Coherence and elicitability

Johanna F. Ziegel*

Abstract

The risk of a financial position is usually summarized by a risk measure. As this risk measure has to be estimated from historical data, it is important to be able to verify and compare competing estimation procedures. In statistical decision theory, risk measures for which such verification and comparison is possible, are called elicitable. It is known that quantile based risk measures such as value at risk are elicitable. In this paper, Gneiting's (2011) result of the non-elicitability of expected shortfall is extended to all law-invariant spectral risk measures unless they reduce to minus the expected value. Hence, it is unclear how to perform forecast verification or comparison. However, the class of elicitable law-invariant coherent risk measures does not reduce to minus the expected value. We show that it consists of certain expectiles.

Keywords: Coherent risk measures; Decision theory; Elicitability; Expected shortfall; Expectiles; Law-invariant risk measures; Spectral risk measures

1 Introduction

Value at Risk (VaR) is the most common risk measure used in banking and finance. The *VaR at level* $\alpha \in (0,1)$ is given by

$$\mathrm{VaR}_\alpha(Y) = -\inf\{x \in \mathbb{R} \mid F_Y(x) \geq \alpha\},$$

where the financial position Y is a real-valued random variable, and F_Y is its cumulative distribution function. In this paper, a positive value of Y

Address correspondence to Johanna F. Ziegel, University of Bern, Department of Mathematics and Statistics, Institute of Mathematical Statistics and Actuarial Science, Sidlerstrasse 5, 3012 Bern, Switzerland, e-mail:johanna.ziegel@stat.unibe.ch

denotes a profit. The sign convention we have chosen for VaR implies that extreme losses correspond to levels α close to zero, and for $Y \leq 0$, the risk VaR$_\alpha(Y)$ will be non-negative. Since the influential paper of Artzner et al. (1999) introduced coherent risk measures, VaR has frequently been criticized as a risk measure because it fails to be subadditive, and hence it is not coherent; see for example Acerbi (2002). Other authors have pointed out the lack of VaR at level α to account for the size of losses beyond the level α (Daníelsson et al., 2001). Median shortfall at level α, or equivalently, VaR at level $(\alpha + 1)/2$ (Kou et al., 2013), does account for the size of losses beyond level α.

The Basel Committe on Banking Supervision (2012) has been investigating the points in favor of and against a change of the regulatory risk measure from VaR to the coherent risk measure *expected shortfall (ES)*, also known as *average* or *conditional value at risk*, which is defined by

$$\mathrm{ES}_\alpha(Y) = \frac{1}{\alpha} \int_0^\alpha \mathrm{VaR}_\tau(Y) d\tau.$$

From the perspective of coherent risk measures, ES is a better alternative to VaR. It remedies both problems mentioned above: It is a coherent risk measure, and it is sensitive to the sizes of the potential losses beyond the threshold α. Other popular coherent risk measures are the so-called *spectral risk measures*, which generalize ES (Acerbi, 2002).

However, despite their theoretical appeal, there are also major drawbacks to using spectral risk measures in risk management, which should not be neglected. Cont et al. (2010) show that there is a fundamental theoretical conflict between subadditivity and robustness of risk measurement procedures for spectral risk measures; see also the related discussion in Kou et al. (2013, Section 5). Cont et al. (2010) state

> We hope to have convinced the reader that there is more to risk measurement than the choice of a 'risk measure': statistical robustness, and not only 'coherence', should be a concern for regulators and end-users when choosing or designing risk measurement procedures. The design of robust risk estimation procedures requires the explicit inclusion of the statistical estimation step in the analysis of the risk measurement procedure.

The next steps beyond estimation are backtesting and forecast verification. Backtesting refers to validating a given estimation procedure for a risk measure on historical data. In this paper, following the ideas of Gneiting (2011), we consider risk measures from a forecasting perspective. With our knowledge of today, we are trying to give the best possible point estimate of the

risk measure for tomorrow, or ten days ahead, or for any other time point in the future. There are numerous choices concerning models, methods and parameters that have to be made to come up with predictions. Hence, for a number of competing forecast or estimation procedures we would like to decide which one performs best. If we restrict our attention to law-invariant coherent risk measures as introduced by Kusuoka (2001) we can view them as functionals on some set $\mathcal{P}$ of probability distributions on $\mathbb{R}$. From the viewpoint of statistical decision theory not all functionals allow for meaningful point forecasts; see Gneiting (2011). Functionals for which meaningful point forecasts and forecast performance comparisons are possible are called *elicitable*; see Section 2 for details. One important example of elicitable functionals are quantiles, hence VaR is elicitable.

Gneiting (2011) has shown that ES is not elicitable, which may be a partial explanation for the difficulties with robust estimation and backtesting. This raises the natural question whether there is a different option. Is there any (interesting) law-invariant coherent risk measure that is also an elicitable functional? We show that the only law-invariant spectral risk measure that is also elicitable is minus the expected value:

$$\rho(Y) = -\mathbb{E}[Y];$$

see Corollary 4.3. However, there are law-invariant coherent risk measures that are elicitable. They are *expectiles* which were first introduced by Newey and Powell (1987). The elicitability of expectiles is a simple corollary of their definition. They have been considered as a risk measure by Kuan et al. (2009). Proposition 4.4 shows that they are coherent risk measures. Very recently, and independently of our work, a proof of this result also appears in Bellini et al. (2013). Proposition 4.4 also identifies the minimal generating set of the Kusuoka representation as defined in Pichler and Shapiro (2012, Definition 2.3). Expectiles are the only elicitable law-invariant coherent risk measures; see Section 4.3.

In the literature, there are procedures for evaluating ES forecasts and that allow for tests; see for example McNeil and Frey (2000); Christoffersen (2003). However, these methods do not allow for a direct comparison and ranking of the predictive performance of competing forecasting methods (Gneiting, 2011).

The non-elicitability of spectral risk measures, and in particular of ES, is the reason that there is no analogue to the quantile regression method (Koenker, 2005) for these functionals, and no M-estimators can be constructed. Chun et al. (2012) construct a mixed quantile estimator for ES as an approximation to an M-estimator. The recent contribution of Rockafellar et al.

(2013) takes this approach further and proposes a framework for 'generalized' regression that is suitable for ES. However, the problem with forecast comparison remains.

The paper is organized as follows. In Section 2 we introduce the notion of elicitability and describe its importance in point forecasting. A brief introduction to law-invariant coherent risk measures is given in Section 3. Section 4 contains the main results of the paper, showing in particular that law-invariant spectral risk measures are not elicitable, and elaborating the prominent role of expectiles as the only elicitable law-invariant coherent risk measures. We conclude the paper with a discussion; see Section 5.

