

DISTRIBUTION-INVARIANT RISK MEASURES, ENTROPY, AND LARGE DEVIATIONS

STEFAN WEBER,* *Cornell University*

Abstract

The simulation of distributions of financial assets is an important issue for financial institutions. If risk measures are evaluated for a simulated distribution instead of the model-implied distribution, the errors in the risk measurements need to be analyzed. For distribution-invariant risk measures which are continuous on compacts, we employ the theory of large deviations to study the probability of large errors. If the approximate risk measurements are based on the empirical distribution of independent samples, then the rate function equals the minimal relative entropy under a risk measure constraint. We solve this minimization problem explicitly for shortfall risk and average value at risk.

Keywords: Risk measure; average value at risk; shortfall risk; Monte Carlo; large deviations principle; Sanov's theorem; relative entropy

2000 Mathematics Subject Classification: Primary 91B30; 49Q20; 62B10; 62D05; 91B28

1. Introduction

The portfolios of banks consist in financial assets such as stocks, bonds, credits, and options. The quantification of the risk associated with these assets is of crucial importance, since banks need to manage their risks and are obliged to respect regulatory constraints. This requires both suitable models of portfolio holdings and appropriate numerical measures of risk. In practice, financial assets are frequently modeled as real-valued random variables on some underlying probability space. In such a setting, the modeling assumptions determine in particular the distributions of the financial assets. A standard approach taken to measure risk is to use certain functionals of these distributions, namely static, distribution-invariant risk measures.

A theory of such risk measures is already well developed. Nevertheless, the implementation of risk measurement requires further analysis. Model distributions are often not directly tractable, but can only be simulated using Monte Carlo methods. If risk measures are evaluated for the simulated distributions instead of the model-implied distributions, actual risk measurements deviate from the model-implied risk and the errors in these measurements need to be analyzed.

In this paper we employ the theory of large deviations to study these errors for various risk measures. We investigate large deviation bounds for a broad class of static risk measures. Specific examples include shortfall risk and average value at risk. For an axiomatic analysis of coherent risk measures, we refer the reader to Artzner *et al.* (1999); for extensions to general probability spaces and convex risk measures, see Delbaen (2002), Föllmer and Schied (2002), (2004, pp. 152–220), and Frittelli and Rosazza (2002).

The paper is structured as follows. In a first step, we describe how the errors in the risk measurements and large deviations are related. In Section 2 we investigate conditions

Received 10 August 2004; revision received 4 December 2006.

* Postal address: School of Operations Research and Industrial Engineering, Cornell University, 279 Rhodes Hall, Ithaca, NY 14853, USA. Email address: sweber@orie.cornell.edu

under which a large deviation principle (LDP) holds for risk measurements. An LDP can be derived from a contraction principle if the risk measure has a certain regularity property, namely continuity on compacts (i.e. compact sets). This notion is introduced in Section 2, and a contraction principle for the corresponding class of risk measures is formulated. In Section 3 we analyze the notion of continuity on compacts. In particular, we characterize coherent, distribution-invariant risk measures which are continuous on compacts. Examples include average value at risk and shortfall risk. Further properties of these risk measures were discussed by Föllmer and Schied (2004, pp. 152–220), Weber (2006), Giesecke *et al.* (2005), and Dunkel and Weber (2005).

For risk measurements based on empirical distributions, the rate function of the LDP can be characterized more explicitly. For independent samples, the rate function of the large deviations of the risk measurements is given as the minimal relative entropy under a risk measure constraint. We calculate the minimal relative entropy explicitly for both shortfall risk and average value at risk in Sections 4 and 5, based on general methods of Csiszár (1975).

For a shortfall risk constraint, the calculation of the minimal relative entropy only involves a linear constraint. A solution to the problem is obtained in Section 4. In the case of average value at risk, which we consider in Section 5, the analysis is more complicated. Our solution is based on a particular representation of average value at risk as the expected loss with respect to the worst-case measure which can be computed by means of the Neyman–Pearson lemma (see (5.1)). The constraint set of the minimization problem is in general not convex, and the calculation is quite involved. Necessary and sufficient conditions for the existence of a solution are formulated in terms of the parameters of the problem. For average value at risk, we solve the original problem in two steps. The first step is to minimize the relative entropy under a linear constraint. Minimizing densities and minimal relative entropies are explicitly calculated. The second step is to solve a minimization problem with three varying parameters.

2. Large deviation bounds

Throughout the paper, we assume that (Ω, \mathcal{F}, P) is a rich probability space, i.e. a probability space on which a random variable with a continuous distribution exists. We recall the following definition.

Definition 2.1. A mapping $\rho: L^\infty(\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}$ is called a distribution-invariant risk measure if it satisfies the following conditions for all $X, Y \in L^\infty$.

- If $X \leq Y$ then $\rho(X) \geq \rho(Y)$ (monotonicity).
- If $m \in \mathbb{R}$ then $\rho(X + m) = \rho(X) - m$ (the translation property).
- If $P \circ X^{-1} = P \circ Y^{-1}$ then $\rho(X) = \rho(Y)$ (distribution invariance).

We denote by $\mathcal{M}_{1,c} \equiv \mathcal{M}_{1,c}(\mathbb{R})$ the space of Borel probability measures on \mathbb{R} with compact support. We endow $\mathcal{M}_{1,c}$ with the weak topology. A distribution-invariant risk measure ρ defines a functional $\rho': \mathcal{M}_{1,c} \rightarrow \mathbb{R}$ by $\rho'(\mu) = \rho(X)$ for some $X \in L^\infty$ with distribution $\mathcal{L}(X) := P \circ X^{-1} = \mu$. For more details, see Weber (2006), (2004).

We consider the following situation. Assume that we are interested in the risk of a financial asset $X \in L^\infty$ with distribution $\mu = \mathcal{L}(X)$. Suppose that the distribution μ is not directly tractable, but that samples of μ can be generated. For example, let (X_i) be a sequence of independent random variables on the probability space (Ω, \mathcal{F}, P) , with common distribution μ . The empirical distribution of the first n samples, X_1, X_2, \dots, X_n , is then given

by the random measure

$$\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \tag{2.1}$$

where δ_x denotes the Dirac measure placing all mass at $x \in \mathbb{R}$. Thus, (μ_n) converges P-almost surely (P-a.s.) to μ in the weak topology.

A naive Monte Carlo procedure for simulating $\rho(X)$ is to calculate $\rho'(\mu_n)$, $n \in \mathbb{N}$. A possible measure of the quality of the n th approximation is the probability that the error of the simulated risk deviates from the true risk of X by more than a given bound $\varepsilon > 0$, i.e.

$$P(|\rho'(\mu_n) - \rho(X)| > \varepsilon).$$

If ρ is regular enough, asymptotic upper and lower exponential bounds for these error probabilities can be obtained from the theory of large deviations.

Definition 2.2. A distribution-invariant risk measure $\rho: L^\infty \rightarrow \mathbb{R}$ is called *continuous on compacts* if, for all compact sets $K \subseteq \mathbb{R}$, the restriction of ρ' to $\mathcal{M}_1(K)$ is continuous. Here $\mathcal{M}_1(K)$ denotes the space of probability measures supported in K .

The notion of continuity on compacts is weaker than continuity of risk measures with respect to the weak topology. For risk measures which are continuous on compacts, an LDP is an immediate consequence of the contraction principle.

Proposition 2.1. *Let ρ be a distribution-invariant risk measure that is continuous on compacts. Assume that $(\mu_n) \subseteq \mathcal{M}_{1,c}$ is a sequence of random measures that satisfies an LDP with rate sequence (γ_n) and rate function I . Additionally, assume that there exists a compact set $K \subseteq \mathbb{R}$ such that $\text{supp } \mu_n \subseteq K$ for all n . Then $(\rho'(\mu_n))_n$ satisfies an LDP with rate sequence (γ_n) and rate function*

$$J(x) := \inf\{I(v) : v \in \mathcal{M}_{1,c}, x = \rho'(v)\}.$$

Proof. The proposition is an immediate consequence of the contraction principle for Hausdorff spaces (see Dembo and Zeitouni (1998, Theorem 4.2.1)).

Remark 2.1. Value at risk is not continuous on compacts. Nevertheless, it is possible to derive upper large deviation bounds by direct calculations; see, e.g. Fu *et al.* (2003) and Weber (2004).

We now specialize to the case of empirical measures μ_n as defined in (2.1). For independent samples, the sequence $(\mu_n)_n$ converges P-a.s. to μ . Thus, a strong law of large numbers holds for the risk measurements. At the same time, we have the following LDP for the risk measures.

Corollary 2.1. *Let ρ be continuous on compacts. Then $(\rho'(\mu_n))_n$ satisfies an LDP with rate n and rate function*

$$J(x) = \inf\{H(v|\mu) : v \in \mathcal{M}_{1,c}, x = \rho'(v)\}.$$

Here $H(v|\mu)$ denotes the relative entropy of the probability measure v with respect to μ , defined by

$$H(v|\mu) := \begin{cases} \int f \log f \, d\mu & \text{if } f := dv/d\mu \text{ exists,} \\ \infty & \text{otherwise.} \end{cases}$$

Proof. The proof is a corollary of Sanov’s theorem (see, e.g. Dembo and Zeitouni (1998, Theorem 6.2.10)) and the contraction principle for empirical measures stated in Proposition 2.1.

3. Continuity on compacts

In this section we characterize risk measures which are continuous on compacts and thus satisfy the contraction principle of the preceding section.

The following theorem is elementary, but allows us to identify examples of risk measures which are continuous on compacts. We recall that a risk measure ρ is said to be *continuous from above* if $X_n \searrow X$ P-a.s. implies that $\rho(X_n) \nearrow \rho(X)$. Analogously, ρ is said to be *continuous from below* if $X_n \nearrow X$ P-a.s. implies that $\rho(X_n) \searrow \rho(X)$.

Theorem 3.1. *Let ρ be a distribution-invariant risk measure. The following conditions are equivalent.*

1. ρ is continuous on compacts.
2. ρ is continuous both from above and from below.
3. ρ is continuous for bounded sequences, i.e. for every bounded sequence (X_n) converging P-a.s. to some X , it holds that $\lim_{n \rightarrow \infty} \rho(X_n) = \rho(X)$.

