

Large deviations for risk measures in finite mixture models*

Valeria Bignozzi[†]

Claudio Macci[‡]

Lea Petrella[§]

Abstract

Due to their heterogeneity, insurance risks can be properly described as a mixture of different fixed models, where the weights assigned to each model may be estimated empirically from a sample of available data. If a risk measure is evaluated on the estimated mixture instead of the (unknown) true one, then it is important to investigate the committed error. In this paper we study the asymptotic behaviour of estimated risk measures, as the data sample size tends to infinity, in the fashion of large deviations. We obtain large deviation results by applying the contraction principle, and the rate functions are given by a suitable variational formula; explicit expressions are available for mixtures of two models. Finally, our results are applied to the most common risk measures, namely the quantiles, the Expected Shortfall and the shortfall risk measure.

AMS Subject Classification. Primary: 60F10, 91B30. Secondary: 62B10, 62D05.

Keywords: contraction principle, Lagrange multipliers, quantile, entropic risk measure, relative entropy.

1 Introduction

Quantitative risk management for financial and insurance companies requires the modelling of financial positions in terms of random variables on a suitable probability space; in mathematical terms, this corresponds to identifying a probability law (model) μ on the real line that describes as accurately as possible, the random behaviour of the position. Model risk, that arises from the uncertainty about the model to adopt, has been largely discussed in various area of the literature, because it may impact substantially companies decision making and performance. We can distinguish three main approaches to deal with model uncertainty: 1) the model is not specified but directly extrapolated from data via the empirical distribution; 2) a model is selected and its parameters are estimated from data (e.g. using Maximum Likelihood Estimation); 3) a class of candidate models is considered (for instance models suggested by expert opinion) and then one or an average of them is applied. The latter approach is probably the most common one and includes for instance: the worst-case approach proposed by Gilboa and Schmeidler (1989) in the theory of utility maximization, where the chosen model is the one providing the most adverse outcome; the Bayesian model averaging approach, developed by Raftery et al. (1997) where (posterior) weights are calculated for each model considering both information arising from data and prior beliefs; the highest posterior approach, where the selected model is the one most favourable according to the posterior weights. Cairns (2000) provided a general framework for dealing with model and

*The support of Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM) is acknowledged.

[†]Dipartimento di Statistica e Metodi Quantitativi, Università di Milano Bicocca, Via Bicocca degli Arcimboldi 8, I-20126 Milano, Italia. e-mail: valeria.bignozzi@unimib.it

[‡]Dipartimento di Matematica, Università di Roma Tor Vergata, Via della Ricerca Scientifica, I-00133 Roma, Italia. e-mail: macci@mat.uniroma2.it

[§]Dipartimento di Metodi e Modelli per l'Economia, il Territorio e la Finanza, Sapienza Università di Roma, Via del Castro Laurenziano 9, I-00161 Roma, Italia. e-mail: lea.petrella@uniroma1.it

parameter uncertainty in an insurance framework, while Pesaran et al. (2009) considered a model averaging approach in a risk management context. As we will see later, in this contribution we consider an average of fixed models where the weights are estimated empirically.

A second fundamental step for internal and external risk management purposes is to quantify the riskiness of the company positions. Once the model has been chosen, this essentially corresponds to applying a suitable risk measure ρ to the financial position. The impact of model uncertainty on risk measurement was discussed among others by Barrieu and Scandolo (2015) and Bignozzi and Tsanakas (2016) where different measures of model risk are considered.

Most of the risk measures generally considered, from both academics and practitioners, are *law-invariant* that is, univocally determined by the probability law of the random variable. These risk measures can then be treated as statistical functionals.

While the mathematical theory of risk measures is by now well developed, we refer for instance to Föllmer and Schied (2016) for an extensive treatment of coherent and convex risk measures, research on the statistical properties of risk measures is fairly recent. The seminal paper by Cont et al. (2010) started a new strand in the literature that investigates the statistical properties of risk measures in terms of robustness with respect to available data and to different model estimation procedures. The main difference between the mathematical and the statistical approaches is that, in the first case risk measures are defined on a space of random variables, while in the second one, on a space of probability measures. Although, under weak technical assumptions, for a random variable X with probability law μ , we can identify $\rho(X)$ and $\rho(\mu)$, it is important to emphasise that properties of risk measures on random variables and on distributions are different. In particular, given two random variables X, Y with distributions μ, ν the convex combination

$$pX + (1 - p)Y,$$

for $p \in (0, 1)$, corresponds to a diversification of the portfolio, therefore it is reasonable to require a risk measure to be convex with respect to random variables. On the other side, the mixture distribution

$$p\mu + (1 - p)\nu$$

represents a higher risk profile and thus a risk measure should not be convex with respect to mixtures of distributions. Properties of the risk measures with respect to mixture distributions, have been investigated by Acciaio and Svindland (2013). Weber (2006) used such properties to characterise dynamic risk measures, while Ziegel (2016), Bellini and Bignozzi (2015) and Delbaen et al. (2016) used them to study elicitable functionals. Bernardi et al. (2017) presented some results on risk measures evaluated on mixtures of Gaussian and Student t distributions.

In this contribution we consider risk measures applied to the mixture distribution

$$\pi_1\mu_1 + \dots + \pi_s\mu_s,$$

where $\{\mu_1, \dots, \mu_s\}$ is a set of s available models, and $\pi_1, \dots, \pi_s \geq 0$ (with $\sum_{j=1}^s \pi_j = 1$) are the weights assigned to each model. Mixture models are particularly relevant when a single model is not sufficient to fully describe the data. They represent a flexible approach for modelling heterogeneous data and to carry out cluster analysis. Further, mixture models represent a ductile way to model unknown distributional shapes. Such situations are quite common in insurance where often a mix of small, medium and large size claims occurs; we refer the interested reader to Klugman et al. (2012) for a full treatment of loss modelling in actuarial science. Bernardi et al. (2012) proposed finite mixtures of Skew Normal distributions to properly characterise insurance data, while Lee and Lin (2010) suggested a mixture of Erlang distributions. In a statistical framework mixture models have a variety of applications; we refer to McLachlan and Peel (2004) for an extensive treatment of the topic.

Throughout this paper the models μ_1, \dots, μ_s are assumed to be fixed, and the weights π_1, \dots, π_s are estimated empirically from independent samples. In an insurance framework, we can assume that each model represents the loss profile of a customer (or a class of customers) and the weights are estimated registering the relative frequency of claims occurring for each model. Then we consider the sequence of *empirical risk measures* $\left\{ \rho\left(\sum_{j=1}^s \hat{\pi}_n(j) \mu_j\right) : n \geq 1 \right\}$, where the weight estimators $\hat{\pi}_n(1), \dots, \hat{\pi}_n(s)$ concern the empirical law of i.i.d. random variables $\{X_1, \dots, X_n\}$ with distribution $\pi = (\pi_1, \dots, \pi_s)$ (see (1) below).

In this paper we prove large deviation results for the empirical risk measures. The theory of large deviations gives an asymptotic computation of small probabilities on an exponential scale (see e.g. Dembo and Zeitouni, 1998 as a reference on this topic). The large deviation principles are obtained by applying the contraction principle; so the rate functions are given by a suitable variational formula. We use the method of the Lagrange multipliers, and explicit expressions are available for $s = 2$. We then apply our results to the most common risk measures, namely the quantiles (also known as Value-at-Risk in the risk management literature), the Expected Shortfall (ES) and the shortfall risk measure.

A different approach for large deviation analysis may be the use of precise large deviation techniques which are beyond the purpose of the paper; among others, a possible reference for the interested reader is Féray et al. (2016).

Our work was inspired by Weber (2007), where the author considered the empirical risk measures $\{\rho(\hat{\mu}_n) : n \geq 1\}$, and

$$\hat{\mu}_n := \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$$

is the empirical law of i.i.d. random variables $\{Y_1, \dots, Y_n\}$ having (unknown) distribution μ with bounded support. The main goal of that paper is to investigate coherent and convex risk measures that are continuous on compacts. This condition yields the large deviation principle of $\{\rho(\hat{\mu}_n) : n \geq 1\}$ by applying the contraction principle (see Proposition 2.1 and Corollary 2.1 in Weber, 2007).

