HIGHER ORDER ELICITABILITY AND OSBAND'S PRINCIPLE[1]

BY TOBIAS FISSLER AND JOHANNA F. ZIEGEL

University of Bern

A statistical functional, such as the mean or the median, is called elicitable if there is a scoring function or loss function such that the correct forecast of the functional is the unique minimizer of the expected score. Such scoring functions are called strictly consistent for the functional. The elicitability of a functional opens the possibility to compare competing forecasts and to rank them in terms of their realized scores. In this paper, we explore the notion of elicitability for multi-dimensional functionals and give both necessary and sufficient conditions for strictly consistent scoring functions. We cover the case of functionals with elicitable components, but we also show that one-dimensional functionals that are not elicitable can be a component of a higher order elicitable functional. In the case of the variance, this is a known result. However, an important result of this paper is that spectral risk measures with a spectral measure with finite support are jointly elicitable if one adds the "correct" quantiles. A direct consequence of applied interest is that the pair (Value at Risk, Expected Shortfall) is jointly elicitable under mild conditions that are usually fulfilled in risk management applications.

1. Introduction. Point forecasts for uncertain future events are issued in a variety of different contexts such as business, government, risk-management or meteorology, and they are often used as the basis for strategic decisions. In all these situations, one has a random quantity Y with unknown distribution F. One is interested in a statistical property of F, that is a functional $T(F)$. Here, Y can be real-valued (GDP growth for next year), vector-valued (wind-speed, income from taxes for all cantons of Switzerland), functional-valued (path of the interchange rate Euro–Swiss franc over one day), or set-valued (area of rain tomorrow, area of influenza in a country). Likewise, also the functional T can have a variety of different sorts of values, among them the real- and vector-valued case (mean, vector of moments, covariance matrix, expectiles), the set-valued case (confidence regions) or also the functional-valued case (distribution functions). This article is concerned with the situation where Y is a d-dimensional random vector and T is a k-dimensional functional, thus also covering the real-valued case.

It is common to assess and compare competing point forecasts in terms of a loss function or *scoring function*. This is a function S such as the squared error or the

absolute error which is negatively oriented in the following sense: If the forecast $x \in \mathbb{R}^k$ is issued and the event $y \in \mathbb{R}^d$ materializes, the forecaster is *penalized* by the real value $S(x, y)$. In the presence of several different forecasters, one can compare their performances by ranking their realized scores. Hence, forecasters have an incentive to minimize their Bayes risk or *expected loss* $\mathbb{E}_F[S(x, Y)]$. Gneiting (2011) demonstrated impressively that scoring functions should be incentive compatible in that they should encourage the forecasters to issue truthful reports; see also Engelberg, Manski and Williams (2009), Murphy and Daan (1985). In other words, the choice of the scoring function S must be *consistent* with the choice of the functional T. We say a scoring function S is *strictly $\mathcal{F}$-consistent* for a functional T if $T(F)$ is the unique minimizer of the expected score $\mathbb{E}_F[S(x, Y)]$ for all $F \in \mathcal{F}$, where the class $\mathcal{F}$ of probability distributions is the domain of T. In some parts of the literature, strictly consistent scoring functions are called proper scoring rules. Our choice of terminology is in line with Gneiting (2011). Following Lambert, Pennock and Shoham (2008) and Gneiting (2011), we call a functional T with domain $\mathcal{F}$ *elicitable* if there exists a strictly $\mathcal{F}$-consistent scoring function for T.

The elicitability of a functional allows for regression, such as quantile regression and expectile regression [Koenker (2005), Newey and Powell (1987)] and for M-estimation [Huber (1964)]. Early work on elicitability is due to Osband (1985), Osband and Reichelstein (1985). More recent advances in the one-dimensional case, that is, $k = d = 1$ are due to Gneiting (2011), Lambert (2013), Steinwart et al. (2014) with the latter showing the intimate relation between elicitability and identifiability. Under mild conditions, many important functionals are elicitable such as moments, ratios of moments, quantiles and expectiles. However, there are also relevant functionals which are not elicitable such as variance, mode, or Expected Shortfall [Gneiting (2011), Heinrich (2014), Osband (1985), Weber (2006)].

With the so-called *revelation principle* Osband (1985) [see also Gneiting (2011), Theorem 4] was one of the first to show that a functional, albeit itself not being elicitable, can be a component of an elicitable vector-valued functional. The most prominent example in this direction is that the pair (mean, variance) is elicitable despite the fact that variance itself is not. However, it is crucial for the validity of the revelation principle that there is a bijection between the pair (mean, variance) and the first two moments. Until now, it appeared as an open problem if there are elicitable functionals with non-elicitable components other than those which can be connected to a functional with elicitable components via a bijection. Frongillo and Kash (2015) conjectured that this is generally not possible. We solve this open problem and can reject their conjecture: Corollary 5.5 shows that the pair (Value at Risk, Expected Shortfall) is elicitable, subject to mild regularity assumptions, improving a recent partial result of Acerbi and Székely (2014). To the best of our knowledge, we provide the first proof of this result in full generality. In

fact, Corollary 5.4 demonstrates more generally that *spectral risk measures* with a spectral measure having finite support in $(0, 1]$ can be a component of an elicitable vector-valued functional. These results may lead to a new direction in the contemporary discussion about what risk measure is best in practice, and in particular about the importance of elicitability in risk measurement contexts [Acerbi and Székely (2014), Davis (2016), Embrechts and Hofert (2014), Emmer, Kratz and Tasche (2015)].

Complementing the question whether a functional is elicitable or not, it is interesting to determine the class of strictly consistent scoring functions for a functional, or at least to characterize necessary and sufficient conditions for the strict consistency of a scoring function. Most of the existing literature focuses on real-valued functionals meaning that $k = 1$. For the case $k > 1$, mainly linear functionals, that is, vectors of expectations of certain transformations, are classified where the only strictly consistent scoring functions are *Bregman functions* [Abernethy and Frongillo (2012), Banerjee, Guo and Wang (2005), Dawid and Sebastiani (1999), Osband and Reichelstein (1985), Savage (1971)]; for a general overview of the existing literature, we refer to Gneiting (2011). To the best of our knowledge, only Osband (1985), Lambert, Pennock and Shoham (2008) and Frongillo and Kash (2015) investigated more general cases of functionals, the latter also treating vectors of ratios of expectations as the first nonlinear functionals. In his doctoral thesis, Osband (1985) established a necessary representation for the first-order derivative of a strictly consistent scoring function with respect to the report x which connects it with identification functions. Following Gneiting (2011), we call results in the same flavor *Osband's principle*. Theorem 3.2 in this paper complements and generalizes Osband (1985), Theorem 2.1. Using our techniques, we retrieve the results mentioned above concerning the Bregman representation, however, under somewhat stronger regularity assumptions than the one in Frongillo and Kash (2015); see Proposition 4.4. On the other hand, we are able to treat a much broader class of functionals; see Proposition 4.2, Remark 4.5 and Theorem 5.2. In particular, we show that under mild richness assumptions on the class $\mathcal{F}$, any strictly $\mathcal{F}$-consistent scoring function for a vector of quantiles and/or expectiles is the sum of strictly $\mathcal{F}$-consistent one-dimensional scoring functions for each quantile/expectile; see Proposition 4.2.

The paper is organized as follows. In Section 2, we introduce notation and derive some basic results concerning the elicitability of k-dimensional functionals. Section 3 is concerned with *Osband's principle*, Theorem 3.2, and its immediate consequences. We investigate the situation where a functional is composed of elicitable components in Section 4, whereas Section 5 is dedicated to the elicitability of spectral risk measures. We end our article with a brief discussion; see Section 6. Most proofs are deferred to Section 7 and the supplementary material Fissler and Ziegel (2016).

2. Properties of higher order elicitability.

2.1. *Notation and definitions.* Following Gneiting (2011), we introduce a decision-theoretic framework for the evaluation of point forecasts. To this end, we introduce an *observation domain* $\mathsf{O} \subseteq \mathbb{R}^d$. We equip O with the Borel σ-algebra $\mathcal{O}$ using the induced topology of $\mathbb{R}^d$. We identify a Borel probability measure P on $(\mathsf{O}, \mathcal{O})$ with its cumulative distribution function (c.d.f.) $F_P \colon \mathsf{O} \to [0, 1]$ defined as $F_P(x) := P((-\infty, x] \cap \mathsf{O})$, where $(-\infty, x] = (-\infty, x_1] \times \cdots \times (-\infty, x_d]$ for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Let $\mathcal{F}$ be a class of distribution functions on $(\mathsf{O}, \mathcal{O})$. Furthermore, for some integer $k \geq 1$, let $\mathsf{A} \subseteq \mathbb{R}^k$ be an *action domain*. To shorten notation, we usually write $F \in \mathcal{F}$ for a c.d.f. and also omit to mention the σ-algebra $\mathcal{O}$.

Let $T \colon \mathcal{F} \to \mathsf{A}$ be a functional. We introduce the notation $T(\mathcal{F}) := \{x \in \mathsf{A} \colon x = T(F) \text{ for some } F \in \mathcal{F}\}$. For a set $M \subseteq \mathbb{R}^k$, we will write $\mathrm{int}(M)$ for its interior with respect to $\mathbb{R}^k$, that is, $\mathrm{int}(M)$ is the biggest open set $U \subseteq \mathbb{R}^k$ such that $U \subseteq M$. The convex hull of M is defined as

$$\mathrm{conv}(M) := \left\{ \sum_{i=1}^n \lambda_i x_i \,\Big|\, n \in \mathbb{N}, x_1, \ldots, x_n \in M, \lambda_1, \ldots, \lambda_n > 0, \sum_{i=1}^n \lambda_i = 1 \right\}.$$

We say that a function $a \colon \mathsf{O} \to \mathbb{R}$ is $\mathcal{F}$-integrable if it is F-integrable for each $F \in \mathcal{F}$. A function $g \colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ is $\mathcal{F}$-integrable if $g(x, \cdot)$ is $\mathcal{F}$-integrable for each $x \in \mathsf{A}$. If g is $\mathcal{F}$-integrable, we introduce the map

$$\bar{g} \colon \mathsf{A} \times \mathcal{F} \to \mathbb{R}, \qquad (x, F) \mapsto \bar{g}(x, F) = \int g(x, y) \, \mathrm{d}F(y).$$

Consequently, for fixed $F \in \mathcal{F}$ we can consider the function $\bar{g}(\cdot, F) \colon \mathsf{A} \to \mathbb{R}, x \mapsto \bar{g}(x, F)$, and for fixed $x \in \mathsf{A}$ we can consider the (linear) functional $\bar{g}(x, \cdot) \colon \mathcal{F} \to \mathbb{R}, F \mapsto \bar{g}(x, F)$.

If we fix $y \in \mathsf{O}$ and g is sufficiently smooth in its first argument, then for $m \in \{1, \ldots, k\}$ we denote the mth partial derivative of the function $g(\cdot, y)$ with $\partial_m g(\cdot, y)$. More formally, we set

$$\partial_m g(\cdot, y) \colon \mathrm{int}(\mathsf{A}) \to \mathbb{R}, \qquad (x_1, \ldots, x_k) \mapsto \frac{\partial}{\partial x_m} g(x_1, \ldots, x_k, y).$$

We denote by $\nabla g(\cdot, y)$ the gradient of $g(\cdot, y)$ defined as $\nabla g(\cdot, y) := (\partial_1 g(\cdot, y), \ldots, \partial_k g(\cdot, y))^\top$; and with $\nabla^2 g(\cdot, y) := (\partial_l \, \partial_m g(\cdot, y))_{l,m=1,\ldots,k}$ the Hessian of $g(\cdot, y)$. *Mutatis mutandis*, we use the same notation for $\bar{g}(\cdot, F), F \in \mathcal{F}$. We call a function on A differentiable if it is differentiable in $\mathrm{int}(\mathsf{A})$ and use the notation as given above. The restriction of a function f to some subset M of its domain is denoted by $f_{|M}$.