2 Elicitability

Let $\mathcal{P}$ be a class of probability measures on $\mathbb{R}$ with the Borel sigma algebra. We consider a functional

$$\nu : \mathcal{P} \to 2^{\mathbb{R}}, \quad P \mapsto \nu(P) \subset \mathbb{R},$$

where $2^{\mathbb{R}}$ denotes the power set of $\mathbb{R}$. Often, but not always, $\nu(P)$ is single valued, for example if we consider the expectation functional $\mathbb{E}$ on the class of all probability measures with finite mean. However, quantile functionals may be set-valued. In the case of single valued functionals we will confound the one-point set $\nu(P)$ with its unique element.

In this paper we are interested in the statistical properties of functionals that are law-invariant coherent risk measures; see Section 3. The following Definitions 2.1 and 2.2 are central in the context of point forecasting; see Gneiting (2011, Section 2) for a discussion of their historical background. Let Y be a real-valued random variable, which models the future observation of interest.

Definition 2.1. A scoring function $s : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ is *consistent* for the functional ν relative to the class $\mathcal{P}$, if

$$\mathbb{E}_P s(t, Y) \leq \mathbb{E}_P s(x, Y) \tag{1}$$

for all $P \in \mathcal{P}$, all $t \in \nu(P)$, and all $x \in \mathbb{R}$. Here, Y has distribution P. It is *strictly consistent* if it is consistent and equality in (1) implies that $x \in \nu(P)$.

Given a consistent scoring function s for a functional ν, an *optimal forecast* $\hat{x}$ for $\nu(P)$ is given by

$$\hat{x} = \arg\min_x \mathbb{E}_P s(x, Y).$$

Competing forecast procedures for ν can be compared using the scoring function s. Suppose that in n forecast cases we have point forecasts $x_1^{(k)}, \ldots, x_n^{(k)}$, $k = 1, \ldots, K$, and realizing observations $y_1, \ldots, y_n$. The index k numbers the K competing forecast procedures. We can rank the procedures by their average scores

$$\bar{s}^{(k)} = \frac{1}{n} \sum_{i=1}^{n} s\big(x_i^{(k)}, y_i\big).$$

The consistency of the scoring rule for the functional ν ensures that accurate forecasts of $\nu(P)$ are rewarded. On the contrary, evaluating point forecasts with respect to 'some' scoring function, which is not consistent for ν, may lead to grossly misguided conclusions about the quality of the forecasts. A drastic example is provided in the simulation study of Gneiting (2011, Section 1.2). Summarized in rough terms, one can construct realistic examples where the performance of skilful statistical forecasts is ranked worse than an ignorant no-change forecast when evaluated by 'some' scoring function, such as the absolute error or the squared error, for example. Therefore, point forecasts for a functional ν have to be evaluated by means of a scoring function, which is consistent for ν.

Definition 2.2. A functional ν is *elicitable* relative to the class $\mathcal{P}$, if there exists a scoring function s which is strictly consistent for ν relative to $\mathcal{P}$.

Many interesting functionals are elicitable and a wealth of examples is given in Gneiting (2011). The most prominent example concerning risk management may be VaR, which is essentially a quantile and as such elicitable. The scoring functions that are consistent for α-quantiles have been characterized by Thomson (1979); Saerens (2000); see also Gneiting (2011, Theorem 9). Subject to some regularity and integrability conditions, they are given by

$$s(x,y) = (\mathbb{1}\{x \geq y\} - \alpha)(g(x) - g(y)), \tag{2}$$

where g is an increasing function and $\mathbb{1}$ denotes the indicator function.

However, not all functionals are elicitable, the most striking example in the present context being ES. The following necessary condition is due to Osband (1985); see also Lambert et al. (2008). As this theorem is central to the results presented in this paper, we provide a proof.

Theorem 2.1 (Osband). *An elicitable functional ν has* convex level sets *in the following sense: If $P_0 \in \mathcal{P}$ and $P_1 \in \mathcal{P}$, and $P^* = pP_0 + (1-p)P_1 \in \mathcal{P}$ for some $p \in (0,1)$, then $t \in \nu(P_0)$ and $t \in \nu(P_1)$ imply $t \in \nu(P^*)$.*

Proof. Let s be a strictly consistent scoring function for ν and let P_0, P_1, P^*, p and t be as required in the Theorem. Then we obtain for any $x \in \mathbb{R}$ that

$$\begin{aligned}
\mathbb{E}_{P^*}s(t,Y) &= p\mathbb{E}_{P_0}s(t,Y) + (1-p)\mathbb{E}_{P_1}s(t,Y) \\
&\leq p\mathbb{E}_{P_0}s(x,Y) + (1-p)\mathbb{E}_{P_1}s(x,Y) = \mathbb{E}_{P^*}s(x,Y),
\end{aligned}$$

hence $t \in \nu(P^*)$. $\square$

3 Coherent risk measures

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a standard probability space without atoms. A *coherent risk measure* is a map $\rho : L^\infty(\Omega, \mathcal{F}, \mathbf{P}) \to \mathbb{R}$, which fulfils the following four properties. It is *monotone*, so $X \geq Y$ implies $\rho(X) \leq \rho(Y)$; it is *subadditive*, that is $\rho(X+Y) \leq \rho(X) + \rho(Y)$ for all $X, Y \in L^\infty(\Omega, \mathcal{F}, \mathbf{P})$; it is *positively homogeneous*, i.e. for $\lambda \geq 0$, it holds that $\rho(\lambda Y) = \lambda\rho(Y)$. Finally, it is *translation invariant* in the sense that for all $a \in \mathbb{R}$, we have $\rho(Y+a) = \rho(Y) - a$.

Coherent risk measures were introduced by Artzner et al. (1999); see also Delbaen (2002) and Föllmer and Schied (2004, Chapter 4). All common coherent risk measures used in applications, share the property of *law-invariance*. That is, if X and Y have the same distribution P on $\mathbb{R}$, then

$$\rho(P) := \rho(X) = \rho(Y).$$

Law-invariant risk measures were characterized by Kusuoka (2001). His result was strengthened by Jouini et al. (2006). We summarize the result that is relevant for this paper in the following theorem; compare Jouini et al. (2006, Theorem 2.1). Let $\mathfrak{P}$ denote the set of all probability measures on $[0,1]$ with the weak topology. For a cumulative distribution function $F : \mathbb{R} \to [0,1]$ we define its *generalized inverse* or *quantile function* by

$$F^{-1} : [0,1] \to \mathbb{R}, \quad \alpha \mapsto \inf\{x \in \mathbb{R} \mid F(x) \geq \alpha\}.$$

Theorem 3.1 (Kusuoka)**.** *Let ρ be a law-invariant coherent risk measure. Then there exists a closed convex set $\mathcal{M} \subset \mathfrak{P}$ such that*

$$\rho(Y) = -\inf_{m \in \mathcal{M}} \int_{[0,1]} U_\alpha(Y) dm(\alpha),$$

where

$$U_\alpha(Y) := \frac{1}{\alpha} \int_0^\alpha F_Y^{-1}(u) du, \quad \alpha \in (0,1],$$

and $U_0(Y) = \operatorname{ess\,inf} Y$.