The paper is organised as follows. Section 2 gathers some preliminaries on large deviations and their applications to our framework with finite mixtures. In Section 3 we present the main results of the paper, while Section 4 presents some examples for the most common risk measures used in practice and in the literature.

2 Preliminaries

In this section we recall some preliminaries on large deviations and a large deviation principle for a sequence of estimators (see Proposition 2.1).

2.1 Preliminaries on large deviations

A sequence of random variables $\{W_n : n \geq 1\}$ taking values on a topological space \mathcal{W} satisfies the large deviation principle (LDP for short) with rate function $I : \mathcal{W} \rightarrow [0, \infty]$ if I is a lower semi-continuous function,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P(W_n \in O) \geq - \inf_{w \in O} I(w) \text{ for all open sets } O$$

and

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P(W_n \in C) \leq - \inf_{w \in C} I(w) \text{ for all closed sets } C.$$

A rate function I is said to be good if all its level sets $\{\{w \in \mathcal{W} : I(w) \leq \eta\} : \eta \geq 0\}$ are compact. Finally we also recall the contraction principle (see e.g. Theorem 4.2.1 in Dembo and Zeitouni, 1998): let \mathcal{Y} be a topological space, and let $f : \mathcal{W} \rightarrow \mathcal{Y}$ be a continuous function; then, if $\{W_n : n \geq 1\}$ satisfies the LDP with good rate function I , and $Y_n := f(W_n)$ (for all $n \geq 1$), $\{Y_n : n \geq 1\}$ satisfies the LDP with good rate function J defined by

$$J(y) := \inf\{I(w) : w \in \mathcal{W}, f(w) = y\}.$$

The LDP for real valued random variables is used to obtain asymptotic evaluations for the logarithm of tail probabilities; indeed, for a wide class of cases, we have

$$\log P(W_n > x) \sim -nI(x)$$

at least for x large enough to have $I(x) = \inf_{w>x} I(w)$ (we use the symbol \sim to mean that the ratio tends to 1 as $n \rightarrow \infty$).

2.2 LDP for estimators of $\rho(\mu)$ when μ is a mixture

We define a law-invariant risk measure as a map

$$\rho : \mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R},$$

that assigns to every probability measure $\mu \in \mathcal{P}(\mathbb{R})$ on the real line a real number $\rho(\mu)$. Such a value, is generally used to summarise the riskiness of the model μ and can be adopted to calculate solvency capital requirements. In the present contribution, we focus on probability distributions that arise as mixtures $\pi_1\mu_1 + \dots + \pi_s\mu_s$ of some fixed models μ_1, \dots, μ_s with weights $\pi = (\pi_1, \dots, \pi_s) \in \Sigma_s$ where

$$\Sigma_s := \{(p_1, \dots, p_s) : p_1, \dots, p_s \geq 0, p_1 + \dots + p_s = 1\}$$

is the simplex; we are then interested in computing $\rho\left(\sum_{j=1}^s \pi_j \mu_j\right)$. In mixture models used for modeling insurance data, it is often the case that the weights π_1, \dots, π_s are unknown and estimated from a set of n available data by $\hat{\pi}_n = (\hat{\pi}_n(1), \dots, \hat{\pi}_n(s))$, see for instance Lee et al. (2012). In order to estimate the error committed in computing the estimated risk measure $\rho\left(\sum_{j=1}^s \hat{\pi}_n(j) \mu_j\right)$ instead of the correct one $\rho\left(\sum_{j=1}^s \pi(j) \mu_j\right)$, we employ the theory of large deviations. In particular, we consider the case where the weights are estimated empirically as

$$\hat{\pi}_n(j) := \frac{1}{n} \sum_{i=1}^n \delta_{X_i=j} \text{ (for all } j \in \{1, \dots, s\}), \quad (1)$$

where $\{X_1, \dots, X_n\}$ are i.i.d. random variables with distribution $\pi = (\pi_1, \dots, \pi_s)$. It is well known that the sequence of empirical measures $(\hat{\pi}_n)_n$ converges P -a.s. to π , and that it satisfies the LDP (see e.g. Theorem 2.1.10 in Dembo and Zeitouni, 1998). Therefore, by applying the contraction principle (see Theorem 4.2.1 in Dembo and Zeitouni, 1998), we obtain the LDP stated in the following proposition.

Proposition 2.1. *Let $\pi = (\pi_1, \dots, \pi_s) \in \Sigma_s$. Moreover assume that the function*

$$\Sigma_s \ni (p_1, \dots, p_s) \mapsto \rho\left(\sum_{j=1}^s p_j \mu_j\right) \quad (2)$$

is continuous. Then $\left\{ \rho \left(\sum_{j=1}^s \hat{\pi}_n(j) \mu_j \right) : n \geq 1 \right\}$ satisfies the LDP (as $n \rightarrow \infty$) with good rate function $H_{\rho, \langle \pi, \mu \rangle}$ defined by

$$H_{\rho, \langle \pi, \mu \rangle}(r) := \inf \left\{ \sum_{j=1}^s p_j \log \frac{p_j}{\pi_j} : (p_1, \dots, p_s) \in \Sigma_s, \rho \left(\sum_{j=1}^s p_j \mu_j \right) = r \right\}. \quad (3)$$

Remark 2.1 (Relative entropy and Sanov's Theorem). *The quantity $\mathbb{E}_\pi \left[\frac{dp}{d\pi} \log \frac{dp}{d\pi} \right] = \sum_{j=1}^s p_j \log \frac{p_j}{\pi_j}$ in (3) is the relative entropy of a general probability measure $p = (p_1, \dots, p_s)$ on the state space $\{1, \dots, s\}$, with respect to the probability measure $\pi = (\pi_1, \dots, \pi_s)$ that gives the actual (but unknown) weights of the mixture model. Large deviation rate functions are indeed often expressed in terms of relative entropy; see e.g. the discussion in Varadhan (2003). The rate function is thus obtained minimising the relative entropy under a constraint on the risk measure.*

Remark 2.2 (The set S_π and the value r_0). *If $\pi_j = 0$ for some $j \in \{1, \dots, s\}$, then we have*

$$p_j \log \frac{p_j}{\pi_j} = \begin{cases} 0 & \text{if } p_j = 0 \\ \infty & \text{if } p_j \in (0, 1]; \end{cases}$$

so, in some sense, the index j is negligible. Then we should consider the set $S_\pi := \{i \in \{1, \dots, s\} : \pi_i > 0\}$ instead of $\{1, \dots, s\}$; however, with a slight abuse of notation, throughout the paper we always refer to $\{1, \dots, s\}$ (and its cardinality s) because we can always rearrange the notation in order to have $S_\pi = \{1, \dots, s\}$. We also remark that $H_{\rho, \langle \pi, \mu \rangle}(r)$ uniquely vanishes at $r = r_0$, where

$$r_0 := \rho \left(\sum_{j=1}^s \pi_j \mu_j \right).$$

Thus we can say that, for every $\delta > 0$, under the hypotheses of Proposition 2.1, the probability

$$P \left(\left| \rho \left(\sum_{j=1}^s \hat{\pi}_n(j) \mu_j \right) - r_0 \right| \geq \delta \right)$$

decays as e^{-nh_δ} , where $h_\delta := \inf \{ H_{\rho, \langle \pi, \mu \rangle}(r) : |r - r_0| \geq \delta \} > 0$, as $n \rightarrow \infty$.

3 Results

In this section we provide, when possible, an explicit expression of the variational formula in (3). Note that in general the constraint $\rho \left(\sum_{j=1}^s p_j \mu_j \right) = r$ cannot be written explicitly in terms of the p_j 's; for this reason we introduce the next Condition 3.1 that requires a sort of linear dependence of the risk measures with respect to the mixture weights. As we shall see in Theorem 3.1, this allows us to handle the variational formula in (3) with the method of Lagrange multipliers. Condition 3.1 does not seem to be restrictive, it is indeed satisfied by many of the risk measures used by academics and practitioners.