DEFINITION 2.1 (Consistency and elicitability). A *scoring function* is an $\mathcal{F}$-integrable function $S \colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$. It is said to be $\mathcal{F}$-*consistent* for a functional $T \colon \mathcal{F} \to \mathsf{A}$ if $\bar{S}(T(F), F) \leq \bar{S}(x, F)$ for all $F \in \mathcal{F}$ and for all $x \in \mathsf{A}$. Furthermore,

S is *strictly $\mathcal{F}$-consistent* for T if it is $\mathcal{F}$-consistent for T and if $\bar{S}(T(F), F) = \bar{S}(x, F)$ implies that $x = T(F)$ for all $F \in \mathcal{F}$ and for all $x \in \mathsf{A}$. Wherever it is convenient, we assume that $S(x, \cdot)$ is locally bounded for all $x \in \mathsf{A}$. A functional $T : \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^k$ is called *k-elicitable*, if there exists a strictly $\mathcal{F}$-consistent scoring function for T.

DEFINITION 2.2 (Identification function).   An *identification function* is an $\mathcal{F}$-integrable function $V : \mathsf{A} \times \mathsf{O} \to \mathbb{R}^k$. It is said to be an *$\mathcal{F}$-identification function* for a functional $T : \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^k$ if $\bar{V}(T(F), F) = 0$ for all $F \in \mathcal{F}$. Furthermore, V is a *strict $\mathcal{F}$-identification function* for T if $\bar{V}(x, F) = 0$ holds if and only if $x = T(F)$ for all $F \in \mathcal{F}$ and for all $x \in \mathsf{A}$. Wherever it is convenient, we assume that $V(x, \cdot)$ is locally bounded for all $x \in \mathsf{A}$ and that $V(\cdot, y)$ is locally Lebesgue-integrable for all $y \in \mathsf{O}$. A functional $T : \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^k$ is said to be *k-identifiable*, if there exists a strict $\mathcal{F}$-identification function for T.

If the dimension k is clear from the context, we say that a functional is elicitable (identifiable) instead of k-elicitable (k-identifiable).

REMARK 2.3.   Depending on the class $\mathcal{F}$, some statistical functionals such as quantiles can be set-valued. In such situations, one can define $T : \mathcal{F} \to 2^\mathsf{A}$. Then a scoring function $S : \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ is called (strictly) $\mathcal{F}$-consistent for T if $\bar{S}(t, F) \le \bar{S}(x, F)$ for all $x \in \mathsf{A}$, $F \in \mathcal{F}$ and $t \in T(F)$ [with equality implying $x \in T(F)$]. The definition of a (strict) $\mathcal{F}$-identification function for T can be generalized *mutatis mutandis*. Many of the results of this paper can be extended to the case of set-valued functionals—at the cost of a more involved notation and analysis. To allow for a clear presentation, we confine ourselves to functionals with values in $\mathbb{R}^k$ in this paper.

2.2. *Basic results.*   The first lemma gives a useful equivalent characterization of strict consistency. Its proof is a direct consequence of the definition.

LEMMA 2.4.   *A scoring function $S : \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ is strictly $\mathcal{F}$-consistent for $T : \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^k$ if and only if the function*

$$\psi : D \to \mathbb{R}, \qquad s \mapsto \bar{S}(t + sv, F)$$

has a global unique minimum at $s = 0$ for all $F \in \mathcal{F}$, $t = T(F)$ and $v \in \mathbb{S}^{k-1}$ where $D = \{s \in \mathbb{R} : t + sv \in \mathsf{A}\}$.

The following result follows directly from the definition of consistency (Definition 2.1). However, it is crucial to understand many of the results of this paper.

LEMMA 2.5.   *Let $T : \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^k$ be a functional with a strictly $\mathcal{F}$-consistent scoring function $S : \mathsf{A} \times \mathsf{O} \to \mathbb{R}$. Then the following two assertions hold:*

(i) *Let $\mathcal{F}' \subseteq \mathcal{F}$ and $T_{|\mathcal{F}'}$ be the restriction of T to $\mathcal{F}'$. Then S is also a strictly $\mathcal{F}'$-consistent scoring function for $T_{|\mathcal{F}'}$.*

(ii) *Let $\mathsf{A}' \subseteq \mathsf{A}$ such that $T(\mathcal{F}) \subseteq \mathsf{A}'$ and $S_{|\mathsf{A}' \times \mathsf{O}}$ be the restriction of S to $\mathsf{A}' \times \mathsf{O}$. Then $S_{|\mathsf{A}' \times \mathsf{O}}$ is also a strictly $\mathcal{F}$-consistent scoring function for T.*

The main results of this paper consist of necessary and sufficient conditions for the strict $\mathcal{F}$-consistency of a scoring function S for some functional T. What are the consequences of Lemma 2.5 for such conditions? Assume that we start with a functional $T': \mathcal{F}' \to \mathsf{A}' \subseteq \mathbb{R}^k$ and deduce some *necessary* conditions for a scoring function $S': \mathsf{A}' \times \mathsf{O} \to \mathbb{R}$ to be strictly $\mathcal{F}'$-consistent for T'. Then Lemma 2.5(i) implies that these conditions continue to be necessary conditions for the strict $\mathcal{F}$-consistency of S' for $T: \mathcal{F} \to \mathsf{A}'$ where $\mathcal{F}' \subseteq \mathcal{F}$, and T is some extension of T' such that $T(\mathcal{F}) \subseteq \mathsf{A}'$. On the other hand, Lemma 2.5(ii) implies that the necessary conditions for the strict $\mathcal{F}'$-consistency of a scoring function $S': \mathsf{A}' \times \mathsf{O} \to \mathbb{R}$ continue to be necessary conditions for the strict $\mathcal{F}'$-consistency of $S: \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ for T', where $\mathsf{A}' \subseteq \mathsf{A}$ and S is some extension of S'.

Summarizing, given a functional $T: \mathcal{F} \to \mathsf{A}$, a collection of necessary conditions for the strict $\mathcal{F}$-consistency of scoring functions for T is the more restrictive the smaller the class $\mathcal{F}$ and the smaller the set A is [provided that $T(\mathcal{F}) \subseteq \mathsf{A}$, of course]. Hence, in the forthcoming results concerning necessary conditions, it is no loss of generality to just mention which distributions must necessarily be in the class $\mathcal{F}$ to guarantee the validity of the results. Furthermore, it is no loss of generality to make the assumption that T is surjective, so $\mathsf{A} = T(\mathcal{F})$.

Some of the subsequent results also provide *sufficient* conditions for the strict $\mathcal{F}$-consistency of a scoring function $S: \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ for a functional $T: \mathcal{F} \to \mathsf{A}$. Those results are the stronger the bigger the class $\mathcal{F}$ and the bigger the set A is. For the notion of elicitability, this means that the assertion that a functional $T: \mathcal{F} \to \mathsf{A}$ is elicitable is also the stronger the bigger the class $\mathcal{F}$ and the bigger the set A is. To demonstrate this reasoning, observe that if the functional $T: \mathcal{F} \to \mathsf{A}$ is degenerate in the sense that it is constant, so $T \equiv t$ for some $t \in \mathsf{A}$ (which covers the particular case that $\mathcal{F}$ contains only one element), then T is automatically elicitable with a strictly $\mathcal{F}$-consistent scoring function $S: \mathsf{A} \times \mathsf{O} \to \mathbb{R}$, defined as $S(x, y) := \|x - t\|$.

As a last result in this section, we present the intuitive observation that a vector of elicitable functionals itself is elicitable.

LEMMA 2.6. *Let $k_1, \ldots, k_l \geq 1$ and let $T_m: \mathcal{F} \to \mathsf{A}_m \subseteq \mathbb{R}^{k_m}$ be a k_m-elicitable functional, $m \in \{1, \ldots, l\}$. Then the functional $T = (T_1, \ldots, T_l): \mathcal{F} \to \mathsf{A}$ is k-elicitable where $k = k_1 + \cdots + k_l$ and $\mathsf{A} = \mathsf{A}_1 \times \cdots \times \mathsf{A}_l \subseteq \mathbb{R}^k$.*

PROOF. For $m \in \{1, \ldots, l\}$, let $S_m: \mathsf{A}_m \times \mathsf{O} \to \mathbb{R}$ be a strictly $\mathcal{F}$-consistent scoring function for T_m. Let $\lambda_1, \ldots, \lambda_l > 0$ be positive real numbers. Then

$$(2.1) \qquad S(x_1, \ldots, x_l, y) := \sum_{m=1}^{l} \lambda_m S_m(x_m, y)$$

is a strictly $\mathcal{F}$-consistent scoring function for T. $\quad\square$

A particularly simple and relevant case of Lemma 2.6 is the situation $k_1 = \cdots = k_l = 1$ such that $k = l$. It is an interesting question whether the scoring functions of the form (2.1) are the only strictly $\mathcal{F}$-consistent scoring functions for T, which amounts to the question of *separability* of scoring rules that was posed by Frongillo and Kash (2015). The answer is generally negative. As mentioned in the Introduction, it is known that all Bregman functions elicit T, if the components of T are all expectations of transformations of Y [Abernethy and Frongillo (2012), Banerjee, Guo and Wang (2005), Dawid and Sebastiani (1999), Osband and Reichelstein (1985), Savage (1971)] or ratios of expectations with the same denominator [Frongillo and Kash (2015)]; see also Proposition 4.4. However, for other situations, such as a combination of different quantiles and/or expectiles, the answer is positive; see Proposition 4.2. These results rely on "Osband's principle" which gives necessary conditions for scoring functions to be strictly $\mathcal{F}$-consistent for a given functional T; see Section 3.

There are more involved functionals that are k-elicitable than combinations of k 1-elicitable components. An immediate example that is the pair (expectation, variance) which can be obtained through the revelation principle from the 2-elicitable pair (expectation, second moment). In Section 5, we show that the concept of k-elicitability is also not restricted to functionals that can be obtained by combining Lemma 2.6 and the revelation principle. It is shown in Weber (2006), Example 3.4 and Gneiting (2011), Theorem 11, that the coherent risk measure Expected Shortfall at level α, $\alpha \in (0, 1)$, does not have convex level sets and is therefore not elicitable. In contrast, we show in Corollary 5.5 that the pair (Value at Risk$_\alpha$, Expected Shortfall$_\alpha$) is 2-elicitable relative to the class of distributions on $\mathbb{R}$ with finite first moment and unique α-quantiles. This refutes Proposition 2.3 of Osband (1985); see Remark 5.3 for a discussion.

3. Osband's principle. In this section, we give necessary conditions for the strict $\mathcal{F}$-consistency of a scoring function S for a functional $T: \mathcal{F} \to \mathsf{A}$. In the light of Lemma 2.5 and the discussion thereafter, we have to impose some richness conditions on the class $\mathcal{F}$ as well as on the "variability" of the functional T. To this end, we establish a link between strictly $\mathcal{F}$-consistent scoring functions and strict $\mathcal{F}$-identification functions. We illustrate the idea in the one-dimensional case. Let $\mathcal{F}$ be a class of distribution functions on $\mathbb{R}$, $T: \mathcal{F} \to \mathbb{R}$ a functional and $S: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ a strictly $\mathcal{F}$-consistent scoring function for T. Furthermore, let $V: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be an oriented strict $\mathcal{F}$-identification function for T. Then, under certain regularity conditions, there is a nonnegative function $h: \mathbb{R} \to \mathbb{R}$ such that

$$(3.1) \qquad \frac{\mathrm{d}}{\mathrm{d}x} S(x, y) = h(x) V(x, y).$$

If we naïvely swap differentiation and expectation and h does not vanish, the form (3.1) plus the identification property of V are sufficient for the first order

condition on $\bar{S}(\cdot, F)$, $F \in \mathcal{F}$, to be satisfied and the orientation of V (see Remark 4.1) as well as the fact that h is positive are sufficient for $\bar{S}(\cdot, F)$ to satisfy the second-order condition for strict $\mathcal{F}$-consistency. So the really interesting part is to show that the form given in (3.1) is necessary for the strict $\mathcal{F}$-consistency of a scoring function for T.