For $m \in \mathfrak{P}$, $Y \in L^\infty(\Omega, \mathcal{F}, \mathbf{P})$ we define

$$\nu_m(Y) = \int_{[0,1]} U_\alpha(Y) m(d\alpha). \tag{3}$$

Up to the sign, the ν_m are exactly the *spectral risk measures* of Acerbi (2002). The following alternative representation of ν_m will be useful in the following. It is a direct consequence of Fubini's theorem.

$$
\begin{aligned}
\nu_m(Y) &= \int_{(0,1]} U_\alpha(Y) m(d\alpha) + m(\{0\}) U_0(Y) \\
&= \int_{(0,1]} \frac{1}{\alpha} \int_{(0,1]} \mathbb{1}\{u \le \alpha\} F_Y^{-1}(u) du \, m(d\alpha) + m(\{0\}) U_0(Y) \\
&= \int_{(0,1]} \int_{[u,1]} \frac{1}{\alpha} m(d\alpha) F_Y^{-1}(u) du + m(\{0\}) U_0(Y) \\
&= \mathbb{E}(g_m(U) F_Y^{-1}(U)) + m(\{0\}) U_0(Y), \tag{4}
\end{aligned}
$$

where U is standard uniformly distributed, and $g_m : (0,1] \to [0,\infty)$ is given by

$$g_m(u) = \int_{[u,1]} \frac{1}{\alpha} m(d\alpha).$$

Following Pichler and Shapiro (2012) we call g_m the *spectral function* of $m \in \mathfrak{P}$. It is left-continuous, decreasing and $\int_0^1 g_m(u) du = m((0,1]) \le 1$. This implies in particular that the functional $\nu_m(Y)$ is finite for all $Y \in L^\infty(\Omega, \mathcal{F}, \mathbf{P})$. We call the function $p \mapsto \int_p^1 g_m(u) du$ the *integrated spectral function* of m. As ν_m is law-invariant, we will also write $\nu_m(P) := \nu_m(Y)$ if Y has distribution P.

4 Coherence and elicitability

4.1 Coherent functionals with convex level sets

If a functional ν is not elicitable with respect to a class of probability distributions $\mathcal{P}_0$, then it cannot be elicitable with respect to any larger class $\mathcal{P} \supset \mathcal{P}_0$. In particular, if the functional ν does not have convex level sets in the sense of Theorem 2.1 for some class $\mathcal{P}_0$, it will fail to have convex level sets for any larger class $\mathcal{P}$ containing $\mathcal{P}_0$. A simple class $\mathcal{P}^*$ of probability distributions, that has proven useful to show the violations of the necessary condition for elicitability in Theorem 2.1 is the class of two-point distributions, that is

$$\mathcal{P}^* = \{p\delta_x + (1-p)\delta_y \mid x, y \in \mathbb{R}, p \in [0,1]\}, \tag{5}$$

where δ_x is the Dirac measure at the point $x \in \mathbb{R}$.

The following theorem summarizes the main results of this paper.

Theorem 4.1. *Let $\mathcal{P}^*$ be the class of two-point distributions on $\mathbb{R}$; see (5). Let $\mathcal{M} \subseteq \mathfrak{P}$ be a closed set of probability measures on $[0, 1]$, that does not contain δ_0. If the functional*

$$\nu : \mathcal{P}^* \to \mathbb{R}, \quad P \mapsto \nu(P) = \inf_{m \in \mathcal{M}} \nu_m(P),$$

with ν_m defined at (3), has convex level sets, then $m(\{0\}) = 0$ for all $m \in \mathcal{M}$, and there exists a $C \in (0, 1]$ such that all probability measures

$$m_p = \frac{p(1-C)}{p(1-C) + C}\delta_p + \frac{C}{p(1-C) + C}\delta_1, \quad p \in (0, 1), \tag{6}$$

are contained in $\mathcal{M}$. Furthermore, for all measures $m \in \mathcal{M}$, there exists a $q \in (0, 1)$ such that

$$\int_p^1 g_m(v)dv \geq \int_p^1 g_{m_q}(v)dv, \quad p \in [0, 1], \tag{7}$$

and

$$\int_p^1 g_m(v)dv \geq \frac{C(1-p)}{C(1-p) + p}, \quad p \in [0, 1]. \tag{8}$$

The lower bound in (8) and the integrated spectral functions of m_p are illustrated in Figure 1. We would like to give a short summary of the proof of Theorem 4.1. The details are deferred to Section 4.4.

For a two-point distribution $p\delta_x + (1 - p)\delta_y$ it is possible to calculate that $\nu(p\delta_x + (1 - p)\delta_y) = \alpha_p x + (1 - \alpha_p)y$ for some $\alpha_p \in [0, 1]$. It follows from the properties of ν as a risk measure that $\alpha_p > 0$. The assumption $\delta_0 \notin \mathcal{M}$ is necessary to guarantee that for some $p \in (0, 1)$ we have $\alpha_p < 1$. Exploiting the convexity of level sets as given by Theorem 2.1, first we show that $\alpha_p \in (0, 1)$ for all $p \in (0, 1)$, and then we derive an explicit formula for α_p in terms of C and p.

Remark. Weber (2006) also studied law-invariant risk measures with convex level sets. His motivation came from considerations of dynamic consistency of risk measures, which he shows to be closely related to convexity of level sets. Weber (2006, Theorem 3.1) is more general than Theorem 4.1 in the sense that not only coherent risk measures are considered and that it provides a characterization instead of a necessary condition. However, his result requires regularity assumptions on the risk measure under consideration, which we do not impose in this work. See also the remarks in Section 4.2 and 4.3.

4.2 Spectral risk measures

The first corollary to Theorem 4.1 shows that none of the coherent risk measures considered in Pichler and Shapiro (2012, Example 3) are elicitable, unless they reduce to minus the expected value.

Corollary 4.2. *Suppose the functional ν in Theorem 4.1 has convex level sets and there is a finite set $\mathcal{M}_0 \subset \mathfrak{P}$, $\delta_0 \notin \mathcal{M}_0$ such that*

$$\nu(P) = \inf_{m \in \mathcal{M}_0} \nu_m(P), \quad P \in \mathcal{P}^*,$$

then $\mathcal{M}_0 = \{\delta_1\}$ and

$$\nu(P) = \mathbb{E}_P(Y).$$

Proof. By Theorem 4.1 the closed set $\mathcal{M}_0$ is uncountable unless $C = 1$. If $C = 1$, then the lower bound in (8) is $1 - p$, which is the integrated spectral function of δ_1. The claim follows directly from Dana (2005, Lemma 2.2). □

Now, it follows easily that elicitable spectral risk measures are essentially the expected value.