Proof. Condition 2 implies condition 3. Let (X_n) be bounded and P-a.s. convergent to X . Then $(\sup_{m \geq n} X_m)_n$ and $(\inf_{m \geq n} X_m)_n$ converge to X from above and below, respectively. Thus,

$$\rho(X) = \lim_n \rho\left(\sup_{m \geq n} X_m\right) \leq \liminf_{n \rightarrow \infty} \rho(X_n) \leq \limsup_{n \rightarrow \infty} \rho(X_n) \leq \lim_n \rho\left(\inf_{m \geq n} X_m\right) = \rho(X).$$

Condition 3 implies condition 2. If (X_n) converges to X from below or from above, then (X_n) is bounded. This implies the claim.

Condition 1 implies condition 3. Let (X_n) be a bounded sequence converging P-a.s. to some X . Then there exists a compact $K \subseteq \mathbb{R}$ such that $X_n, X \in K, n \in \mathbb{N}$, P-a.s. Clearly $\mathcal{L}(X_n), \mathcal{L}(X) \in \mathcal{M}_1(K)$ and $\mathcal{L}(X_n) \rightarrow \mathcal{L}(X)$. Thus, $\rho(X_n) = \rho'(\mathcal{L}(X_n)) \rightarrow \rho'(\mathcal{L}(X)) = \rho(X)$.

Condition 3 implies condition 1. Let $K \subseteq \mathbb{R}$ be a compact set and assume that $\mu_n \Rightarrow \mu$ for $\mu_n, \mu \in \mathcal{M}_1(K)$ (where ‘ \Rightarrow ’ denotes weak convergence of probability measures). Denote by F_n and F the distribution functions of μ_n and μ , respectively. Since (Ω, \mathcal{F}, P) is rich, there exists a random variable Z with $\mathcal{L}(Z) = \text{unif}[0, 1]$. Define $X_n := F_n^{-1}(Z)$ and $X := F^{-1}(Z)$, where F_n^{-1} and F^{-1} are the right-continuous inverses of F_n and F , respectively. Observe that $X_n \rightarrow X$ P-a.s. as $n \rightarrow \infty$. Moreover, $X_n, X \in K$ P-a.s. Hence,

$$\rho'(\mu_n) = \rho(X_n) \rightarrow \rho(X) = \rho'(\mu). \tag{3.1}$$

This completes the proof.

We provide examples for risk measures which are continuous on compacts. The current industry standard, *value at risk*, is not continuous on compacts. In contrast, the coherent risk measure *average value at risk* is continuous on compacts. We recall that a risk measure is coherent if for $X, Y \in L^\infty$ it satisfies both

- $\rho(\alpha X + (1 - \alpha)Y) \leq \alpha\rho(X) + (1 - \alpha)\rho(Y)$ for $\alpha \in [0, 1]$ (convexity) and
- $\rho(\alpha X) = \alpha\rho(X)$ for $\alpha \geq 0$ (positive homogeneity).

Example 3.1. The value at risk at level $\lambda \in (0, 1)$ is defined by

$$\text{VaR}_\lambda(X) = \inf\{m \in \mathbb{R}: \mathbb{P}(m + X < 0) \leq \lambda\}.$$

In order to see that value at risk is not continuous on compacts, we consider the following example. Let the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be given by the unit interval $[0, 1]$ and the Lebesgue measure. For $1/n < 1 - \lambda$, define $X_n = \mathbf{1}_{[\lambda+1/n, 1]}$. Then $X_n \nearrow X := \mathbf{1}_{[\lambda, 1]}$ as $n \rightarrow \infty$. However, $\text{VaR}_\lambda(X_n) = 0$ does not converge to $\text{VaR}_\lambda(X) = -1$ as $n \rightarrow \infty$.

Example 3.2. The average value at risk at level $\lambda \in (0, 1]$ is defined by

$$\text{AVaR}_\lambda(X) = \frac{1}{\lambda} \int_0^\lambda \text{VaR}_\gamma(X) \, d\gamma.$$

Average value at risk is continuous on compacts.

According to Föllmer and Schied (2004, Theorem 4.47), AVaR_λ is continuous from below. Furthermore, by Föllmer and Schied (2004, Theorem 4.31), AVaR_λ is continuous from above. Thus, AVaR is continuous on compacts, by Theorem 3.1.

Average value at risk is an important building block in the construction of coherent, distribution-invariant risk measures. We quote the following result of Kusuoka (2001).

Theorem 3.2. *On a rich probability space, a coherent, distribution-invariant risk measure ρ is continuous from above if and only if*

$$\rho(X) = \sup_{\mu \in \mathcal{M}} \int_{(0,1]} \text{AVaR}_\lambda(X) \mu(d\lambda) \tag{3.2}$$

for some set $\mathcal{M} \subseteq \mathcal{M}_1((0, 1])$. Here $\mathcal{M}_1((0, 1])$ denotes the space of Borel probability measures on $(0, 1]$. The assumption of continuity from above can be dropped if the underlying probability space is standard; see Jouini et al. (2006).

This representation of coherent, distribution-invariant risk measures provides a further perspective on risk measures which are continuous on compacts; see Theorem 3.3 and Corollary 3.2, below.

We denote by $\text{AVaR}_0(X) := \|X^-\|$ the essential infimum of X . Motivated by Theorem 3.2, we introduce the following notation. For a measure $\mu \in \mathcal{M}_1([0, 1])$, i.e. a Borel probability measure on $[0, 1]$, we write

$$\rho_\mu(X) := \int_{[0,1]} \text{AVaR}_\lambda(X) \mu(d\lambda), \quad X \in L^\infty.$$

For coherent risk measures we can now state necessary and sufficient conditions for continuity on compacts.

Theorem 3.3. *For a coherent, distribution-invariant risk measure ρ , the following properties are equivalent.*

1. ρ is continuous from below.
2. ρ is continuous on compacts.
3. ρ is continuous in the Mackey topology $\tau(L^\infty, L^1)$.

4. There exists a law-invariant set $\mathcal{D} \subseteq L^1$ of densities representing ρ such that the supremum is attained as follows:

$$\rho(X) = \max_{h \in \mathcal{D}} E[-hX] \quad \text{for all } X \in L^\infty.$$

5. There exists a $\sigma(L^1, L^\infty)$ -compact, law-invariant set $\mathcal{D} \subseteq L^1$ of densities representing ρ , such that

$$\rho(X) = \sup_{h \in \mathcal{D}} E[-hX] \quad \text{for all } X \in L^\infty.$$

6. The maximal representative set,

$$\mathcal{D}_{\max} := \left\{ h \in L^1 : \|h\|_{L^1} = 1, h \geq 0, \sup_{X \in L^\infty} [E[-hX] - \rho(X)] = 0 \right\},$$

is $\sigma(L^1, L^\infty)$ -compact.

7. There exists a law-invariant set $\mathcal{D} \subseteq L^1$ of densities representing ρ such that the supremum is attained as follows:

$$\rho(X) = \max_{h \in \mathcal{D}} \int_0^1 q_h(t)q_{-X}(t) dt \quad \text{for all } X \in L^\infty.$$

Here q_Y denotes the quantile function of a random variable Y .

8. There exists a set $\mathcal{M} \subseteq \mathcal{M}((0, 1])$ of probability measures representing ρ such that the supremum is attained as follows:

$$\rho(X) = \max_{\mu \in \mathcal{M} \subseteq \mathcal{M}_1((0, 1])} \int_{(0, 1]} AVaR_\gamma(X) \mu(d\gamma) \quad \text{for all } X \in L^\infty.$$

In properties 4, 5, and 7, the set \mathcal{D} may be chosen to be convex. The set \mathcal{D}_{\max} in property 6 is always convex.

Proof. Property 1 is equivalent to property 2. This follows from Föllmer and Schied (2004, Corollary 4.35 and Theorem 4.31) and Theorem 3.1.

Property 5 implies property 1, which is equivalent to property 4. Except for the law invariance of \mathcal{D} in property 4, this is an immediate consequence of Föllmer and Schied (2004, Corollary 4.35). We show that \mathcal{D} can be chosen to be law invariant. By Föllmer and Schied (2004, Corollary 4.34), we may choose \mathcal{D} to be the maximal representative set of densities,

$$\mathcal{D}_{\max} = \{h \in L^1 : \|h\|_{L^1} = 1, h \geq 0, \alpha_{\min}(h) = 0\},$$

defined in terms of the minimal penalty function,

$$\alpha_{\min}(h) = \sup_{X \in L^\infty} [E[-hX] - \rho(X)].$$

By Föllmer and Schied (2004, Theorem 4.54), α_{\min} depends only on the law of its argument. Thus, \mathcal{D}_{\max} is law invariant.

Property 4 implies property 5. This is a consequence of James' theorem (see Föllmer and Schied (2004, Theorem A.66)). The proof parallels the argument in Föllmer and Schied (2004, Corollary 4.35). The set \mathcal{D} is law invariant by assumption.

Property 4 implies property 6. As we pointed out in the second paragraph of the proof, the set \mathcal{D} in property 4 can be assumed to be \mathcal{D}_{\max} . The weak compactness of \mathcal{D}_{\max} follows, again by James' theorem.

Property 6 implies property 4. This is a special case of the implication of property 4 by property 5.

Property 4 implies property 3. This is an immediate consequence of Aliprantis and Border (1999, Theorem 5.102).

Property 3 implies property 4. By Aliprantis and Border (1999, Theorem 5.102) we have $\rho(X) = \max_{h \in \mathcal{D}} E[-hX]$, where

$$\mathcal{D} = \{h \in L^1 : E[-hX] \leq \rho(X) \text{ for all } X \in L^\infty\}$$

is $\sigma(L^1, L^\infty)$ -compact.

If $h \in \mathcal{D}$ then $h \geq 0$ P-a.s. Suppose that $h \geq 0$ P-a.s. does not hold. With $X = \mathbf{1}_{\{h < 0\}}$, for $n \in \mathbb{N}$ we obtain

$$E[-nh\mathbf{1}_{\{h < 0\}}] \leq \rho(nX) \leq \rho(X).$$

The left-hand side of this inequality converges to ∞ as n tends to ∞ , contradicting the fact that $\rho(X) \in \mathbb{R}$.

Finally, with $X = 1$ we obtain $E[-h] \leq \rho(1) = -1$ and, thus, $E[h] \geq 1$, and with $X = -1$ we obtain $E[h] \leq \rho(-1) = 1$. This implies that \mathcal{D} is a set of densities. Since (Ω, \mathcal{F}, P) is a rich probability space and ρ is distribution invariant, we can use Kallenberg (1997, Theorem 6.10) to show that \mathcal{D} is law invariant.