Condition 3.1. *The function in (3) can be written as*

$$H_{\rho, \langle \pi, \mu \rangle}(r) := \inf \left\{ \sum_{j=1}^s p_j \log \frac{p_j}{\pi_j} : (p_1, \dots, p_s) \in \Sigma_s, \sum_{j=1}^s p_j \Psi_\rho(\mu_j, r) = 0 \right\}, \quad (4)$$

for some (strictly) decreasing functions $\Psi_\rho(\mu_1, \cdot), \dots, \Psi_\rho(\mu_s, \cdot)$. Moreover, for all $i \in \{1, \dots, s\}$, there exists a unique $r_i^{(0)}$ such that $\Psi_\rho(\mu_i, r_i^{(0)}) = 0$.

Obviously we could require that $\Psi_\rho(\mu_1, \cdot), \dots, \Psi_\rho(\mu_s, \cdot)$ are increasing functions (instead of decreasing); in such a case we can reduce to Condition 3.1 (namely the functions $\Psi_\rho(\mu_1, \cdot), \dots, \Psi_\rho(\mu_s, \cdot)$ are decreasing) by a change of sign.

We will see in Section 4 that Condition 3.1 is fulfilled by some popular risk measures, such as the quantiles, the mean and the class of convex shortfall risk measures introduced by Föllmer and Schied (2002). The Expected Shortfall satisfies this condition only under some extra requirements. A similar condition recently appeared in the literature about elicitable risk measures under the name of Convex Level Sets (CxLS). A risk measure has CxLS if, given $\rho(\mu_1) = \dots = \rho(\mu_s) = r$, then $\rho(\sum_{j=1}^s p_j \mu_j) = r$. Clearly, if $\rho(\mu_1) = \dots = \rho(\mu_s) = r$, the CxLS property implies our Condition 3.1 with

$$\Psi_\rho(\mu_j, r) := \rho(\mu_j) - r.$$

A full characterization of convex risk measures satisfying the CxLS property is provided in Delbaen et al. (2016).

Remark 3.1 (Consequences of Condition 3.1 for r_0). *If Condition 3.1 holds, then we have*

$$\sum_{j=1}^s \pi_j \Psi_\rho(\mu_j, r_0) = 0. \quad (5)$$

Moreover, if we set $\underline{r}_\rho := \min \{r_i^{(0)} : i \in \{1, \dots, s\}\}$ and $\bar{r}_\rho := \max \{r_i^{(0)} : i \in \{1, \dots, s\}\}$, we have $\underline{r}_\rho \leq \bar{r}_\rho$. Then we can distinguish two cases (see parts (i) and (ii) in the next Theorem 3.1):

- $\underline{r}_\rho = \bar{r}_\rho = \hat{r}_\rho$, which occurs if and only if $r_1^{(0)} = \dots = r_s^{(0)} = \hat{r}_\rho$; in this case we have $r_0 = \hat{r}_\rho$, and the estimators $\left\{ \rho \left(\sum_{j=1}^s \hat{\pi}_n(j) \mu_j \right) ; n \geq 1 \right\}$ are constantly equal to r_0 ;
- $\underline{r}_\rho < \bar{r}_\rho$; in this case we have $r_0 \in (\underline{r}_\rho, \bar{r}_\rho)$, and the estimators $\left\{ \rho \left(\sum_{j=1}^s \hat{\pi}_n(j) \mu_j \right) ; n \geq 1 \right\}$ take values in $[\underline{r}_\rho, \bar{r}_\rho]$.

The first case always occurs if $s = 1$.

Now we are ready to present Theorem 3.1. In general we only have an explicit expression of $H_{\rho, \langle \pi, \mu \rangle}$ for the case $s = 2$; see Remark 3.2 and Remark 3.3. The case with $s = \infty$ will be discussed in Remark 3.4.

Theorem 3.1. *Consider the same hypotheses of Proposition 2.1. Assume that Condition 3.1 holds, and let \underline{r}_ρ and \bar{r}_ρ be as in Remark 3.1.*

(i) *If $\underline{r}_\rho = \bar{r}_\rho$, then*

$$H_{\rho, \langle \pi, \mu \rangle}(r) = \begin{cases} 0 & \text{if } r = r_1^{(0)} = \dots = r_s^{(0)} \\ \infty & \text{otherwise.} \end{cases}$$

(ii) *If $\underline{r}_\rho < \bar{r}_\rho$, then*

$$H_{\rho, \langle \pi, \mu \rangle}(r) = \begin{cases} -\log \left(\sum_{j=1}^s \pi_j e^{-\lambda_*(r) \Psi_\rho(\mu_j, r)} \right) & \text{if } r \in (\underline{r}_\rho, \bar{r}_\rho) \\ -\log \sum_{j: r_j^{(0)} = \bar{r}_\rho} \pi_j & \text{if } r = \bar{r}_\rho \\ -\log \sum_{j: r_j^{(0)} = \underline{r}_\rho} \pi_j & \text{if } r = \underline{r}_\rho \\ \infty & \text{if } r \notin [\underline{r}_\rho, \bar{r}_\rho], \end{cases}$$

where $\lambda_*(r)$ is such that

$$\frac{\sum_{j=1}^s \pi_j \Psi_\rho(\mu_j, r) e^{-\lambda_*(r) \Psi_\rho(\mu_j, r)}}{\sum_{j=1}^s \pi_j e^{-\lambda_*(r) \Psi_\rho(\mu_j, r)}} = 0. \quad (6)$$

Proof. We start with the proof of the statement (i). For $r = r_1^{(0)} = \dots = r_s^{(0)}$ we have

$$H_{\rho, \langle \pi, \mu \rangle}(r) = \inf \left\{ \sum_{j=1}^s p_j \log \frac{p_j}{\pi_j} : (p_1, \dots, p_s) \in \Sigma_s \right\} = 0$$

(the infimum is attained by choosing $(p_1, \dots, p_s) = (\pi_1, \dots, \pi_s)$); on the contrary, for $r \neq r_1^{(0)} = \dots = r_s^{(0)}$, we have $H_{\rho, \langle \pi, \mu \rangle}(r) = \infty$ because the condition $\sum_{j=1}^s p_j \Psi_\rho(\mu_j, r) = 0$ fails for every choice of $(p_1, \dots, p_s) \in \Sigma_s$ (in fact the values $\{\Psi_\rho(\mu_j, r) : j \in \{1, \dots, s\}\}$ are all positive if $r < r_1^{(0)} = \dots = r_s^{(0)}$ and are all negative if $r > r_1^{(0)} = \dots = r_s^{(0)}$), and therefore we have the infimum over the empty set.

Now we concentrate the attention on the proof of the statement (ii). For $r \notin [\underline{r}_\rho, \bar{r}_\rho]$ we have the same argument of the proof of the statement (i), for the case $r \neq r_1^{(0)} = \dots = r_s^{(0)}$. For $r \in (\underline{r}_\rho, \bar{r}_\rho)$ we introduce the function

$$\mathcal{L}(p_1, \dots, p_s, \lambda) = \sum_{j=1}^s p_j \log \frac{p_j}{\pi_j} + \lambda \left(\sum_{j=1}^s p_j \Psi_\rho(\mu_j, r) \right)$$

and, by the Lagrange multipliers method, (p_1, \dots, p_s) attains the infimum in (4) if it is the solution of the system

$$\begin{cases} \log \frac{p_i}{\pi_i} + 1 + \lambda \Psi_\rho(\mu_i, r) = 0, & \text{for all } i \in \{1, \dots, s\} \\ \sum_{j=1}^s p_j \Psi_\rho(\mu_j, r) = 0; \end{cases}$$

then the minimiser $(p_1, \dots, p_s) = (p_1(r), \dots, p_s(r))$, which attains the infimum in (4), is defined by

