting $\mu = \mathbb{P}_0$ again, take now any autonomous integrand for β given by a convex function f as in Section 2. Then the set Γ of distributions given by (9) equals the f-divergence ball $\{\mathbb{P} : D_f(\mathbb{P}||\mathbb{P}_0) \leq k\}$ if f is cofinite, while if $f'(+\infty) := \lim_{s\to\infty} f(s)/s$ is finite, Γ is a proper subset of that ball. We will focus on Γ defined by (9) anyway. Note that if f is not cofinite then this choice of Γ is adequate for assigning model risk only if X is essentially bounded below, see Corollary 2 of Theorem 2. In that case, (θ_1, θ_2) with $\theta_2 < 0$ belongs to int dom K if and only if $\theta_1 + \theta_2 m < f'(+\infty)$.

The most poular f-divergences are the *power divergences*, defined by

$$f_\alpha(s) := [s^\alpha - \alpha(s-1) - 1]/[\alpha(\alpha-1)], \quad \alpha \in \mathbb{R}.$$

Formally, f_α is undefined if $\alpha = 0$ or $\alpha = 1$, but the definition is commonly extended by limiting, thus

$$f_0(s) := \log s + s - 1, \qquad f_1(s) := s \log s - s + 1.$$

This means that also $D_{f_0}(\mathbb{P}||\mathbb{P}_0) = I(\mathbb{P}_0||\mathbb{P})$ and $D_{f_1}(\mathbb{P}||\mathbb{P}_0) = I(\mathbb{P}||\mathbb{P}_0)$ are regarded as power divergences. Note that the function f_α is cofinite if and only if $\alpha \geq 1$, and if $\alpha < 1$ then $f'_\alpha(+\infty) = 1/(1-\alpha)$.

A trite calculation gives that if $\alpha > 1$ then

$$f^*_\alpha(x) = \frac{|x(\alpha-1)+1|_+^{\alpha/(\alpha-1)} - 1}{\alpha}, \tag{46}$$

where $|.|_+$ and $|.|_-$ denote positive resp. negative part: $|a|_+ := \max\{0, a\}$, $|a|_- := \max\{0, -a\}$. If $\alpha < 1$, $\alpha \neq 0$ then (46) holds for $x \leq 1/(1-\alpha)$, and $f^*_\alpha(x) = +\infty$ otherwise.

Substituting this f^*_α in (23) gives the function $K(\theta_1, \theta_2)$ in the current case, but no simpified expression for this K or its effective domain like in the relative entropy case is available. Nevertheless, it is clear that in case $\alpha > 1$ the necessary and sufficient condition for $k_{\max} > 0$ in Theorem 2 holds if and only if $E_{\mathbb{P}_0}(|X|_-^{\alpha/(\alpha-1)}) < +\infty$, while in case $\alpha < 1$ the necessary and sufficient condition is the essential boundedness of X from below, by Corollary 2. Note that the set Θ in (26) is equal to $\operatorname{dom} K$ whenever $\alpha > 0$ but may be a proper subset of $\operatorname{dom} K$ if $\alpha < 0$.

The generalised exponential famiily (27) is now given by

$$p_\theta(r) = |(\theta_1 + \theta_2 X(r))(\alpha - 1) + 1|_+^{1/(\alpha-1)}, \quad \theta = (\theta_1, \theta_2) \in \Theta, \tag{47}$$

where in case $\alpha < 1$ the positive part can be omitted. This follows from (46) because if $\alpha < 1$ then $\theta \in \Theta$ implies by definition that $\theta_1 + \theta_2 X(r) < f'(+\infty) = 1/(1-\alpha)$, μ-a.e.. We note that (47) applies also in the case $\alpha = 0$ that has been excluded in (46).

Unlike for the relative entropy case, no explicit condition is available for $\int p_\theta d\mathbb{P}_0 = 1$, and the two equations in Theorem 1 cannot be reduced to one.

Bregman balls To obtain for the functional H in (10) the Bregman distance $B_{f,\mu}$ of (4), we choose the non-autonomous integrand

$$\beta(r, s) = f(s) - f(p_0(r)) - f'(p_0(r))(s - p_0(r)).$$

For simplicity, f is assumed differentiable. To make sure that the assumptions on β are met, in case $f'(0) = -\infty$ we assume that the default density p_0 is strictly positive; this assumption is not needed if $f'(0) > -\infty$. Note that the case $\mathbb{P}_0 = \mu$ is uninteresting for Bregman distances, since if p_0 identically equals 1, the above integrand reduces to $f(s)$ and Bregman distance reduces to f-divergence (assuming $f(1) = f'(1) = 0$ as usual).