The idea of this characterization originates from Osband (1985). He gives a characterization including $\mathbb{R}^k$-valued functionals, but for his proof he assumes that $\mathcal{F}$ contains all distributions with finite support. This is not a problem *per se*, but in the light of Lemma 2.5 and the discussion thereafter it is desirable to weaken this assumption. In particular, the results in Section 5 on spectral risk measures cannot be derived if $\mathcal{F}$ has to contain all distributions with finite support. Relying on a functional space extension of the Kuhn–Tucker theorem, Osband (1985) conjectures that his characterization continues to hold if $\mathcal{F}$ consists only of absolutely continuous distributions, but we do not believe that his approach is feasible in this case. In Steinwart et al. (2014), Theorem 5, there is a rigorous statement of Osband's principle for the one-dimensional functionals where the distributions in $\mathcal{F}$ must be absolutely continuous with respect to some finite measure. We shall give a proof in the setting of an $\mathbb{R}^k$-valued functional that does not have to specify the kinds of distributions in $\mathcal{F}$, but only uses the following (minimal) collection of regularity assumptions. To this end, we apply a similar technique as in the proof of Osband (1985), Lemma 2.2, which is based on a finite-dimensional argument.

Let $\mathcal{F}$ be a class of distribution functions on $\mathsf{O} \subseteq \mathbb{R}^d$. Fix a functional $T: \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^k$, an identification function $V: \mathsf{A} \times \mathsf{O} \to \mathbb{R}^k$ and a scoring function $S: \mathsf{A} \times \mathsf{O} \to \mathbb{R}$.

ASSUMPTION (V1). Let $\mathcal{F}$ be a convex class of distributions functions on $\mathsf{O} \subseteq \mathbb{R}^d$ and assume that for every $x \in \mathrm{int}(\mathsf{A})$ there are $F_1, \ldots, F_{k+1} \in \mathcal{F}$ such that

$$0 \in \mathrm{int}\big(\mathrm{conv}\big(\{\bar{V}(x, F_1), \ldots, \bar{V}(x, F_{k+1})\}\big)\big).$$

REMARK 3.1. Assumption (V1) implies that for every $x \in \mathrm{int}(\mathsf{A})$ there are $F_1, \ldots, F_k \in \mathcal{F}$ such that the vectors $\bar{V}(x, F_1), \ldots, \bar{V}(x, F_k)$ are linearly independent.

Assumption (V1) ensures that the class $\mathcal{F}$ is "rich" enough meaning that the functional T varies sufficiently in order to derive a necessary form of the scoring function S in Theorem 3.2. Assumptions like (V1) are classical in the literature. For the case of k-elicitability, Osband (1985) assumes that $0 \in \mathrm{int}(\mathrm{conv}(\{V(x, y): y \in \mathsf{O}\}))$. Steinwart et al. (2014), Definition 8 and Lambert (2013) treat the case $k = 1$ and work under the assumption that the functional is *strictly locally nonconstant* which implies assumption (V1) if the functional is identifiable.

ASSUMPTION (V2). For every $F \in \mathcal{F}$, the function $\bar{V}(\cdot, F) : \mathsf{A} \to \mathbb{R}^k$, $x \mapsto \bar{V}(x, F)$, is continuous.

ASSUMPTION (V3). For every $F \in \mathcal{F}$, the function $\bar{V}(\cdot, F)$ is continuously differentiable.

If the function $x \mapsto V(x, y)$, $y \in \mathsf{O}$, is continuous (continuously differentiable), assumption (V2) [assumption (V3)] is satisfied, and it is equivalent to (V2) [(V3)] if $\mathcal{F}$ contains all measures with finite support. However, (V2) and (V3) are much weaker requirements if we move away from distributions with finite support. To illustrate this fact, let $k, d = 1$ and $V(x, y) = \mathbb{1}\{y \le x\} - \alpha$, $\alpha \in (0, 1)$, which is a strict $\mathcal{F}$-identification function for the α-quantile. Of course, $V(\cdot, y)$ is not continuous. But if $\mathcal{F}$ contains only probability distributions F that have a continuous derivative $f = F'$, then $\bar{V}(x, F) = F(x) - \alpha$ and $(\mathrm{d}/\mathrm{d}x)\bar{V}(x, F) = f(x)$ and V satisfies (V2) and (V3). The following assumptions (S1) and (S2) are similar conditions as (V2) and (V3) but for scoring functions instead of identification functions.

ASSUMPTION (S1). For every $F \in \mathcal{F}$, the function $\bar{S}(\cdot, F) : \mathsf{A} \to \mathbb{R}$, $x \mapsto \bar{S}(x, F)$, is continuously differentiable.

ASSUMPTION (S2). For every $F \in \mathcal{F}$, the function $\bar{S}(\cdot, F)$ is continuously differentiable and the gradient is locally Lipschitz continuous. Furthermore, $\bar{S}(\cdot, F)$ is twice continuously differentiable at $t = T(F) \in \mathrm{int}(\mathsf{A})$.

Note that assumption (S2) implies that the gradient of $\bar{S}(\cdot, F)$ is (totally) differentiable for almost all $x \in \mathsf{A}$ by Rademacher's theorem, which in turn indicates that the Hessian of $\bar{S}(\cdot, F)$ exists for almost all $x \in \mathsf{A}$ and is symmetric by Schwarz's theorem; see Grauert and Fischer (1978), page 57.

THEOREM 3.2 (Osband's principle). *Let* $T : \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^k$ *be a surjective, elicitable and identifiable functional with a strict* $\mathcal{F}$-*identification function* $V : \mathsf{A} \times \mathsf{O} \to \mathbb{R}^k$ *and a strictly* $\mathcal{F}$-*consistent scoring function* $S : \mathsf{A} \times \mathsf{O} \to \mathbb{R}$. *If the assumptions* (V1) *and* (S1) *hold, then there exists a matrix-valued function* $h : \mathrm{int}(\mathsf{A}) \to \mathbb{R}^{k \times k}$ *such that for* $l \in \{1, \dots, k\}$

$$(3.2) \qquad \partial_l \bar{S}(x, F) = \sum_{m=1}^{k} h_{lm}(x) \bar{V}_m(x, F)$$

for all $x \in \mathrm{int}(\mathsf{A})$ *and* $F \in \mathcal{F}$. *If in addition, assumption* (V2) *holds, then* h *is continuous. Under the additional assumptions* (V3) *and* (S2), *the function* h *is locally Lipschitz continuous.*

Under the conditions of Theorem 3.2, equation (3.2) gives a characterization of the partial derivatives of the expected score. If we impose more smoothness assumptions on the expected score, we are also able to give a characterization of the second-order derivatives of the expected score.

COROLLARY 3.3. *For a surjective, elicitable and identifiable functional $T: \mathcal{F} \to A \subseteq \mathbb{R}^k$ with a strict $\mathcal{F}$-identification function $V: A \times O \to \mathbb{R}^k$ and a strictly $\mathcal{F}$-consistent scoring function $S: A \times O \to \mathbb{R}$ that satisfy assumptions* (V1), (V3) *and* (S2), *we have the following identities for the second-order derivatives*:

$$\partial_m \partial_l \bar{S}(x, F) = \sum_{i=1}^{k} \partial_m h_{li}(x) \bar{V}_i(x, F) + h_{li}(x) \partial_m \bar{V}_i(x, F)$$

(3.3)

$$= \sum_{i=1}^{k} \partial_l h_{mi}(x) \bar{V}_i(x, F) + h_{mi}(x) \partial_l \bar{V}_i(x, F) = \partial_l \partial_m \bar{S}(x, F),$$

for all $l, m \in \{1, \ldots, k\}$, for all $F \in \mathcal{F}$ and almost all $x \in \mathrm{int}(A)$, where h is the matrix-valued function appearing at (3.2). *In particular,* (3.3) *holds for $x = T(F) \in \mathrm{int}(A)$.*

Theorem 3.2 and Corollary 3.3 establish necessary conditions for strictly $\mathcal{F}$-consistent scoring functions on the level of the expected scores. If the class $\mathcal{F}$ is rich enough and the scoring and identification function smooth enough in the following sense, we can also deduce a necessary condition for S which holds pointwise.

ASSUMPTION (F1). For every $y \in O$, there exists a sequence $(F_n)_{n \in \mathbb{N}}$ of distributions $F_n \in \mathcal{F}$ that converges weakly to the Dirac-measure δ_y such that the support of F_n is contained in a compact set K for all n.

ASSUMPTION (VS1). Suppose that the complement of the set

$$C := \{(x, y) \in A \times O | \ V(x, \cdot) \text{ and } S(x, \cdot) \text{ are continuous at the point } y\}$$

has $(k + d)$-dimensional Lebesgue measure zero.

PROPOSITION 3.4. *Assume that $\mathrm{int}(A) \subseteq \mathbb{R}^k$ is a star domain and let $T: \mathcal{F} \to A$ be a surjective, elicitable and identifiable functional with a strict $\mathcal{F}$-identification function $V: A \times O \to \mathbb{R}^k$ and a strictly $\mathcal{F}$-consistent scoring function $S: A \times O \to \mathbb{R}$. Suppose that assumptions* (V1), (V2), (S1), (F1) *and* (VS1) *hold. Let h be the matrix valued function appearing at* (3.2). *Then the scoring function S is necessarily of the form*

$$S(x, y) = \sum_{r=1}^{k} \sum_{m=1}^{k} \int_{z_r}^{x_r} h_{rm}(x_1, \ldots, x_{r-1}, v, z_{r+1}, \ldots, z_k)$$

(3.4)

$$\times V_m(x_1, \ldots, x_{r-1}, v, z_{r+1}, \ldots, z_k, y) \, dv + a(y)$$

for almost all $(x, y) \in \mathsf{A} \times \mathsf{O}$, *for some star point* $z = (z_1, \ldots, z_k) \in \text{int}(\mathsf{A})$ *and some* $\mathcal{F}$-*integrable function* $a: \mathsf{O} \to \mathbb{R}$. *On the level of the expected score* $\bar{S}(x, F)$, *equation* (3.4) *holds for all* $x \in \text{int}(A)$, $F \in \mathcal{F}$.

While Theorem 3.2, Corollary 3.3 and Proposition 3.4 only establish necessary conditions for strictly $\mathcal{F}$-consistent scoring functions for some functional T, often they guide a way how to construct strictly $\mathcal{F}$-consistent scoring functions starting with a strict $\mathcal{F}$-identification function V for T.

For the one-dimensional case, one can use the fact that, subject to some mild regularity conditions, if V is a strict $\mathcal{F}$-identification function, then either V or $-V$ is oriented; see Remark 4.1. Supposing that V is oriented, we can choose any strictly positive function $h: \mathsf{A} \to \mathbb{R}$ to get the derivative of a strictly $\mathcal{F}$-consistent scoring function. Then integration yields the desired strictly $\mathcal{F}$-consistent scoring function.