Corollary 4.3. *Spectral risk measures, other than minus the expected value, are not elicitable relative to any class of probability distributions that contains the two-point distributions.*

Proof. It is a direct consequence of Theorem 2.1 and Corollary 4.2 that any spectral risk measure, which is not minus the essential infimum, is not elicitable unless it is minus the expected value. Therefore, it only remains to show that $\operatorname{ess\,inf}(Y)$ is not an elicitable functional relative to the class of two-point distributions. Suppose the contrary, and let S be a strictly consistent scoring function. If $Y = a$ almost surely for some $a \in \mathbb{R}$, then $\mathbb{E}S(x, Y) = S(x, a)$ and we obtain $S(a, a) < S(x, a)$ for all $x \in \mathbb{R} \setminus \{a\}$. If Y has distribution $p\delta_a + (1 - p)\delta_b$ with $a < b$ and $p \in (0, 1]$, we obtain

$$pS(x, a) + (1 - p)S(x, b) > pS(a, a) + (1 - p)S(a, b), \quad x \in \mathbb{R} \setminus \{a\}.$$

With $x = b$ and letting $p \to 0$, we obtain $S(b, b) \geq S(a, b)$, a contradiction. □

Remark. While the proof of Corollary 4.3 shows that $\operatorname{ess\,inf}(Y)$ is not elicitable relative to the class of two-point distributions, it is easy to check that the interval $(-\infty, \operatorname{ess\,inf}(Y)]$ is elicitable. Strictly consistent scoring functions are given at (2) with $\alpha = 0$ and any strictly increasing function g.

Gneiting (2011, Theorem 11) shows that ES is not elicitable with respect to any class of probability measures that contains the measures with finite support, or the finite mixtures of absolutely continuous distributions with compact support. In both cases the proof is done by showing a violation of the necessary condition of convex level sets given in Theorem 2.1. We believe that it is possible to modify the proof of Theorem 4.1 using mixtures of absolutely continuous distributions with compact support instead of two-point distributions. However, the details remain to be worked out and are likely to be rather technical.

Remark. While the result that ES is not elicitable is due to Gneiting (2011), the non-convexity of its level sets already appears in Weber (2006, Example 3.4).

4.3 Expectiles

Theorem 4.1 provides an upper and a lower bound on a potentially elicitable law-invariant coherent risk measures ρ via the provided restrictions on the integrated spectral functions. This is illustrated in Figure 1, and details are given below.

Let $\mathcal{M} \subset \mathfrak{P}$ be a closed set such that

$$\rho(X) = -\inf_{m \in \mathcal{M}} \nu_m(X), \quad X \in L^\infty(\Omega, \mathcal{F}, \mathbf{P}).$$

By Dana (2005, Lemma 2.2) equation (8) of Theorem 4.1 implies that there exists a $C \in (0, 1]$ such that

$$\rho(X) \leq \mathfrak{u}_C(X) := -\int_0^1 \frac{C}{(v + C(1 - v))^2} F_X^{-1}(v)dv.$$

The map $\mathfrak{u}_C : L^\infty(\Omega, \mathcal{F}, \mathbf{P}) \to \mathbb{R}$ is a spectral risk measure with spectral function $g(v) = C/(v + C(1 - v))^2$ for $v \in (0, 1]$. The associated measure $m \in \mathfrak{P}$ has density $2C(1 - C)v/(v + C(1 - v))^3$ on $(0, 1)$ and a point mass C at $v = 1$. The integrated spectral function $\int_p^1 g(v)dv$ is illustrated as a dashed line in Figure 1. By Corollary 4.3, the spectral risk measure $\mathfrak{u}_C$ is not elicitable unless $C = 1$. In this case, it reduces to minus the expected value.

We define

$$\mathcal{M}_C := \{m_p \in \mathfrak{P} \mid p \in (0, 1)\},$$

where m_p is given at (6) and

$$\mathfrak{l}_C(X) := -\inf_{m \in \mathcal{M}_C} \nu_m(X), \quad X \in L^\infty(\Omega, \mathcal{F}, \mathbf{P}). \tag{9}$$

By Theorem 4.1 we immediately obtain $\rho(X) \geq \mathfrak{l}_C(X)$. Invoking Dana (2005, Lemma 2.2) equation (7) yields $\rho(X) \leq \mathfrak{l}_C(X)$, hence $\rho(X) = \mathfrak{l}_C(X)$. In the remainder of this section we characterize the law-invariant coherent risk measure $\mathfrak{l}_C$.

As introduced in Newey and Powell (1987), the τ-*expectile* $\mu_\tau(X)$, $\tau \in (0, 1)$, of a random variable with finite mean is the unique solution $x = \mu_\tau(X)$ to the equation

$$\tau \int_x^\infty (y - x)dF_X(y) = (1 - \tau) \int_{-\infty}^x (x - y)dF_X(y). \qquad (10)$$

Bellini et al. (2013) show that (up to the sign) expectiles are law-invariant coherent risk measures for $\tau \in (0, 1/2]$. As mentioned in introduction, expectiles are elicitable. The scoring functions that are consistent for τ-expectiles were recently characterized by Gneiting (2011, Theorem 10). Subject to some regularity and integrability conditions, they are given by

$$s(x, y) = |\mathbb{1}\{x \geq y\} - \tau|(g(y) - g(x) - g'(x)(y - x)),$$

where g is a convex function with subgradient g'. The prominent role of expectiles as the only elicitable law-invariant coherent risk measures is underlined by the following proposition.

Proposition 4.4. *The law-invariant coherent risk measure $\mathfrak{l}_C$ defined at (9) is minus the τ-expectile for*

$$\tau := \frac{C}{C + 1} \ \in (0, \tfrac{1}{2}].$$

Proof. Let X a random variable with finite first moment, $F := F_X$, and $\mu := \mu_\tau(X)$ its τ-expectile with $\tau = C/(C + 1)$. We define

$$p^* := F(\mu).$$

We will show that $\nu_{m_{p^*}}(X) = \mu$ and that $\nu_{m_p}(X)$ is minimal at $p = p^*$. If F is continuous the latter claim can alternatively be shown by methods of calculus. We show the claims directly in order to avoid case distinctions.