$$p_i(r) := \frac{\pi_i e^{-\lambda_*(r) \Psi_\rho(\mu_i, r)}}{\sum_{j=1}^s \pi_j e^{-\lambda_*(r) \Psi_\rho(\mu_j, r)}} \text{ for all } i \in \{1, \dots, s\}, \quad (7)$$

where $\lambda_*(r)$ is such that (6) holds. In conclusion, for $r \in (\underline{r}_\rho, \bar{r}_\rho)$, we have

$$\begin{aligned} H_{\rho, \langle \pi, \mu \rangle}(r) &= \sum_{j=1}^s p_j(r) \log \frac{p_j(r)}{\pi_j} = \sum_{j=1}^s \frac{\pi_j e^{-\lambda_*(r) \Psi_\rho(\mu_j, r)}}{\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}} \log \frac{e^{-\lambda_*(r) \Psi_\rho(\mu_j, r)}}{\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}} \\ &= -\lambda_*(r) \underbrace{\sum_{j=1}^s \frac{\pi_j \Psi_\rho(\mu_j, r) e^{-\lambda_*(r) \Psi_\rho(\mu_j, r)}}{\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}}}_{=0 \text{ by (6)}} - \log \left(\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)} \right), \end{aligned}$$

and therefore

$$H_{\rho, \langle \pi, \mu \rangle}(r) = -\log \left(\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)} \right). \quad (8)$$

Finally the cases $r \in \{\underline{r}_\rho, \bar{r}_\rho\}$. The minimisers $(p_1, \dots, p_s) = (p_1(r), \dots, p_s(r))$, which attain the infimum in (4), are

$$p_i(\bar{r}_\rho) = \begin{cases} \pi_i / (\sum_{j: r_j^{(0)} = \bar{r}_\rho} \pi_j) & \text{if } i \in \{j : r_j^{(0)} = \bar{r}_\rho\} \\ 0 & \text{if } i \notin \{j : r_j^{(0)} = \bar{r}_\rho\} \end{cases}$$

and

$$p_i(\underline{r}_\rho) = \begin{cases} \pi_i / (\sum_{j: r_j^{(0)} = \underline{r}_\rho} \pi_j) & \text{if } i \in \{j : r_j^{(0)} = \underline{r}_\rho\} \\ 0 & \text{if } i \notin \{j : r_j^{(0)} = \underline{r}_\rho\}, \end{cases}$$

respectively; thus we can easily check that

$$H_{\rho, \langle \pi, \mu \rangle}(r) = \sum_{j=1}^s p_j(r) \log \frac{p_j(r)}{\pi_j} = \begin{cases} -\log \sum_{j: r_j^{(0)} = \bar{r}_\rho} \pi_j & \text{if } r = \bar{r}_\rho \\ -\log \sum_{j: r_j^{(0)} = \underline{r}_\rho} \pi_j & \text{if } r = \underline{r}_\rho. \end{cases}$$

□

Remark 3.2 (The rate function $H_{\rho, \langle \pi, \mu \rangle}$ is not explicit). *Obviously Theorem 3.1 does not provide an explicit expression of the rate function because there is not an explicit expression of $\lambda_*(r)$. However $H_{\rho, \langle \pi, \mu \rangle}(r_0) = 0$; so that $(p_1(r_0), \dots, p_s(r_0)) = (\pi_1, \dots, \pi_s)$ and, by taking into account (7), we obtain $\lambda_*(r_0) = 0$. Therefore we recover the known equality $H_{\rho, \langle \pi, \mu \rangle}(r_0) = 0$ by considering (8) and $\lambda_*(r_0) = 0$. Finally we also remark that (6) and $\lambda_*(r_0) = 0$ yield (5).*

Remark 3.3 (On Theorem 3.1(ii) with $s = 2$). *We remark that, if $s = 2$, then $\underline{r}_\rho < \bar{r}_\rho$ if and only if $r_1^{(0)} \neq r_2^{(0)}$. Here we take $r \in (\underline{r}_\rho, \bar{r}_\rho) = (r_1^{(0)} \wedge r_2^{(0)}, r_1^{(0)} \vee r_2^{(0)})$. Then we have*

$$\pi_1 \Psi_\rho(\mu_1, r) e^{-\lambda_*(r) \Psi_\rho(\mu_1, r)} + \pi_2 \Psi_\rho(\mu_2, r) e^{-\lambda_*(r) \Psi_\rho(\mu_2, r)} = 0$$

by (6) (with $s = 2$), and therefore

$$e^{-\lambda_*(r) \Psi_\rho(\mu_1, r)} \left(\pi_1 \Psi_\rho(\mu_1, r) + \pi_2 \Psi_\rho(\mu_2, r) e^{-\lambda_*(r) (\Psi_\rho(\mu_2, r) - \Psi_\rho(\mu_1, r))} \right) = 0;$$

this yields

$$\pi_1 \Psi_\rho(\mu_1, r) + \pi_2 \Psi_\rho(\mu_2, r) e^{-\lambda_*(r) (\Psi_\rho(\mu_2, r) - \Psi_\rho(\mu_1, r))} = 0,$$

and we get

$$e^{-\lambda_*(r) (\Psi_\rho(\mu_2, r) - \Psi_\rho(\mu_1, r))} = -\frac{\pi_1 \Psi_\rho(\mu_1, r)}{\pi_2 \Psi_\rho(\mu_2, r)};$$

thus

$$\lambda_*(r) = -\frac{1}{\Psi_\rho(\mu_2, r) - \Psi_\rho(\mu_1, r)} \log \left(-\frac{\pi_1 \Psi_\rho(\mu_1, r)}{\pi_2 \Psi_\rho(\mu_2, r)} \right). \quad (9)$$

We remark that, for $r \in (\underline{r}_\rho, \bar{r}_\rho)$, we have

$$\Psi_\rho(\mu_2, r), \Psi_\rho(\mu_1, r) \neq 0 \text{ and } \Psi_\rho(\mu_2, r) \Psi_\rho(\mu_1, r) < 0;$$

thus, in particular, $\Psi_\rho(\mu_2, r) - \Psi_\rho(\mu_1, r) \neq 0$. Finally, by (8) (with $s = 2$) and (9), we get

$$H_{\rho, \langle \pi, \mu \rangle}(r) = -\log \left(\sum_{h=1}^2 \pi_h \left(-\frac{\pi_1 \Psi_\rho(\mu_1, r)}{\pi_2 \Psi_\rho(\mu_2, r)} \right)^{\frac{\Psi_\rho(\mu_h, r)}{\Psi_\rho(\mu_2, r) - \Psi_\rho(\mu_1, r)}} \right).$$

Remark 3.4 (On Theorem 3.1 with $s = \infty$). *It is possible to present a version of Theorem 3.1 with $s = \infty$, namely for the case of countable mixture models. The statement and the proof can be easily adapted and we omit the details. However we remark that Condition 3.1 and Remark 3.1 have to be suitably changed; in fact we should require that the quantities $\underline{r}_\rho := \min \{r_i^{(0)} : i \geq 1\}$ and $\bar{r}_\rho := \max \{r_i^{(0)} : i \geq 1\}$ are well-defined (this is not guaranteed as happens for the case $s < \infty$).*

Finally we are also interested in the local comparison between rate functions around the points where they uniquely vanish; in fact the larger $H_{\rho, \langle \pi, \mu \rangle}$ is around r_0 (except r_0), the faster is the convergence of estimators to r_0 (as $n \rightarrow \infty$). We also recall that, under suitable hypotheses which guarantee the existence of the second derivative of $H_{\rho, \langle \pi, \mu \rangle}(r)$ computed at $r = r_0$, the more $H''_{\rho, \langle \pi, \mu \rangle}(r_0)$ is large, the more $H_{\rho, \langle \pi, \mu \rangle}$ is large around r_0 (except r_0). An expression for $H''_{\rho, \langle \pi, \mu \rangle}(r_0)$ is given in the next proposition.