Establishing *sufficient* conditions for scoring functions to be strictly $\mathcal{F}$-consistent for T is generally more involved in the case $k > 1$. First of all, working under assumption (S2), the symmetry of the Hessian $\nabla^2 \bar{S}(x, F)$ imposes strong necessary conditions on the functions h_{lm}; see, for example, Proposition 4.2 which treats the case where all components of the functional $T = (T_1, \ldots, T_k)$ are elicitable and identifiable. The example of spectral risk measures is treated in Section 5. Second, (3.2) and (3.3) are necessary conditions for $\bar{S}(x, F)$ having a *local* minimum in $x = T(F)$, $F \in \mathcal{F}$. Even if we additionally suppose that the Hessian $\nabla^2 \bar{S}(x, F)$ is strictly positive definite at $x = T(F)$, this is a sufficient condition only for a *local* minimum at $x = T(F)$, but does not provide any information concerning a *global* minimum. Consequently, even if the functions h_{lm} satisfy (3.3), one must verify the strict consistency of the scoring function on a case by case basis. This can often be done by showing that the one-dimensional functions $\mathbb{R} \to \mathbb{R}$, $s \mapsto \bar{S}(t + sv, F)$, with $t = T(F)$, have a global minimum in $s = 0$ for all $v \in \mathbb{S}^{k-1}$ and for all $F \in \mathcal{F}$.

4. Functionals with elicitable components.

Suppose that the functional $T = (T_1, \ldots, T_k): \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^k$ consists of 1-elicitable components T_m. As prototypical examples of such 1-elicitable components, we consider the functionals given in Table 1 where we implicitly assume that $\mathsf{O} \subseteq \mathbb{R}$ if a quantile or an expectile are a part of T. If V_m are strict $\mathcal{F}$-identification functions for T_m then $V: \mathsf{A} \times \mathbb{R} \to \mathbb{R}^k$ with

$$(4.1) \qquad V(x_1, \ldots, x_k, y) = \big(V_1(x_1, y), \ldots, V_k(x_k, y)\big)^\top$$

is a strict $\mathcal{F}$-identification function for T. Under (V3), the partial derivatives of $\bar{V}(x, F)$, $x \in \mathsf{A}$ and $F \in \mathcal{F}$ exist, and if the class $\mathcal{F}$ is sufficiently rich T (or some subset of its components) often fulfills the following assumption.

TABLE 1
Strict identification functions for $k = 1$; see Gneiting (2011), Table 9

Functional	Strict identification function
Ratio $\mathbb{E}_F[p(Y)]/\mathbb{E}_F[q(Y)]$	$V(x, y) = xq(y) - p(y)$
α-Quantile	$V(x, y) = \mathbb{1}\{y \leq x\} - \alpha$
τ-Expectile	$V(x, y) = 2|\mathbb{1}\{y \leq x\} - \tau|(x - y)$

ASSUMPTION (V4). Let assumption (V3) hold. For all $r \in \{1, \ldots, k\}$ and for all $t \in \text{int}(\mathsf{A}) \cap T(\mathcal{F})$ there are $F_1, F_2 \in T^{-1}(\{t\})$ such that

$$\partial_l \bar{V}_l(t, F_1) = \partial_l \bar{V}_l(t, F_2) \qquad \forall l \in \{1, \ldots, k\} \setminus \{r\}, \qquad \partial_r \bar{V}_r(t, F_1) \neq \partial_r \bar{V}_r(t, F_2).$$

The following proposition gives a characterization of the class of strictly $\mathcal{F}$-consistent scoring functions under (V4). In particular, the result covers vectors of different quantiles and/or different expectiles (with the exception of the 1/2-expectile), thus answering a question raised in Gneiting and Raftery (2007), page 370.

One relevant exception when (V4) is not satisfied is when T is a ratio of expectations with the same denominator, that is, $q_m = q$ for all m. We treat this case in Proposition 4.4 below.

REMARK 4.1. Steinwart et al. (2014) introduced the notion of an *oriented* strict $\mathcal{F}$-identification function for the case $k = 1$ and $d = 1$. They say that $V: \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ is an oriented strict $\mathcal{F}$-identification function for the functional $T: \mathcal{F} \to \mathsf{A}$ if V is a strict $\mathcal{F}$-identification function for T and, moreover, $\bar{V}(x, F) > 0$ if and only if $x > T(F)$ for all $F \in \mathcal{F}$ and for all $x \in \mathsf{A}$.

PROPOSITION 4.2. *Let $T_m: \mathcal{F} \to \mathsf{A}_m \subseteq \mathbb{R}$ be 1-elicitable and 1-identifiable functionals with oriented strict $\mathcal{F}$-identification functions $V_m: \mathsf{A}_m \times \mathsf{O} \to \mathbb{R}$ for $m \in \{1, \ldots, k\}$. Define $T = (T_1, \ldots, T_k)$ with identification function V as at (4.1) and a strictly $\mathcal{F}$-consistent scoring function $S: \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ with $\mathsf{A} := T(\mathcal{F}) \subseteq \mathsf{A}_1 \times \cdots \times \mathsf{A}_k$. Suppose that $\text{int}(\mathsf{A})$ is a star domain, and assumptions (V1), (V3), (V4), (S2) hold. Define $\mathsf{A}'_m := \{x_m : \exists (z_1, \ldots, z_k) \in \text{int}(A), z_m = x_m\}$.*

(i) *Let $h: \text{int}(\mathsf{A}) \to \mathbb{R}^{k \times k}$ be the function given at (3.2). Then there are functions $g_m: \mathsf{A}'_m \to \mathbb{R}$, $g_m > 0$, such that $h_{mm}(x_1, \ldots, x_k) = g_m(x_m)$ for all $m \in \{1, \ldots, k\}$ and $(x_1, \ldots, x_k) \in \text{int}(\mathsf{A})$ and*

$$(4.2) \qquad h_{rl}(x) = 0$$

for all $r, l \in \{1, \ldots, k\}, l \neq r$, and for all $x \in \text{int}(\mathsf{A})$.

(ii) *Assume that* (F1) *and* (VS1) *hold. Then S is strictly $\mathcal{F}$-consistent for T if and only if it is of the form*

$$S(x_1, \ldots, x_k, y) = \sum_{m=1}^{k} S_m(x_m, y), \tag{4.3}$$

for almost all $(x, y) \in \mathsf{A} \times \mathsf{O}$, where $S_m : \mathsf{A}_m \times \mathsf{O} \to \mathbb{R}$, $m \in \{1, \ldots, k\}$, are strictly $\mathcal{F}$-consistent scoring functions for T_m.

REMARK 4.3. Lambert, Pennock and Shoham (2008), Theorem 5, show that a scoring function is accuracy-rewarding if and only if it is the sum of strictly consistent scoring functions for each component. Their assumptions are different from Proposition 4.2(ii). For example, they assume that all distributions in $\mathcal{F}$ have finite support, and that scoring functions are twice continuously differentiable. Therefore, despite the same form of the characterization in (4.3), neither result implies the other. However, the key components of the proofs of both results is to show that the cross-derivatives of the expected scoring functions are zero. This implies then a decomposition as in (4.3). The converse is trivial in both cases.

If T is a ratio of expectations with the same denominator, it is well known that the class of strictly $\mathcal{F}$-consistent scoring functions is bigger than the one given in Proposition 4.2(ii).

PROPOSITION 4.4. *Let $T : \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^k$ be a ratio of expectations with the same denominator, that is, $T(F) = \mathbb{E}_F[p(Y)]/\mathbb{E}_F[q(Y)]$ for some $\mathcal{F}$-integrable functions $p : \mathsf{O} \to \mathbb{R}^k$, $q : \mathsf{O} \to \mathbb{R}$. Assume that $\bar{q}(F) > 0$ for all $F \in \mathcal{F}$ and let $V : \mathsf{A} \times \mathsf{O} \to \mathbb{R}^k$, $V(x, y) = q(y)x - p(y)$. Let $S : \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ be a strictly $\mathcal{F}$-consistent scoring function for T and $h : \mathrm{int}(\mathsf{A}) \to \mathbb{R}^{k \times k}$ be the function given at (3.2). Suppose that T is surjective, and assumptions* (V1), (V3), (S2) *hold.*

(i) *It holds that*

$$\partial_l h_{rm}(x) = \partial_r h_{lm}(x), \qquad h_{rl}(x) = h_{lr}(x) \tag{4.4}$$

for all $r, l, m \in \{1, \ldots, k\}$, $l \neq r$, where the first identity holds for almost all $x \in \mathrm{int}(\mathsf{A})$ and the second identity for all $x \in \mathrm{int}(\mathsf{A})$. Moreover, the matrix $(h_{rl}(x))_{r,l=1,\ldots,k}$ is positive definite for all $x \in \mathrm{int}(A)$.

(ii) *Let $\mathrm{int}(\mathsf{A})$ be a star domain and assume that* (F1) *and* (VS1) *hold. Then S is strictly $\mathcal{F}$-consistent for T if and only if it is of the form*

$$S(x, y) = -\phi(x)q(y) + \sum_{m=1}^{k} (q(y)x_m - p_m(y))\partial_m \phi(x) + a(y), \tag{4.5}$$

with

$$\phi(x) = \sum_{r=1}^{k} \int_{z_r}^{x_r} \int_{z_r}^{v} h_{rr}(x_1, \ldots, x_{r-1}, w, z_{r+1}, \ldots, z_k) \, dw \, dv, \tag{4.6}$$

for almost all $(x, y) \in \mathsf{A} \times \mathsf{O}$, *where* $(z_1, \ldots, z_k) \in \operatorname{int}(\mathsf{A})$ *is some star point and* $a: \mathsf{O} \to \mathbb{R}$ *is* $\mathcal{F}$-*integrable. Moreover,* ϕ *has Hessian* h *and is strictly convex.*

Part (ii) of this proposition recovers results of Abernethy and Frongillo (2012), Banerjee, Guo and Wang (2005), Osband and Reichelstein (1985) if $q \equiv 1$, which show that all consistent scoring functions for vectors of expectations are so-called *Bregman functions*, that is, functions of the form (4.5) with $q \equiv 1$ and a convex function ϕ. Frongillo and Kash (2015), Theorem 13, also treat the case of more general functions q.

REMARK 4.5. One might wonder about necessary conditions on the matrix-valued function h in the flavor of Propositions 4.2(i) and 4.4(i) if the k components of the functional T can be regrouped into (a) a new functional $T_1': \mathcal{F} \to \mathsf{A}_1' \subset \mathbb{R}^{k_1'}$ with an oriented strict $\mathcal{F}$-identification function $V_1': \mathsf{A}_1' \times \mathsf{O} \to \mathbb{R}^{k_1'}$ which satisfies assumption (V4), and (b) several, say l, new functionals $T_m': \mathcal{F} \to \mathsf{A}_{k_m'}' \subseteq \mathbb{R}^{k_m'}$, $m \in \{2, \ldots, l+1\}$ which are ratios of expectations with the same denominator, and $k_1' + \cdots + k_{l+1}' = k$. We can apply the propositions to obtain necessary conditions for each of the $(k_m' \times k_m')$-valued functions h_m', $m \in \{1, \ldots, l+1\}$. Applying Lemma 2.6, we get a possible choice for a strictly $\mathcal{F}$-consistent scoring function S for T. On the level of the $k \times k$-valued function h associated to S this means that h is a block diagonal matrix of the form $\operatorname{diag}(h_1', \ldots, h_{l+1}')$. But what about the necessity of this form? Indeed, if we assume that the blocks in (b) have maximal size (or equivalently that l is minimal) then one can verify that h must be necessarily of the block diagonal form described above.