Property 7 implies property 8. For a probability density $h \in L^1$, we define

$$[h] = \{\hat{h} \in L^1 : \hat{h} \stackrel{P}{\sim} h\}$$

(where $\stackrel{P}{\sim}$ denotes equality in distribution under the measure P). We show that the equivalence classes

$$\{[h] : h \in L^1, h \geq 0, \|h\|_{L^1} = 1\}$$

and the probability measures $\mathcal{M}((0, 1])$ are in one-to-one correspondence.

Let μ be given. By Föllmer and Schied (2004, Lemma 4.63), μ corresponds uniquely to an increasing, concave function $\psi : [0, 1] \rightarrow [0, 1]$ with $\psi(0) = 0$ and $\psi(1) = 1$. Its right-continuous right-hand derivative, ψ'_+ , is therefore positive and satisfies

$$\int_0^1 \psi'_+(t) dt = \psi(1) - \psi(0) = 1.$$

The derivative ψ'_+ hence defines a probability density on $((0, 1), \mathcal{B}, \lambda)$, where λ denotes the Lebesgue measure. Since (Ω, \mathcal{F}, P) is a rich probability space, there exists an $h \in L^1$ with $\mathcal{L}(h; P) = \mathcal{L}(\psi'_+; \lambda)$, where the notation $\mathcal{L}(\cdot; \cdot)$ makes explicit the dependence of the distribution on the measure. We define $M(\mu) := [h]$.

Conversely, let $h \in L^1$ be a probability density with equivalence class $[h]$. Then $\psi(t) = \int_0^t q_h(1 - u) du$ depends on the distribution of h only. We have $\psi(0) = 0$ and $\psi(1) = E_P[h] = 1$ (where E_P denotes expectation under the probability measure P), since h is a probability density. The function ψ is increasing, since $q_h \geq 0$. Moreover, $t \mapsto q_h(1 - t)$ is decreasing, which implies that ψ is concave. By Föllmer and Schied (2004, Lemma 4.63),

ψ corresponds to a unique probability measure $\mu \in \mathcal{M}_1((0, 1])$. From the construction it is obvious that $[h] = M(\mu)$.

For a fixed $\mu \in \mathcal{M}((0, 1])$, we set $\rho_\mu(X) = \int_{(0,1]} AVaR_\gamma(X)\mu(d\gamma)$. Let ψ be defined, according to Föllmer and Schied (2004, Lemma 4.63), as the function corresponding to μ . By construction, for $h \in M(\mu)$ we have $q_h(t) = \psi'_+(1 - t)$ for $t \in [0, 1]$. By Föllmer and Schied (2004, Lemma 4.63 and Theorem 4.64),

$$\rho_\mu(X) = \int_0^1 q_{-X}(t)\psi'_+(1 - t) dt = \int_0^1 q_{-X}(t)q_h(t) dt.$$

Now assume that property 7 holds. Then

$$\rho(X) = \max_{h \in \mathcal{D}} \int_0^1 q_h(t)q_{-X}(t) dt = \max_{\mu=M^{-1}([h]), h \in \mathcal{D}} \rho_\mu(X) = \max_{\mu \in \mathcal{M}} \rho_\mu(X),$$

where $\mathcal{M} = \{M^{-1}([h]) : h \in \mathcal{D}\}$. Conversely, if property 8 holds we obtain

$$\rho(X) = \max_{\mu \in \mathcal{M}} \rho_\mu(X) = \max_{h \in \mathcal{D}} \int_0^1 q_h(t)q_{-X}(t) dt,$$

where $\mathcal{D} = \bigcup_{\mu \in \mathcal{M}} M(\mu)$.

Property 4 implies property 7. For $h \in \mathcal{D}$, by the Hardy–Littlewood inequality (see Föllmer and Schied (2004, Theorem A.24)) we have

$$E[-hX] \leq \int_0^1 q_h(t)q_{-X}(t) dt.$$

Since (Ω, \mathcal{F}, P) is rich, there exists a pair of random variables (\tilde{h}, Y) such that $\mathcal{L}((\tilde{h}, -Y); P) = \mathcal{L}((q_h, q_{-X}); \lambda)$. Note that of course $\tilde{h} \sim h$ and $Y \sim X$, but that \tilde{h} and Y are anticomontonic. This implies that

$$\int_0^1 q_h(t)q_{-X}(t) dt = E[-\tilde{h}Y] \leq \rho(Y) = \rho(X).$$

Taking suprema yields

$$\rho(X) = \max_{h \in \mathcal{D}} E[-hX] = \sup_{h \in \mathcal{D}} \int_0^1 q_h(t)q_{-X}(t) dt.$$

Then, letting $h^* = \operatorname{argmax}_{h \in \mathcal{D}} E[-hX]$, we have

$$\rho(X) = E[-h^*X] \leq \int_0^1 q_{h^*}(t)q_{-X}(t) dt \leq \rho(X).$$

Hence, the supremum in property 7 is attained.

Property 8 implies property 1. For a fixed $\mu \in \mathcal{M}((0, 1])$, we set

$$\rho_\mu(X) = \int_{(0,1]} AVaR_\gamma(X)\mu(d\gamma).$$

By Föllmer and Schied (2004, Corollary 4.74), there exists a set of densities \mathcal{D}_μ such that $\rho_\mu(X) = \max_{h \in \mathcal{D}_\mu} E[-hX]$. Letting $\mathcal{D} = \bigcup_{\mu \in \mathcal{M}} \mathcal{D}_\mu$, we obtain $\rho(X) = \max_{h \in \mathcal{D}} E[-hX]$. By Föllmer and Schied (2004, Corollary 4.35), it follows that ρ is continuous from below.

Remark 3.1. If (Ω, \mathcal{F}, P) is a standard probability space, then every distribution-invariant, convex risk measure is continuous from above. This has recently been pointed out by Jouini *et al.* (2006). Their result provides another justification for the fact that continuity on compacts is implied by continuity from below.

Remark 3.2. If the maximizing density in Theorem 3.3 can be explicitly calculated, this can be used to explicitly find the minimal relative entropy under the risk measure constraint. This provides a way to obtain the rate function in Corollary 2.1. We will consider the case of average value at risk in Section 5.

Corollary 3.1. *For a risk measure $\rho : L^\infty \rightarrow \mathbb{R}$, the following properties are equivalent.*

1. ρ is distribution-invariant, coherent, and continuous on compacts.
2. The acceptance set, \mathcal{A} , of ρ is a law-invariant, convex cone, and, for increasing sequences $(X_n)_n \subseteq L^\infty$ with $X = \lim_{n \rightarrow \infty} X_n$ P-a.s., $X \in \mathcal{A}$, and any $\varepsilon > 0$, we have $X_n + \varepsilon \in \mathcal{A}$ for large enough n .

Proof. Property 1 implies property 2. This is obvious.

Property 2 implies property 1. The law invariance and coherence of ρ are immediate. By Theorem 3.3 we need to verify continuity from below. Letting $Y_n \nearrow Y$ P-a.s., we have $X = Y + \rho(Y) \in \mathcal{A}$ and $X_n = Y_n + \rho(Y) \nearrow X$ P-a.s. For any $\varepsilon > 0$, we have $\rho(X_n + \varepsilon) \leq 0$ for large enough n ; thus, $\rho(Y_n) \leq \rho(Y) + \varepsilon$. This implies that $\rho(Y_n)$ converges to $\rho(Y)$ as n tends to ∞ .

Corollary 3.2. *Suppose that a coherent, distribution-invariant risk measure ρ admits a representation as in (3.2) for some weakly compact set $\mathcal{M} \subseteq \mathcal{M}_1((0, 1])$. Then the supremum in (3.2) is actually a maximum, and ρ is continuous on compacts.*

Proof. Since $\lambda \mapsto AVaR_\lambda(X)$ is continuous and bounded in $[-\|X\|, \|X\|]$,

$$\mu \mapsto \int_{(0,1]} AVaR_\gamma(X) \mu(d\gamma)$$

defines a weakly continuous functional on $\mathcal{M}((0, 1])$. Since \mathcal{M} is weakly compact, the supremum in (3.2) is attained. By Theorem 3.3, it follows that ρ is continuous on compacts.

The condition of Corollary 3.2 is of course satisfied if the set \mathcal{M} is a singleton. In this case, ρ is simply a mixture of average values at risk at different levels. By a result of Schmeidler (1986), the class of such risk measures is closely related to the family of distribution-invariant risk measures that are comonotonic; see Föllmer and Schied (2004, Section 4.7).

Theorem 3.4. *On a rich probability space, the class of risk measures*

$$\rho_\mu(X) = \int AVaR_\lambda(X) \mu(d\lambda), \quad \mu \in \mathcal{M}_1([0, 1]),$$

is precisely the class of all distribution-invariant, convex risk measures on L^∞ that are comonotonic. In particular, these risk measures are also coherent. Furthermore, ρ_μ is continuous on compacts if and only if $\mu(\{0\}) = 0$.

Proof. For the proof of the first part of the theorem, see Föllmer and Schied (2004, Theorem 4.87). The second part follows from Föllmer and Schied (2004, Corollary 4.74) and Theorem 3.1.

Remark 3.3. The risk measure ρ_μ can also be represented as a Choquet integral with respect to a continuous, concave distortion of the underlying probability measure; see Föllmer and Schied (2004, Corollary 4.71)

Finally, we discuss an important class of distribution-invariant, convex risk measures which are continuous on compacts, namely shortfall risk.

Definition 3.1. Let $\ell: \mathbb{R} \rightarrow \mathbb{R}$ be a convex loss function, i.e. an increasing, nonconstant, convex function. Assume that z is an interior point of the range of ℓ . We define the acceptance set

$$\mathcal{A} = \left\{ X \in L^\infty : \int \ell(-X) \, dP \leq z \right\}.$$

The shortfall risk is defined by

$$\rho(X) = \inf\{m \in \mathbb{R} : X + m \in \mathcal{A}\}.$$

Shortfall risk is a distribution-invariant risk measure which is continuous from above and below (see Föllmer and Schied (2004, Proposition 4.104 and Theorem 4.31)). Thus, by Theorem 3.1, shortfall risk is continuous on compacts. Like average value at risk, shortfall risk has many desirable properties. In contrast to value at risk, it encourages diversification, since it is convex, and does not neglect the size of losses. For a detailed analysis of this risk measure, including applications to dynamic risk measurement and Monte Carlo simulations, we refer the reader to Weber (2006), Giesecke *et al.* (2005), and Dunkel and Weber (2005).