For $p \in (0, 1)$, we obtain with m_p defined at (6)

$$\nu_{m_p}(X) = \frac{1}{p(1 - C) + p} \int_0^p F^{-1}(v)dv + \frac{C}{p(1 - C) + p} \int_p^1 F^{-1}(v)dv$$

$$= \frac{1}{p(1 - C) + p} \int_{-\infty}^{F^{-1}(p)} ydF(y) + \frac{C}{p(1 - C) + p} \int_{F^{-1}(p)}^\infty ydF(y)$$

$$+ \frac{1 - C}{p(1 - C) + p} F^{-1}(p)(p - F(F^{-1}(p))),$$

11

where we used Acerbi and Tasche (2002, Proposition 3.2) in the second step.

Newey and Powell (1987, Theorem 1) show that $p^* \in (0,1)$. Using (10) we obtain

$$\int_{-\infty}^{\mu} ydF(y) + C\int_{\mu}^{\infty} ydF(y)$$
$$= \int_{-\infty}^{\mu} ydF(y) + C\int_{\mu}^{\infty} \mu dF(y) + \int_{-\infty}^{\mu} \mu dF(y) - \int_{-\infty}^{\mu} ydF(y)$$
$$= C\mu(1 - F(\mu)) + \mu F(\mu) = \mu(p^*(1 - C) + C). \qquad (11)$$

Let $\varepsilon \geq 0$ such that $p := p^* + \varepsilon \in (0,1)$. Then, using equation (11) and partial integration, we obtain

$$\nu_{m_p}(X) = \frac{1}{p(1-C)+C}\Big(\int_{-\infty}^{\mu} ydF(y) + C\int_{\mu}^{\infty} ydF(y)$$
$$+ \int_{\mu}^{F^{-1}(p)} ydF(y) - C\int_{\mu}^{F^{-1}(p)} ydF(y)$$
$$+ (1-C)F^{-1}(p)(p - F(F^{-1}(p)))\Big)$$
$$= \frac{\mu(p^*(1-C)+C)}{p(1-C)+C}$$
$$+ \frac{1-C}{p(1-C)+C}\Big(F^{-1}(p)F(F^{-1}(p)) - \mu F(\mu)$$
$$- \int_{\mu}^{F^{-1}(p)} F(t)dt + F^{-1}(p)p - F^{-1}(p)F(F^{-1}(p))\Big)$$
$$= \mu - \frac{\mu\varepsilon(1-C)}{p(1-C)+C}$$
$$+ \frac{1-C}{p(1-C)+C}\Big(F^{-1}(p)p - \mu p + \mu\varepsilon - \int_{\mu}^{F^{-1}(p)} F(t)dt\Big)$$
$$= \mu + \frac{1-C}{p(1-C)+C}\Big(p(F^{-1}(p) - \mu) - \int_{\mu}^{F^{-1}(p)} F(t)dt\Big).$$

The last term in the above equation is always non-negative. It vanishes for $p = p^*$, hence $\nu_{m_{p^*}}(X) \leq \nu_{m_p}(X)$ for all $p \geq p^*$. The argument for $p \leq p^*$ is completely analogous. $\qquad\square$

Remark. Expectiles as coherent risk measures also appear implicitly in Weber (2006, Corollary 3.2). He shows that the shortfall risk measure with loss function $\ell(x) = \alpha x^+ - \beta x^-$ for $\alpha \geq \beta > 0$ is coherent. Such a shortfall risk measure is equal to the minus the τ-expectile with $\tau = \beta/(\alpha +$
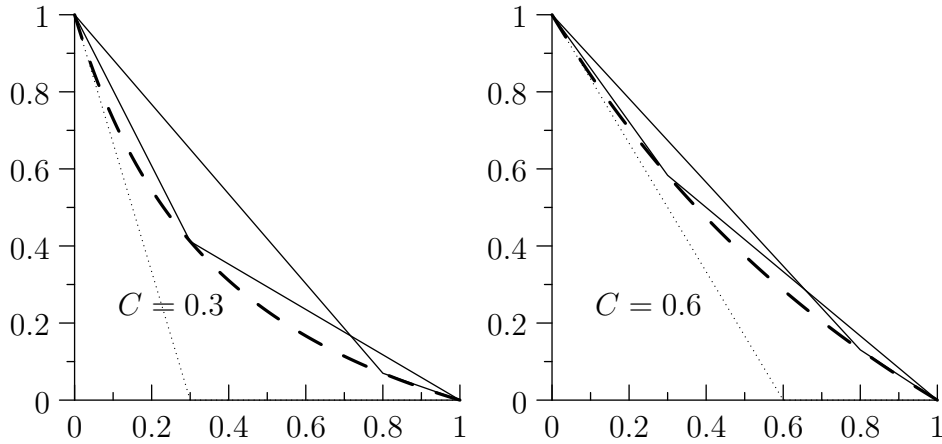
Figure 1: Integrated spectral functions. In both panels, the dashed lines are the integrated spectral functions of $-\mathfrak{u}_C$, and the solid lines are those of $\nu_{m_{0.3}}$ and $\nu_{m_{0.8}}$ as examples. For comparison, the dotted line is the integrated spectral function of $-ES_C$.

β). However, Weber (2006) did not draw the connection to the expectiles (Newey and Powell, 1987) in the statistical literature. Under the additional regularity assumptions (3.1) and (1) of Weber (2006, Theorem 3.1), Weber (2006, Corollary 3.1) characterizes all coherent risk measures with convex level sets as minus τ-expectiles with $\tau \in (0, 1/2]$. Theorem 4.1 shows that these conditions are not necessary for the characterization in the coherent case.

4.4   Proof of Theorem 4.1

For each $m \in \mathcal{M}$, $0 < p_1 \leq p_2 \leq 1$, we define

$$P_m(p_1, p_2) = \int_{p_1}^{p_2} g_m(v)dv, \quad P_m(0, p_2) = \int_0^{p_2} g_m(v)dv + m(\{0\})$$

Using Fubini we obtain

$$\begin{aligned} P_m(p_1, p_2) &= \int_{p_1}^{p_2} \int_{[v,1]} \frac{1}{\alpha} m(d\alpha)dv \\ &= \int_{[p_1,1]} \frac{1}{\alpha} \int_{p_1}^{\alpha \wedge p_2} dvm(d\alpha) = \int_{[p_1,1]} \frac{\alpha \wedge p_2 - p_1}{\alpha} m(d\alpha) \end{aligned} \quad (12)$$