Proposition 3.2. *Consider the same hypotheses and notation of Theorem 3.1(ii). Moreover assume that $\Psi_\rho(\mu_1, \cdot), \dots, \Psi_\rho(\mu_s, \cdot)$ are continuously differentiable functions (at least in a neighborhood of r_0). Then, if we consider the notation $\Psi'_\rho(\mu_h, r) := \frac{d}{dr} \Psi_\rho(\mu_h, r)$, we have*

$$H''_{\rho, \langle \pi, \mu \rangle}(r_0) = \frac{(\sum_{h=1}^s \pi_h \Psi'_\rho(\mu_h, r_0))^2}{\text{Var}_\pi[\Psi_\rho(\mu, r_0)]},$$

where (the last equality holds by (5))

$$\text{Var}_\pi[\Psi_\rho(\mu, r_0)] := \sum_{h=1}^s \pi_h \Psi_\rho^2(\mu_h, r_0) - \left(\sum_{h=1}^s \pi_h \Psi_\rho(\mu_h, r_0) \right)^2 = \sum_{h=1}^s \pi_h \Psi_\rho^2(\mu_h, r_0).$$

Proof. For $r \in (r_\rho, \bar{r}_\rho)$ we have

$$\begin{aligned} H'_{\rho, \langle \pi, \mu \rangle}(r) &= \frac{\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)} [\lambda'_*(r) \Psi_\rho(\mu_h, r) + \lambda_*(r) \Psi'_\rho(\mu_h, r)]}{\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}} \\ &= \lambda'_*(r) \underbrace{\frac{\sum_{h=1}^s \pi_h \Psi_\rho(\mu_h, r) e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}}{\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}}}_{=0 \text{ by (6)}} + \lambda_*(r) \frac{\sum_{h=1}^s \pi_h \Psi'_\rho(\mu_h, r) e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}}{\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}} \\ &= \lambda_*(r) \frac{\sum_{h=1}^s \pi_h \Psi'_\rho(\mu_h, r) e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}}{\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}}; \end{aligned}$$

thus, since $\lambda_*(r_0) = 0$, we get

$$H'_{\rho, \langle \pi, \mu \rangle}(r_0) = \lambda_*(r_0) \frac{\sum_{h=1}^s \pi_h \Psi'_\rho(\mu_h, r_0) e^{-\lambda_*(r_0) \Psi_\rho(\mu_h, r_0)}}{\sum_{h=1}^s \pi_h e^{-\lambda_*(r_0) \Psi_\rho(\mu_h, r_0)}} = 0.$$

Moreover, again for $r \in (r_\rho, \bar{r}_\rho)$, we have

$$H''_{\rho, \langle \pi, \mu \rangle}(r) = \lambda'_*(r) \frac{\sum_{h=1}^s \pi_h \Psi'_\rho(\mu_h, r) e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}}{\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}} + \lambda_*(r) \frac{d}{dr} \left(\frac{\sum_{h=1}^s \pi_h \Psi'_\rho(\mu_h, r) e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}}{\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)}} \right);$$

thus, by taking into account again $\lambda_*(r_0) = 0$, we obtain

$$H''_{\rho, \langle \pi, \mu \rangle}(r_0) = \lambda'_*(r_0) \sum_{h=1}^s \pi_h \Psi'_\rho(\mu_h, r_0). \quad (10)$$

We conclude by computing $\lambda'_*(r_0)$ by means of the implicit function theorem. By (6) we consider the function

$$\Delta(r, \lambda) := \frac{\sum_{j=1}^s \pi_j \Psi_\rho(\mu_j, r) e^{-\lambda \Psi_\rho(\mu_j, r)}}{\sum_{j=1}^s \pi_j e^{-\lambda \Psi_\rho(\mu_j, r)}};$$

the partial derivatives of Δ are

$$\begin{aligned} \Delta_r(r, \lambda) &= \frac{1}{(\sum_{j=1}^s \pi_j e^{-\lambda \Psi_\rho(\mu_j, r)})^2} \cdot \left\{ \left(\sum_{j=1}^s \pi_j \Psi'_\rho(\mu_j, r) e^{-\lambda \Psi_\rho(\mu_j, r)} (1 - \lambda \Psi_\rho(\mu_j, r)) \right) \left(\sum_{j=1}^s \pi_j e^{-\lambda \Psi_\rho(\mu_j, r)} \right) \right. \\ &\quad \left. + \lambda \left(\sum_{j=1}^s \pi_j \Psi'_\rho(\mu_j, r) e^{-\lambda \Psi_\rho(\mu_j, r)} \right) \left(\sum_{j=1}^s \pi_j \Psi_\rho(\mu_j, r) e^{-\lambda \Psi_\rho(\mu_j, r)} \right) \right\} \end{aligned}$$

and

$$\Delta_\lambda(r, \lambda) = \frac{1}{\left(\sum_{j=1}^s \pi_j e^{-\lambda \Psi_\rho(\mu_j, r)}\right)^2} \cdot \left\{ - \left(\sum_{j=1}^s \pi_j \Psi_\rho^2(\mu_j, r) e^{-\lambda \Psi_\rho(\mu_j, r)} \right) \left(\sum_{j=1}^s \pi_j e^{-\lambda \Psi_\rho(\mu_j, r)} \right) + \left(\sum_{j=1}^s \pi_j \Psi_\rho(\mu_j, r) e^{-\lambda \Psi_\rho(\mu_j, r)} \right)^2 \right\};$$

thus, by taking into account (5) and $\lambda_*(r_0) = 0$, we have

$$\Delta_r(r_0, \lambda_*(r_0)) = \sum_{j=1}^s \pi_j \Psi'_\rho(\mu_j, r_0) \text{ and } \Delta_\lambda(r_0, \lambda_*(r_0)) = - \sum_{j=1}^s \pi_j \Psi_\rho^2(\mu_j, r_0),$$

and the implicit function theorem yields

$$\lambda'_*(r_0) = - \frac{\Delta_r(r, \lambda)}{\Delta_\lambda(r, \lambda)} \Big|_{(r, \lambda) = (r_0, \lambda_*(r_0))} = \frac{\sum_{j=1}^s \pi_j \Psi'_\rho(\mu_j, r_0)}{\sum_{j=1}^s \pi_j \Psi_\rho^2(\mu_j, r_0)}.$$

We conclude the proof by combining this equality and (10). \square

4 Examples

In this section we consider some examples of risk measures satisfying Condition 3.1. The first example concerns risk measures that are linearly dependent with respect to the weights, and we present two specific cases. Other examples consider quantiles and the class of shortfall risk measures, that includes the entropic risk measures as a special case. We conclude the section with two examples: one for an insurance application with $s = 2$, and one where we obtain explicit expressions for $s = 3$. In view of what follows we consider the notation F_μ for the distribution function associated with the law μ , namely

$$F_\mu(x) := \mu((-\infty, x]), \quad \text{for all } x \in \mathbb{R}.$$

We remark that, when we deal with a finite mixture $\sum_{j=1}^s p_j \mu_j$ of some laws μ_1, \dots, μ_s (for some $(p_1, \dots, p_s) \in \Sigma_s$), we have $F_{\sum_{j=1}^s p_j \mu_j} = \sum_{j=1}^s p_j F_{\mu_j}$.