By Csiszár and Matúš [2012, Lemma 2.6],

$$\beta^*(r, x) = f^*(x + f'(p_0(r))) - f^*(f'(p_0(r))).$$

The function K from (23) equals

$$K(\theta) := \int_\Omega \left[f^*(\theta_1 + \theta_2 X(r) + f'(p_0(r))) - f^*(f'(p_0(r))\right] d\mu(r).$$

The generalised exponential family $\{p_\theta(r) : \theta \in \Theta\}$ is formed by the functions

$$p_\theta(r) = \beta^{*\prime}(\theta_1 + \theta_2 X(r)) = f^{*\prime}[\theta_1 + \theta_2 X(r) + f'(p_0(r))].$$

Note that while the case of Bregman balls is covered by our general results, it is not apparent that the current special form of β would substantially simplify their application.

6 Evaluation of divergence preferences

Finally, we briefly address divergence preferences, i.e., the problem (3) which is simpler than problem (1) with a divergence ball (2). Divergence preferences include as special case the multiplier preferences of Hansen and Sargent [2001], when we choose the relative entropy I for D. Maccheroni et al. [2006] suggest for D weighted f-divergences

$$D_f^w(\mathbb{P}, \mathbb{P}_0) := \begin{cases} \int_\Omega w(r) f\left(\frac{d\mathbb{P}}{d\mathbb{P}_0}(r)\right) d\mathbb{P}_0(r) & \text{if } \mathbb{P} \ll \mathbb{P}_0, \\ +\infty & \text{otherwise,} \end{cases} \tag{48}$$

where w is a normalised, non-negative weight function.

Below, more generally, the role of D is given to any convex functional as in (10). Introducing a new convex integrand and integral functional by

$$\tilde{\beta}(r, s) := X(r)s + \lambda\beta(r, s), \quad \tilde{H}(p) := \int \tilde{\beta}(r, p(r)) d\mu(r),$$

(where $\lambda > 0$ is fixed), we can write

$$W = \inf_{p:\int p d\mu = 1} \left[\int X p d\mu + \lambda H(p)\right] = \inf_{p:\int p d\mu = 1} \tilde{H}(p). \tag{49}$$

Thus, the problem is to minimize the functional $\tilde{H}(p)$ under the single constraint $\int p d\mu = 1$.

In analogy to (20), consider

$$\tilde{J}(a) := \inf_{p:\int p d\mu = a} \tilde{H}(p), \quad a \in \mathbb{R}.$$

Note that $\tilde{\beta}$ meets the basic assumptions on β (though (13) does not hold for $\tilde{H}$), and that

$$(\tilde{\beta})^*(r, x) = \sup_s [xs - X(r)s - \lambda\beta(r, s)] = \lambda\beta^*\left(r, \frac{x - X(r)}{\lambda}\right).$$

It follows by [Csiszár and Matúš, 2012, Theorem 1.1] that the convex conjugate of $\tilde{J}$ equals

$$\tilde{K}(\theta) := \int (\tilde{\beta})^*(r,\theta) d\mu(r) = \lambda \int \beta^* \left(r, \frac{\theta - X(r)}{\lambda} \right) d\mu(r), \quad \theta \in \mathbb{R},$$

or, with the notation (23),

$$\tilde{J}^*(\theta) = \tilde{K}(\theta) = \lambda K(\frac{\theta}{\lambda}, -\frac{1}{\lambda}), \quad \theta \in \mathbb{R}.$$

As the interior of dom $\tilde{J}$ is $(0, +\infty)$, it follows that $\tilde{J}(a) = \tilde{K}^*(a)$ for each $a > 0$. In particular,

$$\begin{aligned} W = \tilde{J}(1) = \tilde{K}^*(1) &= \sup_{\theta \in \mathbb{R}} (\theta - \tilde{K}(\theta)) = \sup_{\theta \in \mathbb{R}} \left[\theta - \lambda K \left(\frac{\theta}{\lambda}, -\frac{1}{\lambda} \right) \right] \\ &= \lambda \sup_{\theta_1 \in \mathbb{R}} \left[\theta_1 - K \left(\theta_1, -\frac{1}{\lambda} \right) \right]. \end{aligned} \tag{50}$$

Proposition 2. *The necessary and sufficient condition for $W > -\infty$ in (49) is the existence of $\theta_1 \in \mathbb{R}$ with*

$$(\theta_1, -1/\lambda) \in \operatorname{dom} K, \tag{51}$$

and then

$$W = \lambda \sup_{\theta_1} \left[\theta_1 - K(\theta_1, -1/\lambda) \right]. \tag{52}$$

If for some $\theta = (\theta_1, -1/\lambda)$ as in (51) the function p_θ in (27) has integral equal to one, then θ_1 attains the maximum in (52), and $p = p_\theta$ attains the minimum in (49). Otherwise, among the numbers θ_1 satisfying (51) there exists a largest one $\theta_{1\max}$, and p_θ with $\theta = (\theta_{1\max}, -1/\lambda)$ has integral less than one; then $\theta_1 = \theta_{1\max}$ attains the maximum in (52).