13

For all $q \in (0, 1]$, the set

$$\mathcal{C}_q = \{(P_m(0, q), P_m(q, 1)) \in \mathbb{R}^2 \mid m \in \mathcal{M}\}$$

is a subset of the unit simplex in $\mathbb{R}^2$ because $P_m(0, q) + P_m(q, 1) = 1$. The set $\mathcal{C}_q$ is also closed, which can be seen using Helly's theorem, the fact that $\mathcal{M}$ is closed, and the representation of of $P_m(q, 1)$ at (12). Let $\alpha_q := \sup\{\alpha \in [0, 1] \mid (\alpha, 1 - \alpha) \in \mathcal{C}_q\}$ be the lower boundary point of $\mathcal{C}_q$; see Figure 2 for an illustration. As $\mathcal{C}_q$ is closed, the supremum is attained and there is an $m_q \in \mathcal{M}$ such that $P_{m_q}(0, q) = \alpha_q$. Note that $\alpha_q > 0$. Suppose the contrary, then

$$\int_0^q g_{m_q}(v)dv + m_q(\{0\}) = 0,$$

which implies $m_q(\{0\}) = 0$ and $g_{m_q}(v) = 0$ for $v \in (0, q]$, hence $g_{m_q} \equiv 0$ because g_{m_q} is decreasing and non-negative. This is a contradiction because

$$\int_0^1 g_m(v)dv + m(\{0\}) = m([0, 1]) = 1.$$

The function $(0, 1] \to (0, 1], q \mapsto \alpha_q$ is increasing. Define

$$q^* := \inf\{q \in (0, 1] \mid \alpha_q = 1\}. \tag{13}$$

If $q^* = 0$, then $\alpha_q = 1$ for all $q \in (0, 1]$, hence $1 - \alpha_q = \int_q^1 g_{m_q}(v)dv = 0$. This implies $m_q((q, 1]) = 0$ for all $q \in (0, 1]$, and hence m_q converges weakly to δ_0 as $q \to 0$. As $\mathcal{M}$ is closed this is a contradiction to the assumption $\delta_0 \notin \mathcal{M}$. Therefore, $q^* > 0$. We will conclude later that, actually, $q^* = 1$.

Let $\mathcal{A} = \{x = (x_1, x_2) \in \mathbb{R}^2 \mid x_1 \leq x_2\}$. For $(x_1, x_2) \in \mathcal{A}$, $p \in [0, 1]$ the distribution function F of $p\delta_{x_1} + (1 - p)\delta_{x_2}$ has generalized inverse $F^{-1}(v) = x_1 \mathbb{1}\{v \leq p\} + x_2 \mathbb{1}\{v > p\}$, $v \in (0, 1]$. This yields using (4)

$$\nu_m(p\delta_{x_1} + (1 - p)\delta_{x_2}) = x_1 \int_0^p g_m(v)dv + x_2 \int_p^1 g_m(v)dv + m(\{0\})x_1$$
$$= P_m(0, p)x_1 + P_m(p, 1)x_2,$$

hence

$$\nu(p\delta_{x_1} + (1 - p)\delta_{x_2}) = \inf_{m \in \mathcal{M}} \nu_m(p\delta_{x_1} + (1 - p)\delta_{x_2})$$
$$= \inf_{(\alpha, 1-\alpha) \in \mathcal{C}_p} (\alpha x_1 + (1 - \alpha)x_2)$$
$$= \alpha_p x_1 + (1 - \alpha_p)x_2 = \nu_{m_p}(p\delta_{x_1} + (1 - p)\delta_{x_2}).$$
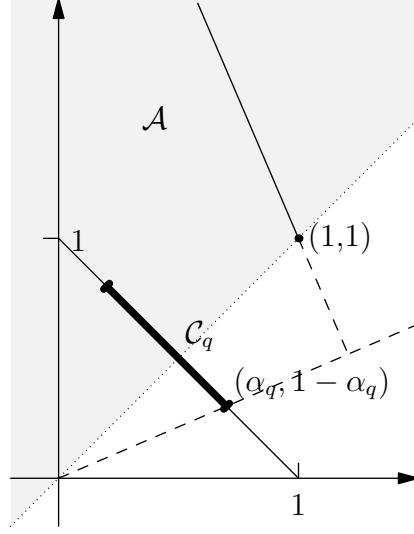
14

Figure 2: Illustration of the construction in the proof of Theorem 4.1.

Let $x_1 = x_2 = 1$, $q \in (0, q^*)$. Then $0 < \alpha_q < 1$. All $y = (y_1, y_2) \in \mathcal{A}$ with $\nu(\delta_1) = 1 = \nu(q\delta_{y_1} + (1-q)\delta_{y_2})$ are given by

$$y_1 = 1 - c, \quad y_2 = My_1 + a$$

for $M = -\alpha_q/(1-\alpha_q) \in (-\infty, 0)$, $a = 1 - M$ and $c \geq 0$; cf. Figure 2. Note that $y_1 \leq x_1 \leq x_2 \leq y_2$. Convexity of the level sets of ν implies that for all $m \in \mathcal{M}$, all $v \in (0, 1]$ and all $c > 0$, we have

$$
\begin{aligned}
1 \leq \nu_m\big(&v(q\delta_{y_1} + (1-q)\delta_{y_2}) + (1-v)(q\delta_{x_1} + (1-q)\delta_{x_2})\big)\\
= {}& P_m(0, qv)y_1 + P_m(qv, q)x_1 + P_m(q, 1-v+qv)x_2 + P_m(1-v+qv, 1)y_2\\
= {}& \nu_m(\delta_1) + P_m(0, qv)(y_1 - x_1) + P_m(1-v+qv, 1)(y_2 - x_2)\\
= {}& 1 - cP_m(0, qv) - cMP_m(1-v+qv, 1),
\end{aligned}
$$

hence,

$$P_m(0, qv) \leq \frac{\alpha_q}{1-\alpha_q} P_m(1-v+qv, 1). \tag{14}$$

We have $\lim_{v\downarrow 0} P_m(0, qv) = m(\{0\})$, and $\lim_{v\downarrow 0} P_m(1-v+qv, 1) = P_m(1, 1) = 0$, hence it follows that $m(\{0\}) = 0$ for all $m \in \mathcal{M}$. Equation (14) also implies that $g_m(v) > 0$ for all $v \in (0, 1)$, $m \in \mathcal{M}$. Suppose the contrary. Then there is a $w^* \in (0, 1)$ such that $g_m(v) = 0$ for all $v \in (w^*, 1]$, which implies $P_m(1 - v^* + qv^*, 1) = 0$ for some $v^* \in (0, 1)$. Now (14) yields $g_m \equiv 0$, which is a contradiction because $m(\{0\}) = 0$. Going back to the definition of q^* at (13) we obtain in particular that $1 - \alpha_p = \int_p^1 g_{m_p}(v)dv > 0$ for all $p \in (0, 1)$. Therefore, $q^* = 1$.