4. Entropy minimization under a shortfall risk constraint

As we have seen in Corollary 2.1, for independently generated samples the rate function of the large deviations of risk measures is determined by the minimal relative entropy under a risk measure constraint. In the current section we consider a first example of the entropy minimization problem under a risk measure constraint: we discuss a shortfall risk constraint. Shortfall risk has many appealing properties. It is distribution invariant, coherent, and sensitive to the size of losses. In contrast to average value at risk, shortfall risk can be used for the weakly consistent dynamic evaluation of financial assets (see Weber (2006)).

The minimization problem. Let $\ell: \mathbb{R} \rightarrow \mathbb{R}$ be a convex loss function, let $z \in \mathbb{R}$ be a point in the interior of the range of ℓ , and let ρ be the shortfall risk associated with ℓ and z . Fix a constant $y \in \mathbb{R}$ and a reference probability $\mu \in \mathcal{M}_{1,c}$ with compact support. We are interested in the problem of minimizing $H(v|\mu)$ where $v \in \mathcal{M}_{1,c}$ and $\rho'(v) = y$. We set $a := \inf\{x \in \mathbb{R} : x \in \text{supp } \mu\}$ and $b := \sup\{x \in \mathbb{R} : x \in \text{supp } \mu\}$; thus, $\text{supp } \mu \subseteq [a, b]$. Observe that $\text{supp } v \not\subseteq \text{supp } \mu$ implies that $v \not\ll \mu$ and, thus, $H(v|\mu) = \infty$. We may therefore restrict our attention to the constraint set

$$\mathcal{C} := \{v \in \mathcal{M}_1([a, b]) : \rho'(v) = y\}.$$

4.1. Existence of solutions

A necessary and sufficient criterion for the existence of solutions can be formulated in terms of the parameters a , b , and y . We need the following general result, which will also be used in Section 5. Since shortfall risk is continuous on compacts, the entropy minimization problem under a shortfall risk constraint represents a special case of the next lemma.

Let $\mu \in \mathcal{M}_{1,c}$, and $a, b, y \in \mathbb{R}$ be given as above. For any distribution-invariant risk measure $\rho: L^\infty \rightarrow \mathbb{R}$ which is continuous on compacts, we define

$$\mathcal{C}_\rho := \{v \in \mathcal{M}_1([a, b]) : \rho'(v) = y\}.$$

We consider the problem of minimizing $H(\cdot|\mu)$ on \mathcal{C}_ρ .

Lemma 4.1. *Suppose that $\mathcal{C}_\rho \neq \emptyset$. There exists a solution to the entropy minimization problem with constraint set \mathcal{C}_ρ . If there exists a $v \in \mathcal{C}_\rho$ such that $H(v|\mu) < \infty$, then the minimizer has finite relative entropy.*

Proof. If $H(v|\mu) = \infty$ for all $v \in \mathcal{C}_\rho$, then any $v \in \mathcal{C}_\rho$ minimizes the relative entropy. Otherwise, observe that $\mathcal{M}_1([a, b])$ is weakly compact since $[a, b]$ is compact. Since ρ is continuous on compacts, \mathcal{C}_ρ is then also a weakly compact set. Since $H(\cdot|\mu)$ is lower semicontinuous, it achieves its minimum on \mathcal{C}_ρ .

In the case of shortfall risk, the following characterization theorem is essential. A proof is contained in the proof of Föllmer and Schied (2004, Proposition 4.104).

Proposition 4.1. *Let $X \in L^\infty$. Then $\rho(X) = y$ if and only if $\int \ell(-X - y) d\mathbb{P} = z$.*

The following proposition characterizes the existence of solutions.

Proposition 4.2. *The following conditions are equivalent.*

1. *There exists a $v \in \mathcal{C}$ such that $H(v|\mu) < \infty$.*
2. *The minimal value of the relative entropy on \mathcal{C} is finite, and is attained for some element of \mathcal{C} .*
3. *The point a is an atom of μ and $\ell(-a - y) = z$, the point b is an atom of μ and $\ell(-b - y) = z$, or*

$$\ell(-b - y) < z < \ell(-a - y). \tag{4.1}$$

If (4.1) holds then there exists a $v \approx \mu$, $v \in \mathcal{C}$, with $H(v|\mu) < \infty$.

Proof. See Appendix A.

4.2. Structure of the solution

Since shortfall risk imposes a linear constraint, the solution to the minimization problem can be characterized. Its density with respect to μ is of exponential form. The exponent is a linear combination of the constraint functions. We quote a result of Csiszár (1975).

Theorem 4.1. *For $i = 1, 2, \dots, I$ with $I \in \mathbb{N}$, let $f_i: \mathbb{R} \rightarrow \mathbb{R}$ be measurable functions and let $a_i \in \mathbb{R}$. Let $\mu \in \mathcal{M}_1(\mathbb{R})$ and define the constraint set*

$$\hat{\mathcal{C}} = \left\{ v \in \mathcal{M}_1(\mathbb{R}) : \int f_i(x)v(dx) = a_i, \quad i = 1, 2, \dots, I \right\}.$$

Assume that there exists a $v \in \hat{\mathcal{C}}$ with $v \approx \mu$ and $H(v|\mu) < \infty$. Then there exists a unique minimizer on $\hat{\mathcal{C}}$ with finite relative entropy. The measure v is the minimizer if and only if its μ -density is of the following form, with normalizing constant $c > 0$ and $h_i \in \mathbb{R}$, $i = 1, 2, \dots, I$:

$$\frac{dv}{d\mu} = c \exp\left(\sum_{i=1}^I h_i f_i\right).$$

Corollary 4.1. *Assume that (4.1) holds. Then ν is the unique minimizer of the relative entropy on \mathcal{C} if and only if its μ -density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \exp(h\ell(-x - y)). \tag{4.2}$$

Here $c > 0$ is a normalizing constant, $h \in \mathbb{R}$, and the following conditions need to be satisfied:

$$z = c \int \ell(-x - y) \exp(h\ell(-x - y))\mu(dx), \tag{4.3}$$

$$1 = c \int \exp(h\ell(-x - y))\mu(dx). \tag{4.4}$$

Proof. By Proposition 4.2 there exists a $\nu \in \mathcal{C}$ with $H(\nu|\mu) < \infty$ and $\nu \approx \mu$. It follows from Theorem 4.1 that there exists a unique minimizer of the relative entropy. By Proposition 4.1, a measure $\nu \in \mathcal{M}_1([a, b])$ is an element of \mathcal{C} if and only if $\int \ell(-x - y)\nu(dx) = z$. Thus, the minimizer ν has μ -density (4.2) by Theorem 4.1. Condition (4.3) is required by the constraint and condition (4.4) is a normalization.

If ν is the minimizing density characterized in Corollary 4.1, then the minimal relative entropy is given by the expression

$$H(\nu|\mu) = \log c + hz.$$

Remark 4.1. If a is an atom of μ and $\ell(-a - y) = z$, or if b is an atom of μ and $\ell(-b - y) = z$, then $\mathcal{C} = \{\delta_x\}$ with $x = a$ or $x = b$, respectively. Here δ_x denotes the Dirac measure at $x \in \mathbb{R}$. Furthermore, the minimizer, ν , of the relative entropy is trivially unique, and $H(\nu|\mu) = -\log \mu\{x\}$ with $x = a$ or $x = b$ in the same respective cases.

5. Entropy minimization under average value at risk constraints

In the current section we discuss the minimization problem for a second class of risk measures, namely average value at risk. Average value at risk also has appealing properties. It is distribution invariant, coherent, and, in the event of a large loss, takes the size of the loss into account. The last fact follows from, for example, the following representation of average value at risk at level λ , where q is some λ -quantile of the random variable X :

$$AVaR_\lambda(X) = \frac{1}{\lambda} E[(q - X)^+] - q.$$

The minimization problem. Fix a reference probability measure $\mu \in \mathcal{M}_{1,c}$ with compact support, let $\lambda \in (0, 1)$ be a level, and let $y \in \mathbb{R}$ be a constant. We are interested in the problem of minimizing $H(\nu|\mu)$ where $\nu \in \mathcal{M}_{1,c}$ and $AVaR_\lambda(\nu) = y$. We set $a := \inf\{x \in \mathbb{R} : x \in \text{supp } \mu\}$ and $b := \sup\{x \in \mathbb{R} : x \in \text{supp } \mu\}$. As in the case of shortfall risk, we may restrict our attention to the constraint set

$$\mathcal{C} := \{\nu \in \mathcal{M}_1([a, b]) : AVaR_\lambda(\nu) = y\}.$$

5.1. Existence of solutions

A necessary and sufficient criterion for the existence of solutions can be formulated in terms of the parameters a , b , and y . The derivation is based on Lemma 4.1.

In the case of average value at risk, the existence of a minimizer with finite relative entropy can be rephrased in terms of the parameters of the problem. For this purpose, it is useful to recall a particular representation of $AVaR_\lambda$ (see Föllmer and Schied (2004)).

Proposition 5.1. *Let $\lambda \in (0, 1)$ and $\nu \in \mathcal{M}_{1,c}$. Then*

$$AVaR_\lambda(\nu) = - \int x f_\nu(x) \nu(dx), \tag{5.1}$$

where f_ν is the following density of a probability measure with respect to ν :

$$f_\nu(x) = \frac{1}{\lambda} (\mathbf{1}_{(-\infty, q)} + \kappa \mathbf{1}_{\{q\}}). \tag{5.2}$$

Here q is a λ -quantile of ν , i.e.

$$\int \mathbf{1}_{(-\infty, q)} d\nu \leq \lambda, \quad \int \mathbf{1}_{(-\infty, q]} d\nu \geq \lambda,$$

and the parameter κ is defined as follows:

$$\kappa = \begin{cases} 0 & \text{if } \nu\{q\} = 0, \\ \frac{\lambda - \nu((-\infty, q))}{\nu(\{q\})} & \text{if } \nu\{q\} \neq 0. \end{cases}$$

The following proposition characterizes the existence of solutions.

Proposition 5.2. *The following conditions are equivalent.*

1. *There exists a $\nu \in \mathcal{C}$ such that $H(\nu|\mu) < \infty$.*
2. *The minimal value of the relative entropy on \mathcal{C} is finite, and is attained for some element of \mathcal{C} .*
3. *Either $a < -y < b$ or $-y$ is an atom of μ .*

Proof. See Appendix A.