5. Spectral risk measures.
Risk measures are a common tool to measure the risk of a financial position Y. A risk measure is usually defined as a mapping ρ from some space of random variables, for example, L^∞, to the real line. Arguably, the most common risk measure in practice is *Value at Risk* at level α ($\operatorname{VaR}_\alpha$) which is the generalized α-quantile $F^{-1}(\alpha)$, that is,

$$\operatorname{VaR}_\alpha(Y) := F^{-1}(\alpha) := \inf\{x \in \mathbb{R}: F(x) \geq \alpha\},$$

where F is the distribution function of Y. An important alternative to $\operatorname{VaR}_\alpha$ is *Expected Shortfall* at level α (ES_α) (also known under the names Conditional Value at Risk or Average Value at Risk). It is defined as

$$(5.1) \qquad \operatorname{ES}_\alpha(Y) := \frac{1}{\alpha} \int_0^\alpha \operatorname{VaR}_u(Y) \, \mathrm{d}u, \qquad \alpha \in (0, 1],$$

and $\operatorname{ES}_0(Y) = \operatorname{ess\,inf} Y$. Since the influential paper of Artzner et al. (1999) introducing coherent risk measures, there has been a lively debate about which risk measure is best in practice, one of the requirements under discussion being the *coherence* of a risk measure. We call a functional ρ *coherent* if it is monotone,

meaning that $Y \le X$ a.s. implies that $\rho(Y) \le \rho(X)$; it is super-additive in the sense that $\rho(X + Y) \ge \rho(X) + \rho(Y)$; it is positively homogeneous which means that $\rho(\lambda Y) = \lambda \rho(Y)$ for all $\lambda \ge 0$; and it is translation invariant which amounts to $\rho(Y + a) = \rho(Y) + a$ for all $a \in \mathbb{R}$. In the literature on risk measures, there are different sign conventions which co-exist. In this paper, a positive value of Y denotes a profit. Moreover, the position Y is considered the more risky the smaller $\rho(Y)$ is. Strictly speaking, we have chosen to work with utility functions instead of risk measures as, for example, in Delbaen (2012). The risk measure ρ is called *comonotonically additive* if $\rho(X + Y) = \rho(X) + \rho(Y)$ for comonotone random variables X and Y. Coherent and comonotonically additive risk measures are also called *spectral* risk measures [Acerbi (2002)]. All risk measures of practical interest are law-invariant, that is, if two random variables X and Y have the same law F, then $\rho(X) = \rho(Y)$. As we are only concerned with law-invariant risk measures in this paper, we will abuse notation and write $\rho(F) := \rho(X)$, if X has distribution F.

One of the main criticisms on VaR_α is its failure to fulfill the super-additivity property in general [Acerbi (2002)]. Furthermore, it fails to take the size of losses beyond the level α into account [Daníelsson et al. (2001)]. In both of these aspects, ES_α is a better alternative as it is coherent and comonotonically additive, that is, a spectral risk measure. However, with respect to robustness, some authors argue that VaR_α should be preferred over ES_α [Cont, Deguest and Scandolo (2010), Kou, Peng and Heyde (2013)], whereas others argue that the classical statistical notions of robustness are not necessarily appropriate in a risk measurement context [Krätschmer, Schied and Zähle (2012, 2015, 2014)]. Finally, ES_α fails to be 1-elicitable [Gneiting (2011), Weber (2006)], whereas VaR_α is 1-elicitable for most classes of distributions $\mathcal{F}$ of practical relevance. In fact, except for the expectation, all spectral risk measures fail to be 1-elicitable [Ziegel (2014)]; further recent results on elicitable risk measures include [Kou and Peng (2014), Wang and Ziegel (2015)] showing that distortion risk measures are rarely elicitable and [Bellini and Bignozzi (2015), Delbaen et al. (2016), Weber (2006)] demonstrating that convex risk measures are only elicitable if they are shortfall risk measures.

We show in Theorem 5.2 (see also Corollaries 5.4 and 5.5) that spectral risk measures having a spectral measure with finite support can be a component of a k-elicitable functional. In particular, the pair $(\text{VaR}_\alpha, \text{ES}_\alpha): \mathcal{F} \to \mathbb{R}^2$ is 2-elicitable for any $\alpha \in (0, 1)$ subject to mild conditions on the class $\mathcal{F}$. We remark that our results substantially generalize the result of Acerbi and Székely (2014) as detailed below.

DEFINITION 5.1 (Spectral risk measures). Let μ be a probability measure on $[0, 1]$ (called *spectral measure*) and let $\mathcal{F}$ be a class of distribution functions on $\mathbb{R}$ with finite first moments. Then the *spectral risk measure* associated to μ is the

functional $\nu_\mu \colon \mathcal{F} \to \mathbb{R}$ defined as

$$\nu_\mu(F) := \int_{[0,1]} \mathrm{ES}_\alpha(F) \mu(\mathrm{d}\alpha).$$

Jouini, Schachermayer and Touzi (2006), Kusuoka (2001) have shown that law-invariant coherent and comonotonically additive risk measures are exactly the spectral risk measures in the sense of Definition 5.1 for distributions with compact support. If $\mu = \delta_\alpha$ for some $\alpha \in [0, 1]$, then $\nu_\mu(F) = \mathrm{ES}_\alpha(F)$. In particular, $\nu_{\delta_1}(F) = \int y \, \mathrm{d}F(y)$ is the expectation of F.

In the following theorem, we show that spectral risk measures whose spectral measure μ has finite support in $(0, 1)$ are k-elicitable for some k. The key to finding the form of the strictly consistent scoring functions at (5.2) is the observation that spectral risk measures jointly with the correct quantiles are identifiable with identification function given at (5.4). It is possible to extend the result to spectral measures with finite support in $(0, 1]$; see Corollary 5.4.

THEOREM 5.2. *Let $\mathcal{F}$ be a class of distribution functions on $\mathbb{R}$ with finite first moments. Let $\nu_\mu \colon \mathcal{F} \to \mathbb{R}$ be a spectral risk measure where μ is given by*

$$\mu = \sum_{m=1}^{k-1} p_m \delta_{q_m},$$

with $p_m \in (0, 1]$, $\sum_{m=1}^{k-1} p_m = 1$, $q_m \in (0, 1)$ and the q_m's are pairwise distinct. Define the functional $T = (T_1, \ldots, T_k) \colon \mathcal{F} \to \mathbb{R}^k$, where $T_m(F) := F^{-1}(q_m)$, $m \in \{1, \ldots, k-1\}$, and $T_k(F) := \nu_\mu(F)$. Then the following assertions are true:

(i) *If the distributions in $\mathcal{F}$ have unique q_m-quantiles, $m \in \{1, \ldots, k-1\}$, then the functional T is k-elicitable with respect to $\mathcal{F}$.*

(ii) *Let $\mathsf{A} \supseteq T(\mathcal{F})$ be convex and set $\mathsf{A}'_r := \{x_r \colon \exists (z_1, \ldots, z_k) \in \mathsf{A}, x_r = z_r\}$, $r \in \{1, \ldots, k\}$. Define the scoring function $S \colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ by*

$$S(x, y) = \sum_{r=1}^{k-1} \big(\mathbb{1}\{y \le x_r\} - q_r\big) G_r(x_r) - \mathbb{1}\{y \le x_r\} G_r(y)$$

$$\text{(5.2)} \qquad + G_k(x_k)\left(x_k + \sum_{m=1}^{k-1} \frac{p_m}{q_m}\big(\mathbb{1}\{y \le x_m\}(x_m - y) - q_m x_m\big)\right)$$

$$- \mathcal{G}_k(x_k) + a(y),$$

where $a \colon \mathbb{R} \to \mathbb{R}$ is $\mathcal{F}$-integrable, $G_r \colon \mathsf{A}'_r \to \mathbb{R}$, $r \in \{1, \ldots, k\}$, $\mathcal{G}_k \colon \mathsf{A}'_k \to \mathbb{R}$ with $\mathcal{G}'_k = G_k$ and for all $r \in \{1, \ldots, k\}$ and all $x_r \in \mathsf{A}'_r$ the functions $\mathbb{1}_{(\infty, x_r]} G_r$ are $\mathcal{F}$-integrable.

If $\mathcal{G}_k$ is convex and for all $r \in \{1, \ldots, k-1\}$ and $x_k \in \mathsf{A}'_k$, the function

$$\text{(5.3)} \qquad \mathsf{A}'_{r, x_k} \to \mathbb{R}, \qquad x_r \mapsto x_r \frac{p_r}{q_r} G_k(x_k) + G_r(x_r)$$

with $\mathsf{A}'_{r,x_k} := \{x_r : \exists (z_1, \ldots, z_k) \in \mathsf{A}, x_r = z_r, x_k = z_k\}$ *is increasing, then S is $\mathcal{F}$-consistent for T. If additionally the distributions in $\mathcal{F}$ have unique q_m-quantiles, $m \in \{1, \ldots, k-1\}$, $\mathcal{G}_k$ is strictly convex and the functions given at* (5.3) *are strictly increasing, then S is strictly $\mathcal{F}$-consistent for T.*

(iii) *Assume the elements of $\mathcal{F}$ have unique q_m-quantiles, $m \in \{1, \ldots, k-1\}$ and continuous densities. Define the function $V : \mathsf{A} \times \mathbb{R} \to \mathbb{R}^k$ with components*

$$V_m(x_1, \ldots, x_k, y) = \mathbb{1}\{y \le x_m\} - q_m, \qquad m \in \{1, \ldots, k-1\},$$

(5.4)

$$V_k(x_1, \ldots, x_k, y) = x_k - \sum_{m=1}^{k-1} \frac{p_m}{q_m} y \mathbb{1}\{y \le x_m\}.$$

Then V is a strict $\mathcal{F}$-identification function for T satisfying assumption (V3).

If additionally the interior of $\mathsf{A} := T(\mathcal{F}) \subseteq \mathbb{R}^k$ is a star domain, (V1) *and* (F1) *hold, and $(V_1, \ldots, V_{k-1})$ satisfies* (V4), *then every strictly $\mathcal{F}$-consistent scoring function $S : \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ for T satisfying* (S2), (VS1) *is necessarily of the form given at* (5.2) *almost everywhere. Additionally, $\mathcal{G}_k$ must be strictly convex and the functions at* (5.3) *must be strictly increasing.*

REMARK 5.3. According to Theorem 5.2, the pair $(\mathrm{VaR}_\alpha(F), \mathrm{ES}_\alpha(F))$, and more generally $(F^{-1}(q_1), \ldots, F^{-1}(q_{k-1}), \nu_\mu(F))$, admits only nonseparable strictly consistent scoring functions. This result gives an example demonstrating that Osband (1985), Proposition 2.3, cannot be correct as it states that any strictly consistent scoring function for a functional with a quantile as a component must be separable in the sense that it must be the sum of a strictly consistent scoring function for the quantile and a strictly consistent scoring function for the rest of the functional.

Using Theorem 5.2 and the revelation principle, we can now state one of the main results of this paper.

COROLLARY 5.4. *Let $\mathcal{F}$ be a class of distribution functions on $\mathbb{R}$ with finite first moments and unique quantiles. Let $\nu_\mu : \mathcal{F} \to \mathbb{R}$ be a spectral risk measure. If the support of μ is finite with L elements and contained in $(0, 1]$, then ν_μ is a component of a k-elicitable functional where*:

(i) $k = 1$, *if μ is concentrated at 1 meaning $\mu(\{1\}) = 1$;*
(ii) $k = 1 + L$, *if $\mu(\{1\}) < 1$.*

In the special case of $T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$, the maximal sensible action domain is $\mathsf{A}_0 := \{x \in \mathbb{R}^2 : x_1 \ge x_2\}$ as we always have $\mathrm{ES}_\alpha(F) \le \mathrm{VaR}_\alpha(F)$. For this action domain, the characterization of consistent scoring functions of Theorem 5.2 simplifies as follows.