Example 4.1 (Linear dependence with respect to the weights). *We assume that the function (2) satisfies the following condition:*

$$\rho \left(\sum_{j=1}^s p_j \mu_j \right) = \sum_{j=1}^s p_j \rho(\mu_j) \text{ for all } (p_1, \dots, p_s) \in \Sigma_s.$$

Obviously we have a continuous function. In this case one has

$$\Psi_\rho(\mu_i, r) = \rho(\mu_i) - r$$

which yields $r_i^{(0)} = \rho(\mu_i)$ and $r_0 = \sum_{j=1}^s \pi_j \rho(\mu_j)$; moreover

$$\underline{r}_\rho := \min\{\rho(\mu_i) : i \in \{1, \dots, s\}\} \text{ and } \bar{r}_\rho := \max\{\rho(\mu_i) : i \in \{1, \dots, s\}\}.$$

Now we present some formulas for the rate function $H_{\rho, \langle \pi, \mu \rangle}$ when $r_\rho < \bar{r}_\rho$; in view of this we remark that, for $s = 2$, we have $r_\rho < \bar{r}_\rho$ if and only if $\rho(\mu_1) \neq \rho(\mu_2)$. By (8), for $r \in (r_\rho, \bar{r}_\rho)$, we get

$$H_{\rho, \langle \pi, \mu \rangle}(r) = -\log \left(\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)} \right) = -r \lambda_*(r) - \log \left(\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \rho(\mu_h)} \right).$$

Moreover, by Remark 3.3 concerning the case $s = 2$, for $r \in (r_\rho, \bar{r}_\rho)$ we get

$$H_{\rho, \langle \pi, \mu \rangle}(r) = -\log \left(\sum_{h=1}^2 \pi_h \left(-\frac{\pi_1(\rho(\mu_1) - r)}{\pi_2(\rho(\mu_2) - r)} \right)^{\frac{\rho(\mu_h) - r}{\rho(\mu_2) - \rho(\mu_1)}} \right);$$

in particular we can easily check that this formula yields $H_{\rho, \langle \pi, \mu \rangle}(r_0) = 0$. Finally, by Proposition 3.2 (and after some computations where we take into account that $r_0 = \sum_{j=1}^s \pi_j \rho(\mu_j)$), we get

$$H''_{\rho, \langle \pi, \mu \rangle}(r_0) = \frac{1}{\sum_{h=1}^s \pi_h (\rho(\mu_h) - r_0)^2} = \frac{1}{\sum_{h=1}^s \pi_h \rho^2(\mu_h) - r_0^2}.$$

Here we briefly present two particular cases concerning Example 4.1.

- The expected value (when μ_1, \dots, μ_s are probability measures of integrable random variables); in fact we have

$$\int_{\mathbb{R}} x \sum_{j=1}^s p_j \mu_j(dx) = \sum_{j=1}^s p_j \int_{\mathbb{R}} x \mu_j(dx).$$

- The Expected Shortfall ES_α , for $\alpha \in (0, 1)$, when μ_1, \dots, μ_s have the same α -quantile, namely when $F_{\mu_1}^{-1}(\alpha) = \dots = F_{\mu_s}^{-1}(\alpha) =: r_\alpha$. We recall that

$$\text{ES}_\alpha(\mu) := \frac{1}{1 - \alpha} \int_{F_\mu^{-1}(\alpha)}^{\infty} x \mu(dx),$$

and that we have $\left(\sum_{j=1}^s p_j F_{\mu_j} \right)^{-1}(\alpha) = r_\alpha$. Therefore

$$\begin{aligned} \text{ES}_\alpha \left(\sum_{j=1}^s p_j \mu_j \right) &= \frac{1}{1 - \alpha} \int_{\left(\sum_{j=1}^s p_j F_{\mu_j} \right)^{-1}(\alpha)}^{\infty} x \sum_{j=1}^s p_j \mu_j(dx) \\ &= \sum_{j=1}^s \frac{p_j}{1 - \alpha} \int_{r_\alpha}^{\infty} x \mu_j(dx) = \sum_{j=1}^s p_j \text{ES}_\alpha(\mu_j). \end{aligned}$$

We conclude with the final examples.

Example 4.2 (Quantiles). *Let us consider $\alpha \in (0, 1)$ and strictly increasing and continuous distribution functions $F_{\mu_1}, \dots, F_{\mu_s}$ on the same interval. We assume that the function (2) is defined by*

$$\Sigma_s \ni (p_1, \dots, p_s) \mapsto \rho \left(\sum_{j=1}^s p_j \mu_j \right) := \left(\sum_{j=1}^s p_j F_{\mu_j} \right)^{-1}(\alpha).$$

This function is continuous (see Appendix for details). In this case one has

$$\Psi_\rho(\mu_i, r) = \alpha - F_{\mu_i}(r)$$

which yields $r_i^{(0)} = F_{\mu_i}^{-1}(\alpha)$ and $r_0 = \left(\sum_{j=1}^s \pi_j F_{\mu_j} \right)^{-1}(\alpha)$; moreover

$$\underline{r}_\rho := \min \{ F_{\mu_i}^{-1}(\alpha) : i \in \{1, \dots, s\} \} \text{ and } \bar{r}_\rho := \max \{ F_{\mu_i}^{-1}(\alpha) : i \in \{1, \dots, s\} \}.$$

Now we present some formulas for the rate function $H_{\rho, \langle \pi, \mu \rangle}$ when $\underline{r}_\rho < \bar{r}_\rho$; in view of this we remark that, for $s = 2$, we have $\underline{r}_\rho < \bar{r}_\rho$ if and only if $F_{\mu_1}^{-1}(\alpha) \neq F_{\mu_2}^{-1}(\alpha)$. By (8), for $r \in (\underline{r}_\rho, \bar{r}_\rho)$, we get

$$H_{\rho, \langle \pi, \mu \rangle}(r) = -\log \left(\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)} \right) = \alpha \lambda_*(r) - \log \left(\sum_{h=1}^s \pi_h e^{\lambda_*(r) F_{\mu_h}(r)} \right).$$

Moreover, by Remark 3.3 concerning the case $s = 2$, for $r \in (\underline{r}_\rho, \bar{r}_\rho)$ we get

$$H_{\rho, \langle \pi, \mu \rangle}(r) = -\log \left(\sum_{h=1}^2 \pi_h \left(-\frac{\pi_1(\alpha - F_{\mu_1}(r))}{\pi_2(\alpha - F_{\mu_2}(r))} \right)^{\frac{\alpha - F_{\mu_h}(r)}{F_{\mu_1}(r) - F_{\mu_2}(r)}} \right);$$

in particular we can easily check that this formula yields $H_{\rho, \langle \pi, \mu \rangle}(r_0) = 0$ because $\sum_{j=1}^s \pi_j F_{\mu_j}(r_0) = \alpha$. Finally, by Proposition 3.2 (and after some computations where we take into account again that $\sum_{j=1}^s \pi_j F_{\mu_j}(r_0) = \alpha$), we get

$$H''_{\rho, \langle \pi, \mu \rangle}(r_0) = \frac{(\sum_{h=1}^s \pi_h F'_{\mu_i}(r))^2}{\sum_{h=1}^s \pi_h (\alpha - F_{\mu_h}(r_0))^2} = \frac{(\sum_{h=1}^s \pi_h F'_{\mu_i}(r))^2}{\sum_{h=1}^s \pi_h F_{\mu_h}^2(r_0) - \alpha^2}.$$

Example 4.3 (Shortfall risk measures). We recall some preliminaries (see Föllmer and Schied (2002)). Given a loss function $\ell : \mathbb{R} \rightarrow \mathbb{R}$ (that is a convex, increasing and not identically constant function) and an interior point x_0 in the range of ℓ , a shortfall risk measure is defined by

$$\rho(\mu) := \inf \left\{ m \in \mathbb{R} : \int_{\mathbb{R}} \ell(x - m) \mu(dx) \leq x_0 \right\};$$

moreover it is the unique solution m to the equation

$$\int_{\mathbb{R}} \ell(x - m) \mu(dx) = x_0.$$

We assume that the function (2) is defined by

$$\Sigma_s \ni (p_1, \dots, p_s) \mapsto \rho \left(\sum_{j=1}^s p_j \mu_j \right) := r, \text{ where } \sum_{j=1}^s p_j \int_{\mathbb{R}} \ell(x - r) \mu_j(dx) = x_0.$$

The continuity of this function can be checked by adapting the proof in the Appendix. In this case one has

$$\Psi_\rho(\mu_i, r) = \int_{\mathbb{R}} \ell(x - r) \mu_i(dx) - x_0.$$

From now on, in order to have explicit results, we continue our analysis for the class of entropic risk measures, that is the case where we have the loss function $\ell(x) = e^{\theta x}$, for $\theta > 0$, and $x_0 = 1$. We can check the following equalities:

$$\rho(\mu) = \frac{1}{\theta} \log \left(\int_{\mathbb{R}} e^{\theta x} \mu(dx) \right);$$

the function (2) becomes

$$\Sigma_s \ni (p_1, \dots, p_s) \mapsto \rho \left(\sum_{j=1}^s p_j \mu_j \right) := \frac{1}{\theta} \log \int_{\mathbb{R}} e^{\theta x} \sum_{j=1}^s p_j \mu_j(dx);$$