5.2. Structure of the solutions

Classical results of Csiszár (1975) determine the general structure of the minimizer. We compute the solution explicitly. In order to avoid trivial cases, we will always assume that one and, thus, all of the equivalent conditions of Proposition 5.2 are satisfied. We distinguish two cases of different complexities: in case (A) μ does not have any atoms, while in case (B) μ possibly does have atoms.

5.2.1. *A two-step procedure. I.* First we focus on case (A). We can restrict our attention to probability measures which are absolutely continuous with respect to μ . A minimizer ν will thus not have any atoms. The formulae characterizing the density f_ν in (5.1) simplify to $f_\nu = 1/\lambda \mathbf{1}_{(-\infty, q)}$, and $\int \mathbf{1}_{(-\infty, q)}(x) \nu(dx) = \lambda$.

The original problem can be reduced to a family of relative entropy minimization problems under linear constraints and a one-dimensional minimization problem.

Step 1. Fix some quantile level $q \in \mathbb{R}$. Minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures $\nu \ll \mu$ which satisfy the constraints

$$-\frac{1}{\lambda} \int x \mathbf{1}_{(-\infty, q)}(x) \nu(dx) = y, \tag{5.3}$$

$$\int \mathbf{1}_{(-\infty, q)}(x) \nu(dx) = \lambda. \tag{5.4}$$

We will provide conditions under which this problem has a solution. The solution is then unique and can be written in terms of an exponential density.

Step 2. As we will see, if for $q \in \mathbb{R}$ the minimization problem in step 1 has a solution with finite relative entropy, then the minimizer will be unique. We denote this minimizer by ν^q . Otherwise, we set $\nu^q = \Delta$ with the convention that $H(\Delta|\mu) = \infty$. With this notation, the solution to the original problem is given by the set $\operatorname{argmin}_{\nu \in \mathcal{D}} H(\nu|\mu)$ with $\mathcal{D} = \{\nu^q : q \in \mathbb{R}\}$.

5.2.2. Entropy minimization under linear constraints. I. We fix an arbitrary reference measure $\mu \in \mathcal{M}_{1,c}$ without atoms, and let $q \in \mathbb{R}$. In this section we consider the following minimization problem: minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures which satisfy the constraints (5.3) and (5.4).

Proposition 5.3. *The following conditions are equivalent.*

1. *There exists a probability measure ν with $H(\nu|\mu) < \infty$ that satisfies constraints (5.3) and (5.4).*
2. *There exists a probability measure ν with $H(\nu|\mu) < \infty$ that is equivalent to μ and satisfies constraints (5.3) and (5.4).*
3. *Under constraints (5.3) and (5.4), there exists a unique minimizer of the relative entropy.*
4. *$a < -y < q < b$ and $\mu(-y, q) > 0$.*

Proof. See Appendix A.

If one and, thus, all of the equivalent conditions of Proposition 5.3 are satisfied, then the unique minimizer can be characterized. The proof follows directly from Theorem 4.1.

Corollary 5.1. *Assume that one of the equivalent conditions of Proposition 5.3 holds. The measure ν is the unique minimizer of the relative entropy under constraints (5.3) and (5.4) if and only if its μ -density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \exp((h_1 + h_2x) \mathbf{1}_{(-\infty, q)}(x)).$$

Here $c > 0$ is a normalizing constant, $h_1, h_2 \in \mathbb{R}$, and the following conditions need to be satisfied:

$$\begin{aligned} -\lambda y &= c \int x \exp(h_1 + h_2x) \mathbf{1}_{(-\infty, q)}(x) \mu(dx), \\ \lambda &= c \int \exp(h_1 + h_2x) \mathbf{1}_{(-\infty, q)}(x) \mu(dx), \\ 1 &= c \int \exp((h_1 + h_2x) \mathbf{1}_{(-\infty, q)}(x)) \mu(dx). \end{aligned}$$

If ν is the minimizing density characterized in Corollary 5.1, then the minimal relative entropy is given by the expression $H(\nu|\mu) = \log c + \lambda(h_1 - h_2y)$.

5.2.3. The original problem. I. Assuming that one and, thus, all of the equivalent conditions of Proposition 5.3 are satisfied, we denote the unique solution to the minimization problem under the linear constraints (5.3) and (5.4) by ν^q . The proof of the following proposition is immediate.

Proposition 5.4. *There exists a solution to the problem of minimizing the relative entropy on \mathcal{C} if and only if $a < -y < b$. The collection of all solutions is given by $\{v^q : q \in Q^*\}$, where*

$$Q^* = \operatorname{argmin}\{H(v^q|\mu) : a < -y < q < b, \mu(-y, q) > 0\}.$$

5.2.4. *A two-step procedure. II.* We now consider case (B) in which μ may have atoms. As in case (A), the problem can be decomposed into two subproblems, but the presence of atoms makes the problem more complicated.

Step 1. Fix some quantile level $q \in \mathbb{R}$. We distinguish two subcases:

- (a) q is not an atom of the reference measure μ ;
- (b) q is an atom of μ .

Case (a) is only slightly more complicated than the situation which we considered in Proposition 5.3. Case (b) involves two additional parameters. Let $d \in [0, \lambda]$ and $u \in [0, 1 - \lambda]$. If $d = u = 0$ then we set $d(d + u)^{-1} = 0$. In step 1 we need to minimize $v \mapsto H(v|\mu)$ over all probability measures $v \ll \mu$ which satisfy the constraints

$$\lambda - d = \int \mathbf{1}_{(-\infty, q)}(x)v(dx), \tag{5.5}$$

$$\lambda + u = \int \mathbf{1}_{(-\infty, q]}(x)v(dx), \tag{5.6}$$

$$y = -\frac{1}{\lambda} \int x \left(\mathbf{1}_{(-\infty, q)}(x) + \frac{d}{d + u} \mathbf{1}_{\{q\}}(x) \right) v(dx). \tag{5.7}$$

We will provide conditions under which the problems have a solution. The solution is then unique and can again be represented by a density which is of exponential form outside the set where it vanishes. The solution will not always be equivalent to the reference measure μ .

Step 2. As we will show, if the minimization problems (a) and (b) in step 1 have a solution with finite relative entropy for fixed parameters q, u , and d , then the minimizer will be unique. Let $q \in \mathbb{R}, d \in [0, \lambda]$, and $u \in [0, 1 - \lambda]$. If q is not an atom of μ , then we are in the situation of case (a). If the minimizer exists and $u = d = 0$, we denote it by $v^{q,d,u} = v^{q,0,0}$. If q is an atom then we consider case (b). If the minimizer exists, we denote it by $v^{q,d,u}$. In all other cases, we set $v^{q,d,u} = \Delta$ with the convention that $H(\Delta|\mu) = \infty$. With this notation, the solution to the original problem is given by the set of minimizers

$$\operatorname{argmin}_{v \in \mathcal{D}} H(v|\mu), \quad \mathcal{D} = \{v^{q,d,u} : q \in \mathbb{R}, d \in [0, \lambda], u \in [0, 1 - \lambda]\}. \tag{5.8}$$

5.2.5. *Entropy minimization under linear constraints. II.* First consider case (a). We fix an arbitrary reference measure $\mu \in \mathcal{M}_{1,c}$ and a quantile level $q \in \mathbb{R}$. We assume that q is *not* an atom of μ ; nevertheless, the measure μ may have atoms. In this section we consider the following minimization problem: minimize $v \mapsto H(v|\mu)$ over all probability measures which satisfy constraints (5.3) and (5.4).

Proposition 5.5. *The following conditions are equivalent.*

1. *There exists a probability measure v with $H(v|\mu) < \infty$ that satisfies constraints (5.3) and (5.4).*
2. *Under constraints (5.3) and (5.4), there exists a unique minimizer of the relative entropy.*

3. One of the following conditions holds:

- (i) $a < -y < q < b$ and $\mu(-y, q) > 0$;
- (ii) $a \leq -y < q < b$ and $-y$ is an atom of μ .

Moreover, if condition 3(i) holds then there exists a probability measure ν with $H(\nu|\mu) < \infty$ that is equivalent to μ and satisfies constraints (5.3) and (5.4).

Proof. See Appendix A.

Definition 5.1. For $i = 1, 2, \dots, I$ with $I \in \mathbb{N}$, let $f_i: \mathbb{R} \rightarrow \mathbb{R}$ be measurable functions and let $a_i \in \mathbb{R}$. Let $\mu \in \mathcal{M}_1(\mathbb{R})$ and define the constraint set

$$\hat{\mathcal{C}} = \left\{ \nu \in \mathcal{M}_1(\mathbb{R}) : \int f_i(x)\nu(dx) = a_i, \quad i = 1, 2, \dots, I \right\}.$$

By $\bar{\mathcal{C}}$ we denote the subset of elements $\nu \in \hat{\mathcal{C}}$ with $H(\nu|\mu) < \infty$. A measurable set $N \subseteq \mathbb{R}$ is called a *maximal common nullset* if

1. $\nu(N) = 0$ for all $\nu \in \bar{\mathcal{C}}$ and
2. there exists a $\nu \in \bar{\mathcal{C}}$ such that $\mathbf{1}_N + d\nu/d\mu > 0$ μ -a.s.

Remark 5.1. Maximal common nullsets always exist if $\bar{\mathcal{C}} \neq \emptyset$. This can be demonstrated as, for example, an application of Zorn’s lemma.

Remark 5.2. A maximal common nullset is indeed a maximal set in the following sense. Let $M \subseteq \mathbb{R}$ be a measurable set that satisfies condition 1 of Definition 5.1, and assume that $N \subseteq M$ is a maximal common nullset. Then M is also a maximal common nullset, and $\mu(M \setminus N) = 0$. For a proof, see Weber (2004).

Remark 5.3. Elements of $\bar{\mathcal{C}}$ that are μ -equivalent exist under the following equivalent conditions (for a proof, see Weber (2004)).

- (i) There exists a $\nu \in \bar{\mathcal{C}}$ with $\nu \approx \mu$.
- (ii) Some maximal common nullset is a μ -nullset.
- (iii) Any maximal common nullset is a μ -nullset.
- (iv) Any μ -nullset is a maximal common nullset.
- (v) Maximal common nullsets and μ -nullsets coincide.
- (vi) The empty set is a maximal common nullset.

In the context of the minimization problem of the current section, maximal common nullsets can be characterized in terms of the parameters of the problem. If condition 3(i) of Proposition 5.5 holds, then maximal common nullset are μ -nullsets. The next proposition investigates the properties maximal common nullsets have when condition 3(i) is *not* satisfied but condition 3(ii) holds.