15

For $m = m_q$ we obtain

$$P_{m_q}(0, qv) \leq \frac{P_{m_q}(0, q)}{P_{m_q}(q, 1)} P_{m_q}(q + (1-q)(1-v), 1),$$

hence

$$P_{m_q}(0, qv)\big(P_{m_q}(q, 1 - v + qv) + P_{m_q}(1 - v + qv, 1)\big)$$
$$\leq \big(P_{m_q}(0, qv) + P_{m_q}(qv, q)\big) P_{m_q}(1 - v + qv, 1)$$

which yields

$$P_{m_q}(0, qv) P_{m_q}(q, 1 - v + qv) \leq P_{m_q}(qv, q) P_{m_q}(1 - v + qv, 1).$$

Monotonicity of g_m yields

$$qv g_{m_q}(qv)(1-q)(1-v) g_{m_q}(1 - v + qv) \leq P_{m_q}(0, qv) P_{m_q}(q, 1 - v + qv)$$
$$\leq P_{m_q}(qv, q) P_{m_q}(q + (1-q)(1-v), 1)$$
$$\leq q(1-v) g_{m_q}(qv+) v(1-q) g_{m_q}((1 - v + qv)+),$$

hence we obtain

$$P_{m_q}(0, qv) P_{m_q}(q, 1 - v + qv) = P_{m_q}(qv, q) P_{m_q}(1 - v + qv, 1). \qquad (15)$$

Equation (15) implies that

$$c_q := \frac{\alpha_q}{1 - \alpha_q} = \frac{P_{m_q}(0, q)}{P_{m_q}(q, 1)} = \frac{P_{m_q}(0, qv)}{P_{m_q}(1 - v + qv, 1)} = \frac{P_{m_q}(qv, q)}{P_{m_q}(q, 1 - v + qv)}, \qquad (16)$$

therefore

$$\frac{q}{1-q} \frac{g_{m_q}(qv)}{g_{m_q}((1 - v + qv)+)} \leq \frac{P_{m_q}(0, qv)}{P_{m_q}(1 - v + qv, 1)} = c_q =$$
$$= \frac{P_{m_q}(qv, q)}{P_{m_q}(q, 1 - v + qv)} \leq \frac{q}{1-q} \frac{g_{m_q}(qv+)}{g_{m_q}(1 - v + qv)},$$

and hence

$$c_q(1-q) g_{m_q}((1 - v(1-q))+) = q g_{m_q}(qv) \qquad (17)$$

for all $v \in (0, 1]$. The left-hand side is increasing in v, whereas the right-hand side is decreasing in v. Both sides are left-continuous. This implies that

$$g_{m_q}(w) = \begin{cases} c_1, & w \in (0, q], \\ c_2, & w \in (q, 1], \end{cases}$$

16

for two constants $c_1 \geq c_2$. We have

$$1 = m_q((0,1]) = \int_0^1 g_m(v)dv = qc_1 + (1-q)c_2$$

and by (17)

$$c_q(1-q)c_2 = qc_1,$$

hence

$$1 = c_q(1-q)c_2 + (1-q)c_2 = c_2(1-q)(1+c_q)$$

which yields

$$g_{m_q}(w) = \begin{cases} \alpha_q/q, & w \in (0,q], \\ (1-\alpha_q)/(1-q), & w \in (q,1]. \end{cases} \tag{18}$$

The inequality $c_1 \geq c_2$ is equivalent to $\alpha_q \geq q$. By the definition of g_{m_q} and (18) we obtain that

$$m_q = d_{q,1}\delta_q + d_{q,2}\delta_1,$$

where $d_{q,1}$, $d_{q,2}$ fulfil $d_{q,1} + d_{q,2} = 1$,

$$d_{q,2} = g_{m_q}(1) = \frac{1-\alpha_q}{1-q}.$$

Therefore

$$d_{q,1} = 1 - d_{q,2} = \frac{\alpha_q - q}{1-q}.$$

For any $m \in \mathcal{M}$ equation (14) implies

$$\frac{q}{1-q}\frac{g_m(qv)}{g_m((1-v+qv)+)} \leq \frac{P_m(0,qv)}{P_m(1-v+qv,1)} \leq \frac{\alpha_q}{1-\alpha_q}. \tag{19}$$

If $p < q \in (0,1)$, let $v = p/q$. Then the above inequality implies

$$\frac{q}{1-q}\frac{\alpha_p(1-p)}{(1-\alpha_p)p} \leq \frac{\alpha_q}{1-\alpha_q}.$$

On the other hand, we also obtain with $v < (1-p)/(1-q)$, that

$$\frac{p}{1-p}\frac{\alpha_q(1-q)}{(1-\alpha_q)q} \leq \frac{\alpha_p}{1-\alpha_p}.$$

Hence for $p, q \in (0, q^*)$ we obtain that

$$\frac{(1-\alpha_q)q}{\alpha_q(1-q)} = \frac{(1-\alpha_p)p}{\alpha_p(1-p)} =: C \in (0,1].$$

17

If $C = 0$, this implies $\alpha_p = 1$, which is a contradiction. This means that we can express α_p in terms of C and $p \in (0, 1)$ as

$$\alpha_p = \frac{p}{C(1-p) + p}.$$

Let $m \in \mathcal{M}$. To show equation (7), observe that $m(\{0\}) = 0$ implies that $\int_0^1 g_m(v)dv = 1$. Therefore, it suffices to show that there exists $q \in (0, 1)$ with $g_m(0+) \leq g_{m_q}(0+)$ and $g_m(1) \geq g_{m_q}(1)$ due to the convexity of the integrated spectral function $p \mapsto \int_p^1 g_m(v)dv$ and the piecewise linearity of $p \mapsto \int_p^1 g_{m_q}(v)dv$; cf. Figure 1.

Equation (19) implies

$$g_m(pv) \leq \frac{1}{C} g_m(1 - v + pv)$$

for $p \in (0, 1)$, $v \in (0, 1]$. Taking the limit as $v \downarrow 0$ yields

$$1 \leq g_m(0+) \leq \frac{1}{C} g_m(1) \leq \frac{1}{C}.$$

Suppose first that $C \in (0, 1)$. In this case we have that $g_m(0+) < 1/C$. If we suppose on the contrary that $g_m(0+) = 1/C$ it follows that $g_m(1) = 1$, hence $m = \delta_1$, which in turn implies $g_m(0+) = 1$, a contradiction. If $g_m(0+) = 1$, then (7) holds for all $q \in (0, 1)$. If $g_m(0+) > 1$, then there exists a $q \in (0, 1)$ such that $g_{m_q}(0+) = 1/(C(1-q) + q) = g_m(0+)$. Furthermore, it follows that

$$C g_m(0+) = \frac{C}{C(1-q) + q} = g_{m_q}(1) \leq g_m(1).$$

The case $C = 1$ is easy.