(so we have $\Sigma_s \ni (p_1, \dots, p_s) \mapsto \frac{1}{\theta} \log \left(\sum_{j=1}^s p_j e^{\theta \rho(\mu_j)} \right)$, which is a continuous function); the function $\Psi_\rho(\mu_i, r)$ can be rewritten as

$$\Psi_\rho(\mu_i, r) = e^{\theta \rho(\mu_i)} - e^{\theta r},$$

which yields $r_i^{(0)} = \rho(\mu_i)$ and $r_0 = \frac{1}{\theta} \log \left(\sum_{j=1}^s \pi_j e^{\theta \rho(\mu_j)} \right)$; moreover

$$r_\rho := \min\{\rho(\mu_i) : i \in \{1, \dots, s\}\} \text{ and } \bar{r}_\rho := \max\{\rho(\mu_i) : i \in \{1, \dots, s\}\}.$$

Now we present some formulas for the rate function $H_{\rho, \langle \pi, \mu \rangle}$ when $r_\rho < \bar{r}_\rho$; in view of this we remark that, for $s = 2$, we have $r_\rho < \bar{r}_\rho$ if and only if $\rho(\mu_1) \neq \rho(\mu_2)$. By (8), for $r \in (r_\rho, \bar{r}_\rho)$, we get

$$H_{\rho, \langle \pi, \mu \rangle}(r) = -\log \left(\sum_{h=1}^s \pi_h e^{-\lambda_*(r) \Psi_\rho(\mu_h, r)} \right) = -e^{\theta r} \lambda_*(r) - \log \left(\sum_{h=1}^s \pi_h e^{-\lambda_*(r) e^{\theta \rho(\mu_i)}} \right).$$

Moreover, by Remark 3.3 concerning the case $s = 2$, for $r \in (r_\rho, \bar{r}_\rho)$ we get

$$H_{\rho, \langle \pi, \mu \rangle}(r) = -\log \left(\sum_{h=1}^2 \pi_h \left(-\frac{\pi_1 (e^{\theta \rho(\mu_1)} - e^{\theta r})}{\pi_2 (e^{\theta \rho(\mu_2)} - e^{\theta r})} \right)^{\frac{e^{\theta \rho(\mu_h)} - e^{\theta r}}{e^{\theta \rho(\mu_2)} - e^{\theta \rho(\mu_1)}}} \right);$$

in particular we can easily check that this formula yields $H_{\rho, \langle \pi, \mu \rangle}(r_0) = 0$ because $e^{\theta r_0} = \sum_{j=1}^s \pi_j e^{\theta \rho(\mu_j)}$. Finally, by Proposition 3.2 (and after some computations where we take into account again that $e^{\theta r_0} = \sum_{j=1}^s \pi_j e^{\theta \rho(\mu_j)}$), we get

$$H''_{\rho, \langle \pi, \mu \rangle}(r_0) = \frac{(\theta e^{\theta r_0})^2}{\sum_{h=1}^s \pi_h (e^{\theta \rho(\mu_h)} - e^{\theta r_0})^2} = \frac{(\theta e^{\theta r_0})^2}{\sum_{h=1}^s \pi_h e^{2\theta \rho(\mu_h)} - e^{2\theta r_0}}.$$

Example 4.4 (An insurance example with $s = 2$). In actuarial science mixture distributions are particularly relevant for modeling different claim sizes. Consider, for instance, a car insurance context where individuals are grouped into s categories depending on their accident history; we assume for convenience that $s = 2$. Each group claim distribution μ_j may be modeled using an exponential distribution with parameter λ_j , $j \in \{1, 2\}$; thus $F_{\mu_j}(x) = 1 - e^{-\lambda_j(x)}$, with $x > 0$ and the probability of arrival of a claim in group j is $\pi_j > 0$, with $\pi_1 + \pi_2 = 1$. We assume that the λ_j 's are given and without loss of generality $\lambda_1 < \lambda_2$ (for $\lambda_1 = \lambda_2$ we find the usual exponential distribution). Instead the π_j 's are estimated empirically from a sample of n claims; denoting X_i a random variable taking values 1 or 2 depending on whether claim i belongs to the group j , we obtain $X_i = j$ with probability π_j and $\hat{\pi}_n(j) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i=j}$ for all $j \in \{1, 2\}$. In this example, we are interested in understanding what happens to $\rho \left(\sum_{j=1}^s \hat{\pi}_n(j) \mu_j \right)$ as $n \rightarrow \infty$ and the risk measure ρ is quantile at level $\alpha \in (0, 1)$, also known as Value-at-Risk (VaR_α) in the risk management literature. We recall that for a model μ with continuous and strictly increasing distribution function F , we have $\rho(\mu) = F^{-1}(\alpha)$, therefore we have

$$\rho(\mu_j) = -\frac{1}{\lambda_j} \log(1 - \alpha), \quad j \in \{1, 2\}.$$

From Example 4.2, we know that

$$\Psi_\rho(\mu_j, r) = \alpha - F_{\mu_j}(r) = \alpha - 1 + e^{-\lambda_j r},$$

and

$$\bar{r}_\rho = r_1^{(0)} = -\frac{1}{\lambda_1} \log(1 - \alpha) > -\frac{1}{\lambda_2} \log(1 - \alpha) = r_2^{(0)} = \underline{r}_\rho.$$

From Remark 3.3, for $r \in (\underline{r}_\rho, \bar{r}_\rho)$ we easily find

$$\lambda_*(r) = -\frac{1}{e^{-\lambda_2 r} - e^{-\lambda_1 r}} \log \left(-\frac{\pi_1(\alpha - 1 + e^{-\lambda_1 r})}{\pi_2(\alpha - 1 + e^{-\lambda_2 r})} \right)$$

and the weights $p = (p_1, p_2)$ in (7) which attain the infimum in (4) are given by

$$p_j(r) = \frac{\pi_j e^{-\lambda_*(r)(\alpha - 1 + e^{-\lambda_j r})}}{\sum_{j=1}^2 \pi_j e^{-\lambda_*(r)(\alpha - 1 + e^{-\lambda_j r})}}, \quad j \in \{1, 2\}.$$

We then obtain the rate function:

$$H_{\rho, \langle \pi, \mu \rangle}(r) = -\log \left(\pi_1 \left(-\frac{\pi_1(\alpha - 1 + e^{-\lambda_1 r})}{\pi_2(\alpha - 1 + e^{-\lambda_2 r})} \right)^{\frac{\alpha - 1 + e^{-\lambda_1 r}}{e^{-\lambda_2 r} - e^{-\lambda_1 r}}} + \pi_2 \left(-\frac{\pi_1(\alpha - 1 + e^{-\lambda_1 r})}{\pi_2(\alpha - 1 + e^{-\lambda_2 r})} \right)^{\frac{\alpha - 1 + e^{-\lambda_2 r}}{e^{-\lambda_2 r} - e^{-\lambda_1 r}}} \right).$$

We refer the interested reader to Lee et al. (2012) for a more detailed analysis of the use of exponential mixture models in insurance.

Example 4.5 (A specific example with $s = 3$). We refer to Condition 3.1 with $s = 3$ and, for some $a > 0$ (and a suitable strictly decreasing function Ψ), we set

$$\Psi_\rho(\mu_i, r) := \Psi(r) + (i - 1)a \text{ (for all } i \in \{1, 2, 3\});$$

(we remark that we are in this situation when we deal with Example 4.1, with $s = 3$; in such a case we have $\Psi(r) = \rho(\mu_1) - r$, $\rho(\mu_2) = \rho(\mu_1) + a$, and $\rho(\mu_3) = \rho(\mu_1) + 2a$).