Proposition 5.6. *Assume that condition 3(i) of Proposition 5.5 holds.*

1. *If $a = -y$ and $\mu(-y, q) = 0$ then any maximal common nullset is a μ -nullset. That is, there exists a μ -equivalent probability measure ν with $H(\nu|\mu) < \infty$ that satisfies constraints (5.3) and (5.4).*
2. *If $a = -y$ and $\mu(-y, q) > 0$ then (a, q) is a maximal common nullset.*
3. *If $a < -y$ and $\mu(-y, q) = 0$ then $[a, -y)$ is a maximal common nullset.*

Proof. See Appendix A.

The minimizers are characterized by the following result of Csiszár (1975).

Theorem 5.1. *For $i = 1, 2, \dots, I$ with $I \in \mathbb{N}$, let $f_i: \mathbb{R} \rightarrow \mathbb{R}$ be measurable functions and let $a_i \in \mathbb{R}$. Let $\mu \in \mathcal{M}_1(\mathbb{R})$ and define the constraint set*

$$\hat{\mathcal{C}} = \left\{ \nu \in \mathcal{M}_1(\mathbb{R}) : \int f_i(x)\nu(dx) = a_i, \quad i = 1, 2, \dots, I \right\}.$$

Assume that there exists a $\nu \in \hat{\mathcal{C}}$ with $H(\nu|\mu) < \infty$. Let N be a maximal common nullset. Then there exists a unique minimizer on $\hat{\mathcal{C}}$ with finite relative entropy. The measure ν is the minimizer if and only if its μ -density is of the following form, with normalizing constant $c > 0$ and $h_i \in \mathbb{R}$, $i = 1, 2, \dots, I$:

$$\frac{d\nu}{d\mu} = c \exp\left(\sum_{i=1}^I h_i f_i\right) \mathbf{1}_{N^c}.$$

Corollary 5.2. *Assume that one and, thus, all of the equivalent conditions of Proposition 5.5 hold. Let N be a maximal common nullset (see Propositions 5.5 and 5.6). The measure ν is the unique minimizer of the relative entropy under constraints (5.3) and (5.4) if and only if its μ -density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \exp((h_1 + h_2x) \mathbf{1}_{(-\infty, q)}(x)) \mathbf{1}_{N^c}(x).$$

Here $c > 0$ is a normalizing constant, $h_1, h_2 \in \mathbb{R}$, and the following conditions need to be satisfied:

$$-\lambda y = c \int x \exp(h_1 + h_2x) \mathbf{1}_{(-\infty, q)}(x) \mathbf{1}_{N^c}(x) \mu(dx), \tag{5.9}$$

$$\lambda = c \int \exp(h_1 + h_2x) \mathbf{1}_{(-\infty, q)}(x) \mathbf{1}_{N^c}(x) \mu(dx), \tag{5.10}$$

$$1 = c \int \exp((h_1 + h_2x) \mathbf{1}_{(-\infty, q)}(x)) \mathbf{1}_{N^c}(x) \mu(dx). \tag{5.11}$$

Proof. The proof follows directly from Theorem 5.1. Condition (5.11) is a normalization, and conditions (5.9) and (5.10) are required by the constraint.

If ν is the minimizing density characterized in Corollary 5.2, then the minimal relative entropy is given by the expression $H(\nu|\mu) = \log c + \lambda(h_1 - h_2y)$.

Now consider case (b). We fix an arbitrary reference measure $\mu \in \mathcal{M}_{1,c}$ and parameters $q \in \mathbb{R}$, $d \in [0, \lambda]$, and $u \in [0, 1 - \lambda]$, and assume that q now is an atom of μ . In this section we consider the following minimization problem: minimize $\nu \mapsto H(\nu|\mu)$ over all probability measures which satisfy the constraints (5.5), (5.6), and (5.7).

Proposition 5.7. *The following conditions are equivalent.*

1. *There exists a probability measure ν with $H(\nu|\mu) < \infty$ that satisfies constraints (5.5), (5.6), and (5.7).*
2. *Under constraints (5.5), (5.6), and (5.7), there exists a unique minimizer of the relative entropy.*
3. *$a \leq -y \leq q \leq b$ and one of the following conditions holds:*
 - (i) *$d = 0$, $-y < q$, and $-y$ is an atom of μ ;*
 - (ii) *$d = 0$, $a < -y < q$, and $\mu(-y, q) > 0$;*
 - (iii) *$d > 0$, $a < -y$, and, with $\bar{a} := \sup\{x \in \text{supp } \mu : x < q\}$,*

$$-y > \frac{\lambda - d}{\lambda}a + \frac{d}{\lambda}q, \quad -y < \frac{\lambda - d}{\lambda}\bar{a} + \frac{d}{\lambda}q;$$

- (iv) *$d > 0$ and, for some atom $r \in \mathbb{R}$ of μ ,*

$$-y = \frac{\lambda - d}{\lambda}r + \frac{d}{\lambda}q.$$

Moreover, if condition 3(ii) or 3(iii) holds, then there exists a probability measure ν with $H(\nu|\mu) < \infty$ that is equivalent to μ and satisfies constraints (5.5), (5.6), and (5.7).

Proof. See Appendix A.

The next proposition investigates the properties maximal common nullsets have when condition 3 of Proposition 5.7 is satisfied. This characterization, together with Theorem 5.1, allows us to specify the solution to the minimization problem under constraints (5.5), (5.6), and (5.7).

Proposition 5.8. *Assume that $a \leq -y \leq q \leq b$ holds.*

1. *Suppose that $d = 0$ and that $-y = a < q$ is an atom of μ . Then $(-y, q)$ is a maximal common nullset.*
2. *Suppose that $d = 0$ and that $a < -y < q$. If $\mu(-y, q) = 0$ and $-y$ is an atom of μ , then $[a, -y)$ is a maximal common nullset. If $\mu(-y, q) > 0$ then the empty set is a maximal common nullset.*
3. *If condition 3(iii) of Proposition 5.7 holds, then the empty set is a maximal common nullset.*
4. *Suppose that condition 3(iv) of Proposition 5.7 holds.*
 - (i) *If $\lambda = d$ then $(-\infty, q)$ is a maximal common nullset.*
 - (ii) *If $\lambda \neq d$ and $r = a$ then (a, q) is a maximal common nullset.*
 - (iii) *If $\lambda \neq d$, $r > a$, and $\mu(r, q) = 0$, then $[a, r)$ is a maximal common nullset.*

Proof. See Appendix A.

Remark 5.4. Proposition 5.8 covers all cases which are considered in Proposition 5.7. The proof of this is left to the reader.

As a corollary of Proposition 5.8 and Theorem 5.1, we finally obtain a characterization of the solution.

Corollary 5.3. *Assume that one of the equivalent conditions of Proposition 5.7 holds. Let N be a maximal common nullset (see Proposition 5.8). The measure ν is the unique minimizer of the relative entropy under constraints (5.5), (5.6), and (5.7) if and only if its μ -density is of the following form:*

$$\frac{d\nu}{d\mu}(x) = c \exp\left((h_1 + h_2 + h_3x) \mathbf{1}_{(-\infty, q)}(x) + \left(h_2 + h_3 \frac{d}{d+u} x \right) \mathbf{1}_{\{q\}}(x) \right) \mathbf{1}_{N^c}(x).$$

Here $c > 0$ is a normalizing constant, $h_1, h_2, h_3 \in \mathbb{R}$, and conditions (5.5), (5.6), and (5.7) must be satisfied.

In particular, the minimizer ν is equivalent to μ if and only if the empty set is a maximal common nullset.

Proof. The proof follows directly from Theorem 5.1.

If ν is the minimizing density characterized in Corollary 5.3, then the minimal relative entropy is given by the expression $H(\nu|\mu) = \log c + h_1(\lambda - d) + h_2(\lambda + u) - h_3\lambda y$.

5.2.6. *The original problem. II. The general case.* The solution to the entropy minimization problem on the constraint set \mathcal{C} can be obtained by minimizing over the solutions under linear constraints. The proof is now immediate.

Proposition 5.9. *There exists a solution to the problem of minimizing the relative entropy on \mathcal{C} if and only if $a < -y < b$ or $-y$ is an atom of μ . The collection of all solutions is given in (5.8).*

Appendix A. Proofs for Sections 4 and 5

Proof of Proposition 4.2. Conditions 1 and 2 are clearly equivalent, by Lemma 4.1. Assume now that condition 1 is satisfied. Suppose that neither a nor b is an atom of μ . Then neither a nor b is an atom of ν . Since $\nu \in \mathcal{C}$, we have $\text{supp } \nu \subseteq [a, b]$. If $\ell(-a - y) < z$ then $\ell(-x - y) < z$ for $x \in (a, b]$. If $\ell(-a - y) = z$ then $\ell(-x - y) < z$ for $x \in (a, b]$, since $\ell^{-1}(\{z\})$ is a singleton and ℓ is increasing. Thus, if $\ell(-a - y) \leq z$ then $\int \ell(-x - y)\nu(dx) < z$, since a is not an atom of ν . By Proposition 4.1, $y \neq \rho'(\nu)$, which is a contradiction. Thus, $\ell(-a - y) > z$. Analogously, it can be shown that $\ell(-b - y) < z$. If a or b is an atom then additionally $\ell(-a - y) = z$ or $\ell(-b - y) = z$ is possible. This proves that condition 1 implies condition 3.

Conversely, assume that condition 3 holds. If a is an atom of μ and $\ell(-a - y) = z$, then define $\nu \ll \mu$ by $d\nu/d\mu = (1/\mu(\{a\})) \mathbf{1}_{\{a\}}$. Then $H(\nu|\mu) < \infty$ and $\int \ell(-x - y)\nu(dx) = \ell(-a - y) = z$, and, thus, $\rho'(\nu) = y$ by Proposition 4.1. Analogously, if b is an atom of μ and $\ell(-b - y) = z$, then $d\nu/d\mu = (1/\mu(\{b\})) \mathbf{1}_{\{b\}}$ defines $\nu \in \mathcal{C}$ with $H(\nu|\mu) < \infty$.