Proof. Clearly, $W = \tilde{J}(1) > -\infty$ if and only if $\tilde{J}$ never equals $-\infty$, thus its conjugate $\tilde{K}$ is not identically $+\infty$; by the formula for $\tilde{K}$, this proves the first assertion. The second assertion follows from (50). As the supremum in (52) is the same as the supremum defining $\tilde{K}^*(1)$ in (50) (with θ/λ substituted by θ_1), the next assertion follows from the simple instance of [Csiszár and Matúš, 2012, Lemma 4.10] used in Remark 4 (note that the function $(\beta^*)'(r,\theta)$ there, replacing β by $\tilde{\beta}$ and θ by $\theta_1\lambda$, gives the function p_θ in the Proposition). For the last assertion, recall that the maximum in the definition of $\tilde{K}^*(1)$, and therefore in (52), is always attained, because $a = 1$ is in the interior of dom $\tilde{K}^*$ (as in Remark 4). Then the (left) derivative by θ_1 of $K(\theta_1, -1/\lambda)$ at the maximiser, say θ_1^*, has to be ≤ 1, and the strict inequality can hold only if $\theta^* = \theta_{1\max}$. As the mentioned derivative equals the integral of p_{θ^*} with $\theta^* = (\theta_1^*, -1/\lambda)$, this completes the proof. □

As an example, let us reproduce a result of Hansen and Sargent [2001] about the objective function of an agent with multiplier preferences, i.e., about W given by (3) with relative entropy chosen for D. Recalling the subsection about relative entropy balls, Proposition 2 with $\beta(r, s) = s \log s - s + 1$, $\mu = \mathbb{P}_0$ gives the following: The necessary and sufficient condition for $W > -\infty$ is $-1/\lambda \in \operatorname{dom} \Lambda$. Under this condition, the function $\exp(\theta_1 + (-1/\lambda)X(r))$ with $\theta_1 = -\Lambda(-1/\lambda)$, i.e., the member of the exponential family (43) with parameter $\tau = -1/\lambda$ attains the minimum in (49), and

$$W = \lambda\overline{\theta}_1 = -\lambda\Lambda(-1/\lambda).$$

Acknowledgement This work has been supported by the Hungarian National Foundation for Scientific Research under Grant K105840 and by the Austrian Forschungsförderungsgesellschaft in its Josef Ressel Centre for Optimisation under Uncertainty. The second author has benefited from discussions with František Matúš.

References

A. Ahmadi-Javid. Entropic Value at Risk: a new coherent risk measure. *Journal of Optimizaton Theory and Applications*, 155(3):1105–1123, 2011.

S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society Ser. B*, 28:131–142, 1966.

P. Artzner, F. Delbaen, J.-M. Ebner, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

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

L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.

T. Breuer and I. Csiszár. Systematic stress tests with entropic plausibility constraints. *Journal of Banking and Finance*, 37:1552–1559, 2013.

R. Cont. Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16:519–547, 2006.

I. Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publications* *of the Mathematical Institute of the Hungarian Academy of Sciences*, 8: 85–108, 1963.

I. Csiszár. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Annals of Statistics*, 19 (4):2032–2066, 1991.

I. Csiszár and F. Matúš. On minimization of entropy functionals under moment constraints. *Kybernetika*, 48:637–689, 2012.

F. Delbaen. Coherent risk measures on general probability spaces. In K. Sandmann and P. J. Schönbucher, editors, *Advances in Stochastics and Finance: Essays in Honour of Dieter Sondermann*, pages 1–37. Springer, 2002.

H. Föllmer and A. Schied. Robust preferences and convex measures of risk. In K. Sandmann and P. Schonbucher, editors, *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*. Springer, 2002.

H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*, volume 27 of *de Gruyter Studies in Mathematics*. Walter de Gruyter, 2nd edition, 2004.

C. Friedman. Confronting model misspecification in finance: Tractable collections of scenario probability measures for robust financial optimization problems. *International Journal of Theoretical and Applied Finance*, 5: 33–54, 2002.

R. Gibson. *Model Risk: Concepts, Calibration and Pricing*. Risk Books, 2000.

I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

P. D. Grunwald and A. P. Dawid. Game theory, maximum entropy, minimum discrepancy and robust bayesian decision theory. *Annals of Statistics*, 32 (4):1367–1433, 2004.

L. Hansen and T. Sargent. Robust control and model uncertainty. *American Economic Review*, 91:60–66, 2001.

J. Kerkhof, B. Melenberg, and H. Schumacher. Model risk and capital reserves. *Journal of Banking and Finance*, 34:267–279, 2010.

S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

F. Liese and I. Vajda. *Convex Statistical Distances*. Teubner, 1987.

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

M. Nutz and H. M. Soner. Superhedging and dynamic risk measures under volatility uncertainty. *SIAM Journal of Control and Optimization*, 50: 2065–2089, 2012.

J. B. Paris and A. Vencovská. A note on the inevitability of maximum entropy. *International Journal in Inexact Reasoning*, 4:183–223, 1990.

R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer, 1997.

J. E. Shore and R. W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, IT-26:26–37, 1980. Correction IT-29 (1983), 942–943.