COROLLARY 5.5. *Let* $\alpha \in (0, 1)$. *Let* $\mathcal{F}$ *be a class of distribution functions on* $\mathbb{R}$ *with finite first moments and unique* α-*quantiles. Let* $\mathsf{A}_0 = \{x \in \mathbb{R}^2 : x_1 \geq x_2\}$. *A scoring function* $S : \mathsf{A}_0 \times \mathbb{R} \to \mathbb{R}$ *of the form*

$$
\begin{aligned}
S(x_1, x_2, y) = {} & \big(\mathbb{1}\{y \leq x_1\} - \alpha\big) G_1(x_1) - \mathbb{1}\{y \leq x_1\} G_1(y) \\
& + G_2(x_2)\Big(x_2 - x_1 + \frac{1}{\alpha}\mathbb{1}\{y \leq x_1\}(x_1 - y)\Big) \\
& - \mathcal{G}_2(x_2) + a(y),
\end{aligned}
$$
(5.5)

where $G_1, G_2, \mathcal{G}_2, a : \mathbb{R} \to \mathbb{R}$, $\mathcal{G}_2' = G_2$, a *is* $\mathcal{F}$-*integrable and* $\mathbb{1}_{(-\infty, x_1]} G_1$ *is* $\mathcal{F}$-*integrable for all* $x_1 \in \mathbb{R}$, *is* $\mathcal{F}$-*consistent for* $T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ *if* G_1 *is increasing and* $\mathcal{G}_2$ *is increasing and convex. If* $\mathcal{G}_2$ *is strictly increasing and strictly convex, then* S *is strictly* $\mathcal{F}$-*consistent for* T.

Under the conditions of Theorem 5.2(iii) *all strictly* $\mathcal{F}$-*consistent scoring functions for* T *are of the form* (5.5) *almost everywhere.*

Acerbi and Székely (2014) also give an example of a scoring function for the pair $T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha) : \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^2$. They use a different sign convention for VaR_α and ES_α than we do in this paper. Using our sign convention, their proposed scoring function $S^W : \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ reads

$$
\begin{aligned}
S^W(x_1, x_2, y) = {} & \alpha\big(x_2^2/2 + W x_1^2/2 - x_1 x_2\big) \\
& + \mathbb{1}\{y \leq x_1\}\big(-x_2(y - x_1) + W(y^2 - x_1^2)/2\big),
\end{aligned}
$$
(5.6)

where $W \in \mathbb{R}$. The authors claim that S^W is a strictly $\mathcal{F}$-consistent scoring function for $T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ provided that

$$
\mathrm{ES}_\alpha(F) > W \mathrm{VaR}_\alpha(F)
$$
(5.7)

for all $F \in \mathcal{F}$. This means that they consider a strictly smaller action domain than A_0 in Corollary 5.5. They assume that the distributions in $\mathcal{F}$ have continuous densities, unique α quantiles, and that $F(x) \in (0, 1)$ implies $f(x) > 0$ for all $F \in F$ with density f. Furthermore, in order to ensure that $\bar{S}^W(\cdot, F)$ is finite, one needs to impose the assumption that $\int_{-\infty}^x y^2 \, dF(y)$ is finite for all $x \in \mathbb{R}$ and $F \in \mathcal{F}$. This is slightly less than requiring finite second moments. As a matter of fact, they only show that $\nabla \bar{S}^W(t_1, t_2, F) = 0$ for $F \in \mathcal{F}$ and $(t_1, t_2) = T(F)$ and that $\nabla^2 \bar{S}^W(t_1, t_2, F)$ is positive definite. This only shows that $\bar{S}^W(x, F)$ has a *local* minimum at $x = T(F)$ but does not provide a proof concerning a *global* minimum; see also the discussion after Corollary 3.3. However, we can use Theorem 5.2(ii) to verify their claims with $G_1(x_1) = -(W/2)x_1^2$, $\mathcal{G}_2(x_2) = (\alpha/2)x_2^2$ and $a = 0$. Hence, $\mathcal{G}_2$ is strictly convex, and the function $x_1 \mapsto x_1 G_2(x_2)/\alpha + G_1(x_1)$ is strictly increasing in x_1 if and only if $x_2 > W x_1$ as at (5.7).

The scoring function S^W has one property which is potentially relevant in applications. If x_1, x_2 and y are expressed in the same units of measurement,

then $S^W(x_1, x_2, y)$ is a quantity with these units squared. If one insists that we should only add quantities with the same units, then the necessary condition that $x_1 \mapsto x_1 G_2(x_2)/\alpha + G_1(x_1)$ is strictly increasing enforces a condition of the type (5.7). The action domain is restricted for S^W and the choice of W may not be obvious in practice. Similarly, for the maximal action domain A_0, an open question of practical interest is the choice of the functions G_1 and $\mathcal{G}_2$ in (5.5). We would like to remark that S remains strictly consistent upon choosing $G_1 = 0$ and $\mathcal{G}_2$ strictly increasing and strictly convex.

6. Discussion. We have investigated necessary and sufficient conditions for the elicitability of k-dimensional functionals of d-dimensio nal distributions. In order to derive necessary conditions, we have adapted Osband's principle for the case where the class $\mathcal{F}$ of distributions does not necessarily contain distributions with finite support. This comes at the cost of certain smoothness assumptions on the expected scores $\bar{S}(\cdot, F)$. For particular situations, for example, when characterizing the class of strictly $\mathcal{F}$-consistent scoring functions for ratios of expectations, it is possible to weaken the smoothness assumptions; see Frongillo and Kash (2015). While moving away from distributions with finite support is not a great gain in the case of linear functionals or ratios of expectations, it comes in handy when considering spectral risk measures. Value at Risk, VaR_α, being defined as the smallest α-quantile, is generally not elicitable for distributions where the α-quantile is not unique. Therefore, we believe that it is also not possible to show joint elicitability of $(\text{VaR}_\alpha, \text{ES}_\alpha)$ for classes $\mathcal{F}$ of distributions with nonunique α-quantiles. However, we can give consistent scoring functions which become strictly consistent as soon as the elements of $\mathcal{F}$ have unique quantiles. Fortunately, the classes $\mathcal{F}$ of distributions that are relevant in risk management usually consist of absolutely continuous distributions having unique quantiles.

Emmer, Kratz and Tasche (2015) have remarked that ES_α is *conditionally elicitable*. Slightly generalizing their definition, a functional $T_k: \mathcal{F} \to \mathsf{A}_k \subseteq \mathbb{R}$ is called *conditionally elicitable of order k*, $k \geq 1$, if there are $k - 1$ elicitable functionals $T_m: \mathcal{F} \to \mathsf{A}_m \subseteq \mathbb{R}$, $m \in \{1, \ldots, k-1\}$, such that T_k is elicitable restricted to the class $\mathcal{F}_{x_1, \ldots, x_{k-1}} := \{F \in \mathcal{F}: T_1(F) = x_1, \ldots, T_{k-1}(F) = x_{k-1}\}$ for any $(x_1, \ldots, x_{k-1}) \in \mathsf{A}_1 \times \cdots \times \mathsf{A}_{k-1}$. *Mutatis mutandis*, one can define a notion of *conditional identifiability* by replacing the term "elicitable" with "identifiable" in the above definition. It is not difficult to check that any conditionally identifiable functional T_k of order k is a component of an identifiable functional $T = (T_1, \ldots, T_k)$. Spectral risk measures ν_μ with spectral measure μ with finite support in $(0, 1)$ provide an example of a conditionally elicitable functional of order $L + 1$, where L is the cardinality of the support of μ; see Theorem 5.2. However, we would like to stress that it is generally an open question whether any conditionally elicitable and identifiable functional T_k of order $k \geq 2$ is always a component of a k-elicitable functional.

Slightly modifying Lambert, Pennock and Shoham (2008), Definition 11, one could define the *elicitability order* of a real-valued functional T as the smallest number k such that the functional is a component of a k-elicitable functional. It is clear that the elicitability order of the variance is two, and we have shown that the same is true for ES_α for reasonably large classes $\mathcal{F}$. For spectral risk measures ν_μ, the elicitability order is at most $L + 1$, where L is the cardinality of the support; see Corollary 5.4.

In the one-dimensional case, Steinwart et al. (2014) have shown that convex level sets in the sense of Osband (1985), Proposition 2.5 [see also Gneiting (2011), Theorem 6] is a sufficient condition for the elicitability of a functional T under continuity assumptions on T. Without such continuity assumptions, the converse of Osband (1985), Proposition 2.5, is generally false; see Heinrich (2014) for the example of the mode functional. It is an open (and potentially difficult) question under which conditions a converse of Osband (1985), Proposition 2.5, is true for higher order elicitability.

7. Proofs.

PROOF OF THEOREM 3.2. Let $x \in \text{int}(\mathsf{A})$. The identifiability property of V plus the first-order condition stemming from the strict $\mathcal{F}$-consistency of S yields the relation $\bar{V}(x, F) = 0 \implies \nabla \bar{S}(x, F) = 0$ for all $F \in \mathcal{F}$. Let $l \in \{1, \ldots, k\}$. To show (3.2), consider the composed functional

$$\bar{B}(x, \cdot) : \mathcal{F} \to \mathbb{R}^{k+1}, \qquad F \mapsto \left(\partial_l \bar{S}(x, F), \bar{V}(x, F) \right).$$

By construction, we know that

$$(7.1) \qquad\qquad \bar{V}(x, F) = 0 \iff \bar{B}(x, F) = 0$$

for all $F \in \mathcal{F}$. Assumption (V1) implies that there are $F_1, \ldots, F_{k+1} \in \mathcal{F}$ such that the matrix $\mathbb{V} = \text{mat}(\bar{V}(x, F_1), \ldots, \bar{V}(x, F_{k+1})) \in \mathbb{R}^{k \times (k+1)}$ has maximal rank, meaning $\text{rank}(\mathbb{V}) = k$. If $\text{rank}(\mathbb{V}) < k$, then the space $\text{span}\{\bar{V}(x, F_1), \ldots, \bar{V}(x, F_{k+1})\}$ would be a linear subspace such that the interior of $\text{conv}(\{\bar{V}(x, F_1), \ldots, \bar{V}(x, F_{k+1})\})$ would be empty. Let $G \in \mathcal{F}$. Then still $0 \in \text{int}(\text{conv}(\{\bar{V}(x, G), \bar{V}(x, F_1), \ldots, \bar{V}(x, F_{k+1})\}))$, so $\text{rank}(\mathbb{V}_G) = k$ where $\mathbb{V}_G = \text{mat}(\bar{V}(x, G), \bar{V}(x, F_1), \ldots, \bar{V}(x, F_{k+1})) \in \mathbb{R}^{k \times (k+2)}$. Define the matrix

$$\mathbb{B}_G = \begin{pmatrix} \partial_l \bar{S}(x, G) & \partial_l \bar{S}(x, F_1) & \cdots & \partial_l \bar{S}(x, F_{k+1}) \\ & \mathbb{V}_G & & \end{pmatrix} \in \mathbb{R}^{(k+1) \times (k+2)}.$$

We use (7.1) to show that $\ker(\mathbb{B}_G) = \ker(\mathbb{V}_G)$. First, observe that the relation $\ker(\mathbb{B}_G) \subseteq \ker(\mathbb{V}_G)$ is clear by construction. To show the other inclusion, let $\theta \in \ker(\mathbb{V}_G)$ be an element of the simplex. Then (7.1) and the convexity of $\mathcal{F}$ yields that $\theta \in \ker(\mathbb{B}_G)$. By linearity, the inclusion holds also for all $\theta \in \ker(\mathbb{V}_G)$ with nonnegative components. Finally, let $\theta \in \ker(\mathbb{V}_G)$ be arbitrary. Assumption (V1) implies that there is $\theta^* \in \ker(\mathbb{V}_G)$ with strictly positive components.

Hence, there is an $\varepsilon > 0$ such that $\theta^* + \varepsilon\theta$ has nonnegative components. Since $\mathbb{V}_G(\theta^* + \varepsilon\theta) = \mathbb{V}_G\theta^* + \varepsilon\mathbb{V}_G\theta = 0$, we know that $\theta^* + \varepsilon\theta \in \ker(\mathbb{B}_G)$. Again using linearity and the fact that $\theta^* \in \ker(\mathbb{B}_G)$, we obtain that $\theta \in \ker(\mathbb{B}_G)$.