Finally, assume that $\ell(-b - y) < z < \ell(-a - y)$. Since $\ell^{-1}(\{z\})$ is a singleton, we obtain $\ell(-x - y) < z$ for $x > q$ and $\ell(-x - y) > z$ for $x < q$, where $q := -\ell^{-1}(\{z\}) - y \in (a, b)$. Thus,

$$u_+ := \int \mathbf{1}_{[a, q]}(x)\ell(-x - y)\mu(dx) > z, \quad u_- := \int \mathbf{1}_{\{q, b\}}(x)\ell(-x - y)\mu(dx) < z.$$

Choose an $\alpha \in (0, 1)$ such that $\alpha u_+ + (1 - \alpha)u_- = z$. Then we define $\nu \ll \mu$ by

$$\frac{d\nu}{d\mu} = \alpha \mathbf{1}_{[a,q]} + (1 - \alpha) \mathbf{1}_{(q,b]}.$$

Clearly, $H(\nu|\mu) < \infty$ and $\nu \in \mathcal{C}$ by Proposition 4.1.

Proof of Proposition 5.2. Conditions 1 and 2 are clearly equivalent, by Lemma 4.1. We will show that conditions 1 and 3 are equivalent. First we suppose that $-y$ is an atom of μ . Then we set

$$\frac{d\nu}{d\mu} = \frac{1}{\mu(\{-y\})} \mathbf{1}_{\{-y\}}.$$

In this case, $-y$ is a λ -quantile of ν and $AVaR_\lambda(\nu) = y$. Next we suppose that $-y$ is not an atom of μ and that condition 1 holds. Then $-y \leq a$, $a < -y < b$, or $-y \geq b$. Let $\nu \in \mathcal{C}$ with $H(\nu|\mu) < \infty$. In particular, $\nu \ll \mu$. Since $\text{supp } \nu \subseteq [a, b]$, it follows that $-y = -AVaR_\lambda(\nu) \in [a, b]$. If $a = -y = -AVaR_\lambda(\nu)$ then a must be an atom of ν . Since $\nu \ll \mu$, a is then also an atom of μ , which is a contradiction. Analogously, it can be shown that $b \neq -y$. We therefore obtain $a < -y < b$. Finally, we have to show that for $a < -y < b$ there always exists a $\nu \in \mathcal{C}$ such that $H(\nu|\mu) < \infty$. We consider the following two cases.

- (i) There exists a $q \in \mathbb{R}$ with $a < -y < q < b$ such that $\mu(-y, q) > 0$. Since μ has at most countably many atoms, we may and will assume that q is not an atom of μ .
- (ii) There exists no such $q \in \mathbb{R}$. This implies that $\mu(-y, b) = 0$. Then b must be an atom of μ , since $b = \sup\{x \in \mathbb{R} : x \in \text{supp } \mu\}$.

We first consider case (i). Since $a < -y < q$, there exists an $\alpha' \in (0, 1)$ such that

$$-y = \alpha' \frac{1}{\mu([a, -y])} \int x \mathbf{1}_{[a,-y]}(x) \mu(dx) + (1 - \alpha') \frac{1}{\mu([-y, q])} \int x \mathbf{1}_{[-y,q]}(x) \mu(dx). \tag{A.1}$$

Define the weights

$$\alpha = \frac{\lambda \alpha'}{\mu([a, -y])}, \quad \beta = \frac{\lambda(1 - \alpha')}{\mu([-y, q])}, \quad \gamma = \frac{1 - \lambda}{\mu([q, b])}. \tag{A.2}$$

We define a probability measure by

$$\frac{d\nu}{d\mu} = \alpha \mathbf{1}_{[a,-y]} + \beta \mathbf{1}_{[-y,q]} + \gamma \mathbf{1}_{[q,b]}. \tag{A.3}$$

Then $H(\nu|\mu) < \infty$. We show that $\nu \in \mathcal{C}$. First, by direct calculation we find that $\nu((-\infty, q)) = \nu((-\infty, q]) = \lambda$. Thus, q is a λ -quantile of ν . Second,

$$\begin{aligned} AVaR_\lambda(\nu) &= -\frac{1}{\lambda} \int x \mathbf{1}_{[a,q]}(x) \nu(dx) \\ &= -\left(\frac{\alpha'}{\mu([a, -y])} \int x \mathbf{1}_{[a,-y]}(x) \mu(dx) + \frac{1 - \alpha'}{\mu([-y, q])} \int x \mathbf{1}_{[-y,q]}(x) \mu(dx) \right) \\ &= y. \end{aligned}$$

Next we consider case (b). In this case, we let

$$\bar{a} := \frac{1}{\mu((-\infty, -y])} \int x \mathbf{1}_{(-\infty, -y]} \mu(dx) < -y.$$

Let $\gamma := \lambda(b + y)/(b - \bar{a}) \in (0, \lambda)$. We define a probability measure ν via its density,

$$\frac{d\nu}{d\mu} = \frac{\gamma}{\mu([a, -y])} \mathbf{1}_{[a,b]} + \frac{1 - \gamma}{\mu(\{b\})} \mathbf{1}_{\{b\}}.$$

Then $H(\nu|\mu) < \infty$. We now verify that $\nu \in \mathcal{C}$. Observe that

$$\int \mathbf{1}_{(-\infty,b)}(x)\nu(dx) = \frac{\gamma}{\mu([a, -y])} \int \mathbf{1}_{(-\infty,-y]}(x)\mu(dx) = \gamma < \lambda.$$

This implies that b is a λ -quantile of ν . We can now calculate $AVaR_\lambda$ using (5.2) with $\kappa = (\lambda - \gamma)/\nu(\{b\})$. We find that

$$\begin{aligned} AVaR_\lambda(\nu) &= \frac{-\gamma}{\lambda\mu([a, -y])} \int x \mathbf{1}_{(-\infty,b)}(x)\mu(dx) - \frac{(\lambda - \gamma)b}{\lambda} \\ &= \frac{-\bar{a}\gamma}{\lambda} - \frac{(\lambda - \gamma)b}{\lambda} = \frac{(b - \bar{a})\gamma}{\lambda} - b = y. \end{aligned}$$

Proof of Proposition 5.3. Condition 3 trivially implies condition 1. In order to show that condition 1 implies condition 3, observe that the constraint set defined by (5.3) and (5.4) is variation closed and convex. By Csiszár (1975, Theorem 2.1), it follows that the minimization problem has a solution with finite relative entropy. The uniqueness of the minimizer follows, since the constraint set is convex and $H(\cdot|\mu)$ is strictly convex on its essential domain. Together, these results show that conditions 1 and 3 are equivalent.

Next we show that condition 2 implies condition 1, which in turn implies condition 4, which itself implies condition 2. The first implication is clear. Assume that condition 1 holds. We show that this implies condition 4. By assumption, $\nu \ll \mu$ and ν does not have any atoms. Since $\text{supp } \nu \subseteq [a, b]$, we find that $q \in (a, b)$. Thus,

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \frac{1}{\nu([a, q])} \int x \mathbf{1}_{[a,q)}(x)\nu(dx) \in [a, q). \tag{A.4}$$

Since $-y$ is not an atom of μ , (A.4) implies that $a < -y$. Suppose moreover that $\mu((-\infty, q)) = 0$ and, thus, $\nu([-\infty, q)) = 0$. Then

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \frac{1}{\nu((a, -y))} \int x \mathbf{1}_{(a,-y)}(x)\nu(dx) < -y,$$

which is a contradiction.

Finally, we show that condition 4 implies condition 2. Define the density of ν with respect to μ by (A.3) with coefficients given by (A.1) and (A.2). This defines a measure ν which is equivalent to μ and satisfies constraints (5.3) and (5.4).

Proof of Proposition 5.5. The proof that conditions 1 and 2 are equivalent is analogous to the proof of the equivalence of conditions 1 and 3 of Proposition 5.3.

Next we show that conditions 1 and 3 here are equivalent. Assume that condition 1 holds. By assumption, $\nu \ll \mu$ and ν does not have an atom at q . Since $\text{supp } \nu \subseteq [a, b]$, we find that $q \in (a, b)$. Thus,

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty,q)}(x)\nu(dx) = \frac{1}{\nu([a, q])} \int x \mathbf{1}_{[a,q)}(x)\nu(dx) \in [a, q). \tag{A.5}$$

Suppose that $-y$ is not an atom of μ . Then (A.5) implies that $a < -y$. Suppose moreover that $\mu((-\infty, q)) = 0$ and, thus, $\nu([-\infty, q)) = 0$. Then

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty, q)}(x) \nu(dx) = \frac{1}{\nu([a, -y))} \int x \mathbf{1}_{[a, -y)}(x) \nu(dx) < -y,$$

which is a contradiction.

Finally, we show that condition 3 implies condition 1. If $-y$ is an atom of μ , then we set

$$\frac{d\nu}{d\mu} = \frac{\lambda}{\mu(\{-y\})} \mathbf{1}_{\{-y\}} + \frac{1-\lambda}{\mu((q, b])} \mathbf{1}_{(q, b]}. \tag{A.6}$$

The measure ν satisfies constraints (5.3) and (5.4) and has finite relative entropy. Nevertheless, it might not be equivalent to μ . If $-y$ is not an atom of μ , then $a < -y$ and $\mu(-y, q) > 0$. Define the density of ν with respect to μ by (A.3) with coefficients given by (A.1) and (A.2). This defines a measure ν which is equivalent to μ and satisfies constraints (5.3) and (5.4).

Proof of Proposition 5.6. In part 1, (A.6) defines a density of a μ -equivalent probability measure ν with $H(\nu|\mu) < \infty$ that satisfies the constraints. In order to verify part 2, set $N := (a, q)$. Let ν be a measure with $H(\nu|\mu) < \infty$ that satisfies the constraints. If $\nu(N) > 0$ then

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty, q)} \nu(dx) > -y, \tag{A.7}$$

which is a contradiction. Next, define a measure ν with density (A.6). As shown in the proof of Proposition 5.5, ν satisfies constraints (5.3) and (5.4), and $H(\nu|\mu) < \infty$. Moreover, μ -a.s.,

$$\mathbf{1}_N + \frac{d\nu}{d\mu} > 0.$$

The proof of part 3 is completely analogous to the proof of part 2. We simply have to set $N := [a, -y)$ and reverse the inequality in (A.7).

Proof of Proposition 5.7. The proof that conditions 1 and 2 are equivalent is analogous to the proof of the equivalence of conditions 1 and 3 of Proposition 5.3.