The last claim of the theorem follows because

$$1 - \alpha_p = \frac{C(1-p)}{C(1-p) + p} = \inf_{m \in \mathcal{M}} \int_p^1 g_m(v)dv.$$

$\square$

5 Discussion

In this paper we have shown that spectral risk measures are not elicitable, so it is unclear if and how it is possible to rank different point forecasts for such measures in a decision theoretically sound manner. In other words, objective comparison of competing estimation procedures for spectral risk measures

is difficult, if not impossible. This does not imply that it is impossible to perform backtests for a specific estimation procedure under fixed model assumptions. However, if one needs to decide between two estimation methods, the values of the test statistic used for backtesting should not be used for a quality ranking of the methods. Such a ranking should only be done through consistent scoring functions, which do not exist for spectral risk measures.

A possible solution to the problem could be working with probabilistic forecasts for Y in the following way. Suppose that an empirical loss distribution $\hat{F}_Y$ for F_Y has been estimated. Usually, using this estimated distribution, the forecast $\rho(\hat{F}_Y)$ for the risk measure $\rho(Y) = \rho(F_Y)$ of Y is calculated, and all further verification and backtesting of the model and estimation procedure are solely based on $\rho(\hat{F}_Y)$. Alternatively, one could directly assess the probabilistic forecast $\hat{F}_Y$; see for example Diebold et al. (1998); Berkowitz (2001). Competing forecasts could be compared using proper scoring rules; see for example Gneiting and Raftery (2007). Gneiting and Ranjan (2011) propose a method to compare density forecasts with emphasis on different regions of interest, such as the center or the tails of the distributions. If the aim is to accurately predict a functional focussed on the tails, such as ES, this approach seems promising.

McNeil and Frey (2000) propose a statistic, called *exceedance residuals*, for backtesting ES, which can be interpreted as a score, and is used as such in Chun et al. (2012, Section 5); see also McNeil et al. (2005, Section 4.4.3). This score is not a scoring function in the sense of this paper, as it depends on two functionals of the future outcome Y, namely on $\text{VaR}_\alpha(Y)$ and $\text{ES}_\alpha(Y)$. Potentially, it is necessary and useful to extend the concept of elicitability in order to put their construction into a decision theoretic framework. A useful notion may be *k-elicitability* as introduced in Lambert et al. (2008, Definition 11). It is an interesting open question whether the bivariate functional of VaR_α and ES_α is 2-elicitable.

Finally, Cont et al. (2010) propose to rethink the necessity of the subadditivity axiom as it is in conflict with robustness of risk measurement procedures for spectral risk measures; cf. Section 1. Supporting their suggestion, we believe that elicitability is also a crucial requirement that should be taken into consideration when choosing a risk measure, if probabilistic forecast evaluation is not an option. Given their appealing statistical properties, the potential of expectiles as risk measures should be further investigated.

References

C. Acerbi. Spectral measures of risk: A coherent representation of subjective risk aversion. *J. Bank. Financ.*, 26:1505–1518, 2002.

C. Acerbi and D. Tasche. On the coherence of expected shortfall. *J. Bank. Financ.*, 26:1487–1503, 2002.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9:203–228, 1999.

Basel Committe on Banking Supervision. Fundamental review of the trading book. Available from `www.bis.org/publ/bcbs219.htm`, May 2012.

F. Bellini, B. Klar, A. Müller, and E. Rosazza Gianin. Generalized quantiles as risk measures. *Preprint, available at* $http://dx.doi.org/10.2139/ssrn.2225751$, 2013.

J. Berkowitz. Testing density forecasts, with applications to risk management. *J. Bus. Econ. Stat.*, 19:465–474, 2001.

P. F. Christoffersen. *Elements of Financial Risk Management*. Academic Press, San Diego, 2003.

S. Y. Chun, A. Shapiro, and S. Uryasev. Conditional value-at-risk and average value-at-risk: Estimation and asymptotics. *Oper. Res.*, 60:739–756, 2012.

R. Cont, R. Deguest, and G. Scandolo. Robustness and sensitivity analysis of risk measurement procedures. *Quant. Finance*, 10:593–606, 2010.

R.-A. Dana. A representation result for concave Schur concave functions. *Math. Finance*, 15:613–634, 2005.

J. Daníelsson, P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, and H. S. Shin. An academic response to Basel II. Special paper no. 130, Financial Markets Group, London School of Economics, 2001.

F. Delbaen. Coherent risk measures on general probability spaces. In *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann*, pages 1–37. Springer, Berlin, 2002.

F. X. Diebold, T. A. Gunther, and A. S. Tay. Evaluating density forecasts with applications to financial risk management. *Int. Econ. Rev.*, 39:863–883, 1998.

H. Föllmer and A. Schied. *Stochastic Finance.* de Gruyter, Berlin, 2nd edition, 2004.

T. Gneiting. Making and evaluating point forecasts. *J. Amer. Statist. Assoc.*, 106:746–762, 2011.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.*, 102:359–378, 2007.

T. Gneiting and R. Ranjan. Comparing density forecasts using threshold- and quantile-weighted scoring rules. *J. Bus. Econ. Stat.*, 29:411–422, 2011.

E. Jouini, W. Schachermayer, and N. Touzi. Law invariant risk measures have the Fatou property. In *Advances in Mathematical Economics*, volume 9, pages 46–71. Springer, Tokyo, 2006.

R. Koenker. *Quantile Regression.* Cambridge University Press, Cambridge, 2005.

S. Kou, X. Peng, and C. C. Heyde. External risk measures and basel accords. *Math. Oper. Res.*, 2013. to appear.

C.-M. Kuan, J.-H. Yeh, and Y.-C. Hsu. Assessing value at risk with CARE, the Conditional Autoregressive Expectile models. *J. Econometrics*, 150: 261–270, 2009.

S. Kusuoka. On law-invariant coherent risk measures. *Adv. Math. Econ.*, 3: 83–95, 2001.

N. Lambert, D. M. Pennock, and Y. Shoham. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138, Chicago, Il, USA, 2008. extended abstract.

A. J. McNeil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *J. Empir. Financ.*, 7:271–300, 2000.

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management.* Princeton University Press, Princeton, 2005.

W. K. Newey and J. L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847, 1987.

K. H. Osband. *Providing Incentives for Better Cost Forecasting.* PhD thesis, University of California, Berkeley, 1985.

A. Pichler and A. Shapiro. Uniqueness of Kusuoka representations. *Preprint, arXiv: 1210. 7257*, 2012.

R. T. Rockafellar, J. O. Royset, and S. I. Miranda. Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional Value-at-Risk. *Preprint, available at http://www.math.washington.edu/~rtr/papers.html*, 2013.

M. Saerens. Building cost functions minimizing to some summary statistics. *IEEE Trans. Neural Netw.*, 11:1263–1271, 2000.

W. Thomson. Eliciting production possibilities from a well-informed manager. *J. Econ. Theory*, 20(3):360–380, 1979.

S. Weber. Distribution-invariant risk measures, information, and dynamic consistency. *Math. Finance*, 16:419–441, 2006.