With the rank-nullity theorem, this gives $\operatorname{rank}(\mathbb{B}_G) = \operatorname{rank}(\mathbb{V}_G) = k$. Hence, there is a unique vector $(h_{l1}(x), \ldots, h_{lk}(x)) \in \mathbb{R}^k$ such that one has $\partial_l \bar{S}(x, G) = \sum_{m=1}^{k} h_{lm}(x)\bar{V}_m(x, G)$. Since $G \in \mathcal{F}$ was arbitrary, the assertion at (3.2) follows.

The second part of the claim can be seen as follows. For $x \in \operatorname{int}(\mathsf{A})$, pick $F_1, \ldots, F_k \in \mathcal{F}$ such that $\bar{V}(x, F_1), \ldots, \bar{V}(x, F_k)$ are linearly independent and let $\mathbb{V}(z)$ be the matrix with columns $\bar{V}(z, F_i)$, $i \in \{1, \ldots, k\}$ for $z \in \operatorname{int}(\mathsf{A})$. Due to assumption (V2) or (V3), $\mathbb{V}(z)$ has full rank in some neighborhood U of x. Let $r \in \{1, \ldots, k\}$ and let e_r be the rth standard unit vector of $\mathbb{R}^k$. We define $\lambda(z) := \mathbb{V}(z)^{-1}e_r$ for $z \in U$. Taking the inverse of a matrix is a continuously differentiable operation, so it is in particular locally Lipschitz continuous. Therefore, the vector λ inherits the regularity properties of $\bar{V}(z, F_i)$, that is, under (V2) λ is continuous, and under (V3) λ is locally Lipschitz continuous. Therefore, these properties carry over to h because for $l \in \{1, \ldots, k\}$, $z \in U$

$$h_{lr}(z) = \sum_{i=1}^{k} \lambda_i(z) \sum_{m=1}^{k} h_{lm}(z)\bar{V}_m(z, F_i) = \sum_{i=1}^{k} \lambda_i(z)\partial_l \bar{S}_m(z, F_i)$$

using the assumptions on S. $\square$

PROOF OF PROPOSITIONS 4.2 AND 4.4.   We show parts (i) of the two propositions simultaneously. We have that $\partial_l \bar{V}_r(x, F) = 0$ for all $l, r \in \{1, \ldots, k\}$, $l \neq r$, and $x \in \operatorname{int}(\mathsf{A})$, $F \in \mathcal{F}$. Equation (3.3) evaluated at $x = t = T(F)$ yields

$$(7.2) \qquad\qquad h_{rl}(t)\partial_l \bar{V}_l(t, F) = h_{lr}(t)\partial_r \bar{V}_r(t, F).$$

If (V4) holds, then (7.2) implies that $h_{rl}(t) = 0$ for $r \neq l$, hence we obtain (4.2) with the surjectivity of T. On the other hand, if $V_r(x, y) = q(y)x_m - p_m(y)$, (7.2) implies that $h_{rl}(t) = h_{lr}(t)$, whence the second part of (4.4) is shown, again using the surjectivity of T. In both cases, (3.3) is equivalent to

$$(7.3) \qquad\qquad \sum_{m=1}^{k} \big(\partial_l h_{rm}(x) - \partial_r h_{lm}(x)\big)\bar{V}_m(x, F) = 0.$$

Using assumption (V1), there are $F_1, \ldots, F_k \in \mathcal{F}$ such that the vectors $\bar{V}(x, F_1), \ldots, \bar{V}(x, F_k)$ are linearly independent. This yields that $\partial_l h_{rm}(x) = \partial_r h_{lm}(x)$ for almost all $x \in \operatorname{int}(\mathsf{A})$. For Proposition 4.2, we can conclude that $\partial_l h_{rr}(x) = \partial_r h_{lr}(x) = 0$ for $r \neq l$ for almost all $x \in \operatorname{int}(\mathsf{A})$. Consequently, invoking that A is connected, the functions h_{mm} only depend on x_m and we can write $h_{mm}(x) = g_m(x_m)$ for some function $g_m : \mathsf{A}'_m \to \mathbb{R}$. By Lemma 2.4(i), for $v \in \mathbb{S}^{k-1}$, $t =$

$T(F) \in \text{int}(\mathsf{A})$, the function $s \mapsto \bar{S}(t + sv, F)$ has a global unique minimum at $s = 0$, hence

$$v^\top \nabla \bar{S}(t + sv, F) = \sum_{m=1}^{k} g_m(t_m + sv_m) \bar{V}_m(t_m + sv_m, F) v_m$$

vanishes for $s = 0$, is negative for $s < 0$ and positive for $s > 0$, where s is in some neighborhood of zero. Choosing v as the lth standard basis vector of $\mathbb{R}^k$ we obtain that $g_l > 0$ exploiting the orientation of V_l and the surjectivity of T.

For Proposition 4.4(i) to show the assertion about the definiteness of h, observe that for $v \in \mathbb{S}^{k-1}$, $t = T(F) \in \text{int}(\mathsf{A})$ we have $\bar{V}(t + sv, F) = \bar{q}(F) sv$ where $\bar{q}(F) > 0$. Hence, $v^\top \nabla \bar{S}(t + sv, F) = \bar{q}(F) sv^\top h(t + sv)v$, which implies the claim using again the surjectivity of T.

For part (ii) of Proposition 4.2, the sufficiency is immediate; see the proof of Lemma 2.6. For necessity, we apply Proposition 3.4 and part (i) such that

$$S(x, y) = \sum_{m=1}^{k} \int_{z_m}^{x_m} g_m(v) V_m(v, y)\, \mathrm{d}v + a(y),$$

for almost all $(x, y) \in \mathsf{A} \times \mathsf{O}$, where $z \in \text{int}(\mathsf{A})$ is a star point of $\text{int}(\mathsf{A})$ and a is an $\mathcal{F}$-integrable function. Let $t = T(F)$ and $x_m \neq t_m$. The strict consistency of S implies that $\bar{S}(t, F) < \bar{S}(t_1, \ldots, t_{m-1}, x_m, t_{m+1}, \ldots, t_k)$. This means $\bar{S}_m(t_m, F) < \bar{S}_m(x_m, F)$ with $S_m(x_m, y) := \int_{z_m}^{x_m} g_m(v) V_m(v, y)\, \mathrm{d}v + (1/k)a(y)$.

For part (ii) of Proposition 4.4, observe that due to part (i) h is the Hessian of ϕ, and thus, ϕ is strictly convex. For the sufficiency of the form (4.5), let $x \neq t = T(F)$ for some $F \in \mathcal{F}$. Then

$$\bar{S}(x, F) - \bar{S}(t, F) = \bar{q}(F)\big(\phi(t) - \phi(x) + \langle \nabla\phi(x), x - t \rangle\big) > 0$$

due to the strict convexity of ϕ and $\bar{q}(F) > 0$. For the necessity of the form (4.5), apply Proposition 3.4 and use partial integration. □

PROOF OF THEOREM 5.2. (i) The second part of Theorem 5.2(ii) implies the k-elicitability of T.

(ii) Let $S: \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ be of the form (5.2), $\mathcal{G}_k$ be convex and the functions at (5.3) be increasing. Let $F \in \mathcal{F}$, $x = (x_1, \ldots, x_k) \in \mathsf{A}$ and set $t = (t_1, \ldots, t_k) = T(F)$, $w = \min(x_k, t_k)$. Then we obtain

$$S(x, y) = -\mathcal{G}_k(x_k) + G_k(w)(x_k - y) + a(y)$$

$$+ \sum_{r=1}^{k-1} \big(\mathbb{1}\{y \leq x_r\} - q_r\big)\Big(G_r(x_r) + \frac{p_r}{q_r} G_k(w)(x_r - y)\Big)$$

$$- \mathbb{1}\{y \leq x_r\} G_r(y) + \big(G_k(x_k) - G_k(w)\big)$$

$$\times \left(x_k + \sum_{m=1}^{k-1} \frac{p_m}{q_m}\big(\mathbb{1}\{y \leq x_m\}(x_m - y) - q_m x_m\big)\right).$$

This implies that $\bar{S}(x, F) - \bar{S}(t, F) = R_1 + R_2$ with

$$R_1 = \sum_{r=1}^{k-1} (F(x_r) - q_r)\left(G_r(x_r) + \frac{p_r}{q_r}G_k(w)x_r\right)$$

$$- \int_{t_r}^{x_r}\left(G_r(y) + \frac{p_r}{q_r}G_k(w)y\right)dF(y),$$

$$R_2 = (G_k(x_k) - G_k(w))\left(x_k + \sum_{m=1}^{k-1}\frac{p_m}{q_m}\left(\int_{-\infty}^{x_m}(x_m - y)\,dF(y) - q_m x_m\right)\right)$$

$$- \mathcal{G}_k(x_k) + \mathcal{G}_k(t_k) + G_k(w)(x_k - t_k).$$

We denote the rth summand of R_1 by ξ_r and suppose that $t_r < x_r$. Due to the assumptions, the term $G_r(y) + (p_r/q_r)G_k(w)y$ is increasing in $y \in [t_r, x_r]$ which implies that $\xi_r \geq (F(x_r) - q_r)(G_r(x_r) + (p_r/q_r)G_k(w)x_r) - (F(x_r) - F(t_r))(G_r(x_r) + (p_r/q_r)G_k(w)x_r) = 0$. Analogously, one can show that $\xi_r \geq 0$ if $x_r < t_r$. If F has a unique q_r-quantile and the term $G_r(y) + (p_r/q_r)G_k(w)y$ is strictly increasing in y, then we even get $\xi_r > 0$ if $x_r \neq t_r$.

Now consider the term R_2. Splitting the integrals from ∞ to x_m into integrals from $-\infty$ to t_m and from t_m to x_m and partially integrating the latter, we obtain

$$R_2 = (G_k(x_k) - G_k(w))$$

$$\times \left(x_k + \sum_{m=1}^{k-1}p_m\left(t_m - x_m - \frac{1}{q_m}\int_{-\infty}^{t_m}y\,dF(y) + \frac{1}{q_m}\int_{t_m}^{x_m}F(y)\,dy\right)\right)$$

$$- \mathcal{G}_k(x_k) + \mathcal{G}_k(t_k) + G_k(w)(x_k - t_k)$$

$$= (G_k(x_k) - G_k(w))\left(x_k - t_k + \sum_{m=1}^{k-1}p_m\left(t_m - x_m + \frac{1}{q_m}\int_{t_m}^{x_m}F(y)\,dy\right)\right)$$

$$- \mathcal{G}_k(x_k) + \mathcal{G}_k(t_k) + G_k(w)(x_k - t_k)$$

$$\geq (G_k(x_k) - G_k(w))(x_k - t_k) - \mathcal{G}_k(x_k) + \mathcal{G}_k(t_k) + G_k(w)(x_k - t_k)$$

$$= \mathcal{G}_k(t_k) - \mathcal{G}_k(x_k) - G_k(x_k)(t_k - x_k) \geq 0.$$

The first inequality is due to the fact that (i) G_k is increasing and (ii) for $x_m \neq t_m$ we have $(1/q_m)\int_{t_m}^{x_m} F(y)\,dy \geq x_m - t_m$ with strict inequality if F has a unique q_m-quantile. The last inequality is due to the fact that $\mathcal{G}_k$ is convex. The inequality is strict if $x_k \neq t_k$ and if $\mathcal{G}_k$ is strictly convex.