We now show that condition 1 implies condition 3. First consider the case in which $d = 0$. We have

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty, q)}(x) \nu(dx) = \frac{1}{\nu((-\infty, q))} \int x \mathbf{1}_{(-\infty, q)}(x) \nu(dx) < q. \tag{A.8}$$

If $-y$ is an atom of μ then condition 3(i) holds. Next, suppose that $-y$ is not an atom of μ ; thus, neither is it an atom of ν . From (A.8) it follows that $-y > a$. Suppose that $\mu((-\infty, q)) = 0$. Then $\nu((-\infty, -y)) = \nu((-\infty, q)) = \lambda$. Hence,

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty, q)}(x) \nu(dx) = \frac{1}{\nu((-\infty, -y))} \int x \mathbf{1}_{(-\infty, -y)}(x) \nu(dx) < -y,$$

which is a contradiction. This implies condition 3(ii).

Next consider the case in which $d > 0$. Then q is an atom of ν . If $\lambda = d$ then condition 3(iv) holds with $r = q$. Otherwise $\nu((-\infty, q)) > 0$ and $a < q$. From (5.7), we obtain

$$-y = \frac{\lambda - d}{\lambda} \underbrace{\frac{1}{\nu((-\infty, q))} \int x \mathbf{1}_{(-\infty, q)}(x) \nu(dx)}_{=:r} + \frac{d}{\lambda} q,$$

from which it follows that $r \in [a, q)$ and $a < -y$. If r is not an atom of μ then it is not an atom of ν and $a < r < \bar{a}$. This implies condition 3(iii). If r is an atom of μ then condition 3(iv) holds.

Finally we have to show that condition 3 implies condition 1. In all cases we will specify a density with respect to μ such that the resulting measure ν satisfies both $H(\nu|\mu) < \infty$ and constraints (5.5), (5.6), and (5.7). In case (i) choose

$$\frac{d\nu}{d\mu} = \frac{\lambda}{\mu(\{-y\})} \mathbf{1}_{\{-y\}} + \frac{1 - \lambda}{\nu([q, \infty))} \mathbf{1}_{[q, \infty)}. \tag{A.9}$$

Next consider case (ii). Clearly

$$\frac{1}{\mu([-y, q))} \int x \mathbf{1}_{[-y, q)}(x) \mu(dx) > -y.$$

Hence, there exists an $\alpha' \in (0, 1)$ which satisfies (A.1). We choose a density according to (A.2) and (A.3). As in the proof of Proposition 5.2, simple calculations show that the constraints are satisfied. Observe that the measure ν specified by (A.3) is equivalent to μ .

Assume that condition 3(iii) is satisfied. Then there exists an $r \in (a, \bar{a})$ such that

$$-y = \frac{\lambda - d}{\lambda} r + \frac{d}{\lambda} q.$$

By the definition of a and \bar{a} , it holds that $\mu([a, r)) > 0$ and $\mu([r, q)) > 0$. Moreover,

$$\begin{aligned} r &> \frac{1}{\mu([a, r))} \int x \mathbf{1}_{[a, r)}(x) \mu(dx) =: g_-, \\ r &< \frac{1}{\mu([r, q))} \int x \mathbf{1}_{[r, q)}(x) \mu(dx) =: g_+. \end{aligned}$$

Thus, there exists an $\alpha \in (0, 1)$ such that $r = \alpha g_- + (1 - \alpha) g_+$. Let

$$\frac{d\nu}{d\mu} = (\lambda - d) \left(\frac{\alpha}{\mu([a, r))} \mathbf{1}_{[a, r)} + \frac{1 - \alpha}{\mu([r, q))} \mathbf{1}_{[r, q)} \right) + \frac{d + u}{\mu(\{q\})} \mathbf{1}_{\{q\}} + \frac{1 - \lambda - u}{\mu((q, \infty))} \mathbf{1}_{(q, \infty)}.$$

Then ν satisfies constraints (5.5), (5.6), and (5.7). Observe that ν is equivalent to μ .

Finally, if condition 3(iv) holds, then the following density, for example, defines an appropriate measure ν :

$$\frac{d\nu}{d\mu} = \frac{\lambda - d}{\mu(\{r\})} \mathbf{1}_{\{r\}} + \frac{d + u}{\mu(\{q\})} \mathbf{1}_{\{q\}} + \frac{1 - \lambda - u}{\mu((q, \infty))} \mathbf{1}_{(q, \infty)}. \tag{A.10}$$

Proof of Proposition 5.8. Denote by $\bar{\mathcal{C}}$ the set of measures ν with $H(\nu|\mu) < \infty$ that satisfy constraints (5.5), (5.6), and (5.7). It follows from Proposition 5.7 that $\bar{\mathcal{C}}$ is never empty in the cases considered in the current proposition.

1. Let $\nu \in \bar{\mathcal{C}}$. If $\nu(-y, q) > 0$ then (5.7) implies that

$$-y = \frac{\nu(\{-y\})}{\nu([-y, q])}(-y) + \frac{\nu((-y, q))}{\nu([-y, q])} \frac{1}{\nu((-y, q))} \int x \mathbf{1}_{(-y, q)}(x)\nu(dx) > -y,$$

which is a contradiction. Thus, $\nu((-y, q)) = 0$.

Next, define $\nu \in \bar{\mathcal{C}}$ using density (A.9). Then, μ -a.s., $\mathbf{1}_{(-y, q)} + d\nu/d\mu > 0$.

2. We first consider the case in which $\mu((-y, q)) = 0$. Let $\nu \in \bar{\mathcal{C}}$ and assume that $\nu([a, -y]) > 0$. Then (5.7) implies that

$$-y = \frac{\nu(\{-y\})}{\nu([a, -y])}(-y) + \frac{\nu([a, -y])}{\nu([a, -y])} \frac{1}{\nu([a, -y])} \int x \mathbf{1}_{[a, -y]}(x)\nu(dx) < -y,$$

which is a contradiction. Thus, $\nu([a, -y]) = 0$.

If $\nu \in \bar{\mathcal{C}}$ is specified using density (A.9) then, μ -a.s., $\mathbf{1}_{[a, -y]} + d\nu/d\mu > 0$.

Second, we consider the case in which $\mu((-y, q)) > 0$. By Proposition 5.7, there exists a μ -equivalent measure $\nu \in \bar{\mathcal{C}}$. The claim follows from Remark 5.3.

3. This follows from Proposition 5.7 and Remark 5.3.

4. If $\lambda = d$ then $\nu((-\infty, q)) = 0$ for $\nu \in \bar{\mathcal{C}}$, by (5.5). In case (ii), assume for $\nu \in \bar{\mathcal{C}}$ that $\nu((a, q)) > 0$. From (5.7) we then obtain

$$-y = \frac{1}{\lambda} \int x \mathbf{1}_{(-\infty, q)}(x)\nu(dx) + \frac{d}{\lambda}q = \frac{\lambda - d}{\lambda} \frac{1}{\lambda - d} \int x \mathbf{1}_{(r, q)}(x)\nu(dx) + \frac{d}{\lambda}q > -y,$$

which is a contradiction. Thus, $\nu((a, q)) = 0$. In case (iii), assume for $\nu \in \bar{\mathcal{C}}$ that $\nu([a, r]) > 0$. Arguing analogously, we find that $\nu([a, r]) = 0$.

Finally, set $N := (-\infty, q)$ in case (i), $N := (a, q)$ in case (ii), and $N := [a, r)$ in case (iii). Define $\nu \in \bar{\mathcal{C}}$ according to density (A.10). Then, μ -a.s., $\mathbf{1}_N + d\nu/d\mu > 0$.

Acknowledgements

I would like to thank Boris Buchmann, Hans Föllmer, and Ulrich Horst for helpful discussions. I am very grateful for the useful comments of an anonymous referee. I acknowledge financial support from Deutsche Forschungsgemeinschaft via Graduiertenkolleg 251 ‘Stochastische Prozesse und Probabilistische Analysis’ and SFB 373 ‘Quantifikation und Simulation ökonomischer Prozesse’.

References

ALIPRANTIS, C. D. AND BORDER, K. C. (1999). *Infinite Dimensional Analysis: A Hitchhiker’s Guide*. Springer, Berlin.
 ARTZNER, P., DELBAEN, F., EBER, J.-M. AND HEATH, D. (1999). Coherent measures of risk. *Math. Finance* **9**, 203–228.
 CSISZÁR, I. (1975). *I*-divergence geometry of probability distributions and minimization problems. *Ann. Prob.* **3**, 146–158.
 DELBAEN, F. (2002). Coherent risk measures on general probability spaces. In *Advances in Finance and Stochastics*, eds K. Sandmann and P. J. Schönbucher, Springer, Berlin, pp. 1–38.
 DEMBO, A. AND ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*. Springer, New York.

- DUNKEL, J. AND WEBER, S. (2005). Efficient Monte Carlo methods for risk measures. Working paper, Cornell University. Available at <http://people.orie.cornell.edu/~sweber/research.htm>.
- FÖLLMER, H. AND SCHIED, A. (2002). Robust preferences and convex measures of risk. In *Advances in Finance and Stochastics*, eds K. Sandmann and P. J. Schönbucher, Springer, Berlin, pp. 39–56.
- FÖLLMER, H. AND SCHIED, A. (2004). *Stochastic Finance. An Introduction in Discrete Time*, 2nd edn. De Gruyter, Berlin.
- FRITTELLI, M. AND ROSAZZA, G. E. (2002). Putting order in risk measures. *J. Banking Finance* **26**, 1473–1486.
- FU, M. C., JIN, X. AND XIONG, X. (2003). Probabilistic error bounds for simulation quantile estimators. *Manag. Sci.* **49**, 230–246.
- GIESECKE, K., SCHMIDT, T. AND WEBER, S. (2005). Measuring the risk of extreme events. Working paper, Cornell University. Available at <http://people.orie.cornell.edu/~sweber/research.htm>.
- JOUINI, E., SCHACHERMAYER, W. AND TOUZI, N. (2006). Law invariant risk measures have the Fatou property. *Adv. Math. Econom.* **9**, 49–71.
- KALLENBERG, O. (1997). *Foundations of Modern Probability*. Springer, New York.
- KUSUOKA, S. (2001). On law invariant coherent risk measures. *Adv. Math. Econom.* **3**, 83–95.
- SCHMEIDLER, D. (1986). Integral representation without additivity. *Proc. Amer. Math. Soc.* **97**, 255–261.
- WEBER, S. (2004). Measures and models of financial risk. Doctoral Thesis, Humboldt-Universität zu Berlin.
- WEBER, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Math. Finance* **16**, 419–442.