We recall that

$$\underline{r}_\rho \leq r_1^{(0)} < r_2^{(0)} < r_3^{(0)} \leq \bar{r}_\rho,$$

where $r_i^{(0)} = \Psi^{-1}((i - 1)a)$ (for all $i \in \{1, 2, 3\}$). Moreover, after some computations, (5) with $s = 3$ yields

$$r_0 = \Psi^{-1}(-a(\pi_2 + 2\pi_3)).$$

Now we compute the rate function $H_{\rho, \langle \pi, \mu \rangle}$ in Theorem 3.1(ii). We have

$$H_{\rho, \langle \pi, \mu \rangle}(\Psi^{-1}(0)) = -\log \pi_1 \text{ and } H_{\rho, \langle \pi, \mu \rangle}(\Psi^{-1}(-2a)) = -\log \pi_3,$$

which concern the cases $r = \underline{r}_\rho$ and $r = \bar{r}_\rho$. In what follows we take $r \in (\underline{r}_\rho, \bar{r}_\rho) = (\Psi^{-1}(0), \Psi^{-1}(-2a))$. Firstly (6) yields

$$e^{-\lambda_*(r)\Psi(r)} \left(\pi_1 \Psi(r) + \pi_2(\Psi(r) + a)e^{-\lambda_*(r)a} + \pi_3(\Psi(r) + 2a)e^{-2\lambda_*(r)a} \right) = 0,$$

and therefore

$$e^{-\lambda_*(r)a} = \frac{-\pi_2(\Psi(r) + a) + \sqrt{\pi_2^2(\Psi(r) + a)^2 - 4\pi_1\pi_3\Psi(r)(\Psi(r) + 2a)}}{2\pi_3(\Psi(r) + 2a)};$$

we remark that $\pi_2^2(\Psi(r) + a)^2 - 4\pi_1\pi_3\Psi(r)(\Psi(r) + 2a) \geq 0$ because $\Psi(r) < 0$ and $\Psi(r) + 2a > 0$, and

$$\sqrt{\pi_2^2(\Psi(r) + a)^2 - 4\pi_1\pi_3\Psi(r)(\Psi(r) + 2a)} \geq |\pi_2(\Psi(r) + a)|.$$

Thus we easily obtain $\lambda_*(r)$ and the rate function expression is

$$H_{\rho, \langle \pi, \mu \rangle}(r) = -\log \left(\sum_{j=1}^3 \pi_j \left(\frac{-\pi_2(\Psi(r) + a) + \sqrt{\pi_2^2(\Psi(r) + a)^2 - 4\pi_1\pi_3\Psi(r)(\Psi(r) + 2a)}}{2\pi_3(\Psi(r) + 2a)} \right)^{\Psi_\rho(\mu_j, r)/a} \right).$$

A final remark concerns the condition $\lambda_*(r_0) = 0$ (see Remark 3.2). The above formulas yield

$$\frac{-\pi_2(\Psi(r_0) + a) + \sqrt{\pi_2^2(\Psi(r_0) + a)^2 - 4\pi_1\pi_3\Psi(r_0)(\Psi(r_0) + 2a)}}{2\pi_3(\Psi(r_0) + 2a)} = 1$$

and, after some computations, we get

$$\pi_1\Psi(r_0) + \pi_2(\Psi(r_0) + a) + \pi_3(\Psi(r_0) + 2a) = 0$$

which recovers (5) (with $s = 3$).

Acknowledgements. The authors are grateful to anonymous reviewers for their careful reports which highly improved the paper.

References

- Acciaio, B. and G. Svindland (2013). Are law-invariant risk functions concave on distributions? *Dependence Modeling* 1, 54–64.
- Barrieu, P. and G. Scandolo (2015). Assessing financial model risk. *European Journal of Operational Research* 242(2), 546–556.
- Bellini, F. and V. Bignozzi (2015). On elicitable risk measures. *Quantitative Finance* 15(5), 725–733.
- Bernardi, M., A. Maruotti, and L. Petrella (2012). Skew mixture models for loss distributions: a Bayesian approach. *Insurance: Mathematics and Economics* 51(3), 617–623.
- Bernardi, M., A. Maruotti, and L. Petrella (2017). Multiple risk measures for multivariate dynamic heavy-tailed models. *Journal of Empirical Finance* 43, 1–32.
- Bignozzi, V. and A. Tsanakas (2016). Model uncertainty in risk capital measurement. *The Journal of Risk* 18(3), 1–24.
- Cairns, A. J. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics* 27(3), 313–330.
- Cont, R., R. Deguest, and G. Scandolo (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance* 10(6), 593–606.
- Delbaen, F., F. Bellini, V. Bignozzi, and J. F. Ziegel (2016). Risk measures with the CxLS property. *Finance and Stochastics* 20(2), 433–453.
- Dembo, A. and O. Zeitouni (1998). *Large Deviations Techniques and Applications* (2nd ed.). Springer.

- Féray, V., P.-L. Méliot, and A. Nikeghbali (2016). *Mod- ϕ Convergence. Normality Zones and Precise Deviations*. Springer.
- Föllmer, H. and A. Schied (2002) Convex measures of risk and trading constraints. *Finance and stochastic* 6(4), 429–447.
- Föllmer, H. and A. Schied (2016). *Stochastic Finance: An Introduction in Discrete Time*. Walter De Gruyter.
- Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2), 141–153.
- Klugman, S. A., H. H. Panjer, and G. E. Willmot (2012). *Loss Models: From Data to Decisions*. John Wiley & Sons.
- Lee, D., Li, W. K. and T.S.T. Wong (2012). Modeling insurance claims via a mixture exponential model combined with peaks-over-threshold approach. *Insurance: Mathematics and Economics* 51(3), 538–550.
- Lee, S. C. and X. S. Lin (2010). Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal* 14(1), 107–130.
- McLachlan, G. and D. Peel (2004). *Finite Mixture Models*. John Wiley & Sons.
- Pesaran, M. H., C. Schleicher, and P. Zaffaroni (2009). Model averaging in risk management with an application to futures markets. *Journal of Empirical Finance* 16(2), 280–305.
- Raftery, A. E., D. Madigan, and J. A. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92(437), 179–191.
- Varadhan, S. R. S. (2003). Large deviations and entropy. In *Entropy*, Princeton Ser. Appl. Math., pp. 199–214. Princeton University Press.
- Weber, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance* 16(2), 419–441.
- Weber, S. (2007). Distribution-invariant risk measures, entropy, and large deviations. *Journal of Applied Probability* 44(1), 16–40.
- Ziegel, J. F. (2016). Coherence and elicibility. *Mathematical Finance* 26(4), 901–918.

Appendix: The continuity of $\rho\left(\sum_{j=1}^s p_j \mu_j\right)$ in Example 4.2

We remark that, if we set $\varphi(x, p_1, \dots, p_s) := \sum_{j=1}^s p_j F_{\mu_j}(x)$, we are interested in the function

$$x(p_1, \dots, p_s) := \left(\sum_{j=1}^s p_j F_{\mu_j} \right)^{-1}(\alpha),$$

which is the implicit function defined by the condition $\varphi(x, p_1, \dots, p_s) = \alpha$.

We assume that $x(p_1, \dots, p_s)$ is not continuous at some point (q_1, \dots, q_s) . Then we can find a sequence $\{(p_1^{(n)}, \dots, p_s^{(n)}) : n \geq 1\}$ which converges to (q_1, \dots, q_s) , and $\{x(p_1^{(n)}, \dots, p_s^{(n)}) : n \geq 1\}$ converges to some limit ℓ , with $\ell \neq x(q_1, \dots, q_s)$; moreover we have

$$\varphi(x(p_1^{(n)}, \dots, p_s^{(n)}), p_1^{(n)}, \dots, p_s^{(n)}) = \alpha \text{ (for all } n \geq 1)$$

and, if we take the limit as $n \rightarrow \infty$, we get $\sum_{j=1}^s q_j F_{\mu_j}(\ell) = \alpha$ (by the continuity of the functions $F_{\mu_1}, \dots, F_{\mu_s}$). The last equality yields $\ell = x(q_1, \dots, q_s)$ by construction, and this is a contradiction.