(iii) If f denotes the density of F, it holds that

$$(7.4) \qquad \mathrm{ES}_\alpha(F) = \frac{1}{\alpha}\int_{-\infty}^{F^{-1}(\alpha)} yf(y)\,dy, \qquad \alpha \in (0, 1].$$

We first show the assertions concerning V given at (5.4). Let $F \in \mathcal{F}$ with density $f = F'$ and let $t = T(F)$. Then we have for $m \in \{1, \ldots, k-1\}$, $x \in \mathsf{A}$, that

$\bar{V}_m(x, F) = F(x_m) - q_m$ which is zero if and only if $x_m = t_m$. On the other hand, using the identity at (7.4)

$$\bar{V}_k(t_1, \ldots, t_{k-1}, x_k, F) = x_k - \sum_{m=1}^{k-1} \frac{p_m}{q_m} \int_{-\infty}^{t_m} y f(y) \, dy = x_k - t_k.$$

Hence, it follows that V is a strict $\mathcal{F}$-identification function for T. Moreover, V satisfies assumption (V3), and we have for $m \in \{1, \ldots, k-1\}$, $l \in \{1, \ldots, k\}$ and $x \in \text{int}(A)$ that $\partial_l \bar{V}_m(x, F) = 0$ if $l \neq m$ and $\partial_m \bar{V}_m(x, F) = f(x_m)$, $\partial_m \bar{V}_k(x, F) = -(p_m/q_m)x_m f(x_m)$ and $\partial_k \bar{V}_k(x, F) = 1$.

From now on, we assume that $t = T(F) \in \text{int}(A)$. Let S be a strictly $\mathcal{F}$-consistent scoring function for T satisfying (S2). Then we can apply Theorem 3.2 and Corollary 3.3 to get that there are locally Lipschitz continuous functions $h_{lm} : \text{int}(A) \to \mathbb{R}$ such that (3.2) and (3.3) hold. If we evaluate (3.3) for $l = k$, $m \in \{1, \ldots, k-1\}$ at the point $x = t$, we get

$$h_{km}(t) \partial_m \bar{V}_m(t, F) + h_{kk}(t) \partial_m \bar{V}_k(t, F) = h_{mk}(t) \partial_k \bar{V}_k(t, F),$$

which takes the form $h_{km}(t) f(t_m) - h_{kk}(t)(p_m/q_m)t_m f(t_m) = h_{mk}(t)$. Invoking assumption (V4) for $(V_1, \ldots, V_{k-1})$, we get that necessarily $h_{mk}(t) = 0$ and $h_{km}(t) = (p_m/q_m)t_m h_{kk}(t)$. So with the surjectivity of T, we get for $x \in \text{int}(A)$ that

$$(7.5) \quad h_{mk}(x) = 0, \qquad h_{km}(x) = \frac{p_m}{q_m} x_m h_{kk}(x) \qquad \text{for all } m \in \{1, \ldots, k-1\}.$$

Now, we can evaluate (3.3) for $m, l \in \{1, \ldots, k-1\}$, $m \neq l$, at $x = t$ and use the first part of (7.5) to get that $h_{ml}(t) f(t_l) = h_{lm}(t) f(t_m)$. Using again the same argument, we get for $x \in \text{int}(A)$ that

$$(7.6) \qquad h_{ml}(x) = 0 \qquad \text{for all } m, l \in \{1, \ldots, k-1\}, l \neq m.$$

At this stage, we can evaluate (3.3) for $l \in \{1, \ldots, k-1\}$, $m \in \{1, \ldots, k\}$, $m \neq l$, for some $x \in \text{int}(A)$. Using (7.5) and (7.6), we obtain

$$\sum_{i=1}^{k} \left(\partial_l h_{mi}(x) - \partial_m h_{li}(x) \right) \bar{V}_i(x_i, F) = 0.$$

Invoking assumption (V1) and using (7.5) and (7.6), we can conclude that for almost all $x \in A$,

$$(7.7) \qquad \partial_l h_{mm}(x) = 0 \qquad \text{for all } l \in \{1, \ldots, k-1\}, m \in \{1, \ldots, k\}, l \neq m$$

and

$$(7.8) \qquad \partial_k h_{ll}(x) = \frac{p_l}{q_l} h_{kk}(x) \qquad \text{for all } l \in \{1, \ldots, k-1\}.$$

Equation (7.7) for $m = k$ shows that there is a locally Lipschitz continuous function $g_k : A'_k \to \mathbb{R}$ such that for all $(x_1, \ldots, x_k) \in \text{int}(A)$, we have $h_{kk}(x_1, \ldots, x_k) =$

$g_k(x_k)$. Equation (7.8) together with (7.7) gives that for $l \in \{1, \ldots, k-1\}$, and $(x_1, \ldots, x_k) \in \text{int}(\mathsf{A})$, we obtain $h_{ll}(x_1, \ldots, x_k) = (p_l/q_l)G_k(x_k) + g_l(x_l)$, where $g_l \colon \mathsf{A}'_l \to \mathbb{R}$ is locally Lipschitz continuous and $G_k \colon \mathsf{A}'_k \to \mathbb{R}$ is such that $G'_k = g_k$.

Knowing the form of the matrix-valued function h, we can apply Proposition 3.4. Let $z \in \text{int}(\mathsf{A})$ be some star point. Then there is some $\mathcal{F}$-integrable function $b \colon \mathbb{R} \to \mathbb{R}$ such that

$$
\begin{aligned}
S(x, y) = \sum_{r=1}^{k-1} \int_{z_r}^{x_r} \left(\frac{p_r}{q_r} G_k(z_k) + g_r(v) \right) \big(\mathbb{1}\{y \leq v\} - q_r \big) \, dv
\end{aligned}
$$

$$
\begin{aligned}
(7.9) \qquad + \big(G_k(x_k) - G_k(z_k) \big) \sum_{m=1}^{k-1} \frac{p_m}{q_m} \big(x_m (\mathbb{1}\{y \leq x_m\} - q_m) - y\mathbb{1}\{y \leq x_m\} \big)
\end{aligned}
$$

$$
\begin{aligned}
+ G_k(x_k)x_k - \mathcal{G}_k(x_k) + b(y),
\end{aligned}
$$

for almost all (x, y) where $\mathcal{G}_k \colon \mathsf{A}'_k \to \mathbb{R}$ is such that $\mathcal{G}'_k = G_k$. One can check by a straightforward computation that the representation of S at (7.9) is equivalent to the one at (5.2) upon choosing a suitable $\mathcal{F}$-integrable function $a \colon \mathbb{R} \to \mathbb{R}$.

It remains to show that $\mathcal{G}_k$ is strictly convex and that the functions given at (5.3) are strictly increasing. To this end, we use Lemma 2.4. Let $D = \{s \in \mathbb{R} \colon t + sv \in \text{int}(\mathsf{A})\}$, and let $v = (v_1, \ldots, v_k) \in \mathbb{S}^{k-1}$ and without loss of generality assume $v_k \geq 0$. We define $\psi \colon D \to \mathbb{R}$ by $\psi(s) := \bar{S}(t + sv, F)$, that is,

$$
\begin{aligned}
\psi(s) = \sum_{r=1}^{k-1} \int_{z_r}^{\bar{s}_r} \left(\frac{p_r}{q_r} G_k(z_k) + g_r(v) \right) \big(F(v) - q_r \big) \, dv
\end{aligned}
$$

$$
\begin{aligned}
+ \big(G_k(\bar{s}_k) - G_k(z_k) \big) \sum_{m=1}^{k-1} \frac{p_m}{q_m} \left(\bar{s}_m \big(F(\bar{s}_m) - q_m \big) - \int_{-\infty}^{\bar{s}_m} yf(y) \, dy \right)
\end{aligned}
$$

$$
\begin{aligned}
+ \bar{s}_k G_k(\bar{s}_k) - \mathcal{G}_k(\bar{s}_k) + \bar{b}(F),
\end{aligned}
$$

where we use the notation $\bar{s} = t + sv$. The function ψ has a minimum at $s = 0$. Hence, there is $\varepsilon > 0$ such that $\psi'(s) < 0$ for $s \in (-\varepsilon, 0)$ and $\psi'(s) > 0$ for $s \in (0, \varepsilon)$. If $v_k = 0$, then

$$
\psi'(s) = \sum_{r=1}^{k-1} (F(\bar{s}_r) - q_r)v_r \left(g_r(\bar{s}_r) + \frac{p_r}{q_r} G_k(\bar{s}_k) \right).
$$

Choosing v as the rth standard basis vector of $\mathbb{R}^k$ for $r \in \{1, \ldots, k-1\}$, we obtain that $g_r(\bar{s}_r) + (p_r/q_r)G_k(\bar{s}_k) > 0$. Exploiting the surjectivity of T we can deduce that the functions at (5.3) are strictly increasing. On the other hand, if v is the kth standard basis vector, we obtain that $\psi'(s) = g_k(\bar{s}_k)s$. Again using the surjectivity of T, we get that $g_k > 0$ which shows the strict convexity of $\mathcal{G}_k$. $\quad\square$

PROOF OF COROLLARY 5.5. The sufficiency follows directly from Theorem 5.2. We will show that G_2 is necessarily bounded below. Suppose the contrary.

For the action domain A_0, we have $\mathsf{A}'_{1,x_2} = [x_2, \infty)$, therefore, for $x_2 \le x_1 < x'_1$ (5.3) yields $-\infty < G_1(x_1) - G_1(x'_1) \le (1/\alpha)G_2(x_2)(x'_1 - x_1)$. Letting $x_2 \to -\infty$, one obtains a contradiction. Let $C_2 = \lim_{x_2 \to -\infty} G_2(x_2) > -\infty$. Then, by (5.3), we obtain that $G_1(x_1) + (C_2/\alpha)x_1$ is increasing in $x_1 \in \mathbb{R}$. We can write S at (5.5) as

$$S(x_1, x_2, y) = (G_2(x_2) - C_2)\left(\frac{1}{\alpha}\mathbb{1}\{y \le x_1\}(x_1 - y) - (x_1 - x_2)\right)$$

$$+ (\mathbb{1}\{y \le x_1\} - \alpha)\left(G_1(x_1) + \frac{C_2}{\alpha}x_1\right)$$

$$- \mathbb{1}\{y \le x_1\}\left(G_1(y) + \frac{C_2}{\alpha}y\right)$$

$$- (\mathcal{G}_2(x_2) - C_2 x_2) + a(y).$$

The last expression is again of the form at (5.5) with increasing functions $\tilde{G}_1(x_1) = G_1(x_1) + (C_2/\alpha)x_1$ and $\tilde{G}_2(x_2) = G_2(x_2) - C_2 \ge 0$. $\quad\square$

SUPPLEMENTARY MATERIAL

Supplement to "Higher order elicitability and Osband's principle" (DOI: 10.1214/16-AOS1439SUPP; .pdf). The proofs of Proposition 3.4 and Corollary 5.4 are deferred to this supplement.

REFERENCES

ABERNETHY, J. D. and FRONGILLO, R. M. (2012). A characterization of scoring rules for linear properties. In *Proceedings of the 25th Annual Conference on Learning Theory. Journal of Machine Learning Research*: *Workshop and Conference Proceedings*, *Vol.* 23 (S. Mannor, N. Srebro and R. C. Williamson, eds.) 27.1–27.13.

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

ACERBI, C. and SZÉKELY, B. (2014). Backtesting expected shortfall. *Risk Mag.* **27** 76–81.

ARTZNER, P., DELBAEN, F., EBER, J.-M. and HEATH, D. (1999). Coherent measures of risk. *Math. Finance* **9** 203–228. MR1850791

BANERJEE, A., GUO, X. and WANG, H. (2005). On the optimality of conditional expectation as a Bregman predictor. *IEEE Trans. Inform. Theory* **51** 2664–2669. MR2246384

BELLINI, F. and BIGNOZZI, V. (2015). On elicitable risk measures. *Quant. Finance* **15** 725–733. MR3334566

CONT, R., DEGUEST, R. and SCANDOLO, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quant. Finance* **10** 593–606. MR2676786

DANÍELSSON, J., EMBRECHTS, P., GOODHART, C., KEATING, C., MUENNICH, F., RENAULT, O. and SHIN, H. S. (2001). An academic response to Basel II. Special Paper No. 130.

DAVIS, M. H. A. (2016). Verification of internal risk measure estimates. *Stat. Risk Model*. To appear.

DAWID, A. P. and SEBASTIANI, P. (1999). Coherent dispersion criteria for optimal experimental design. *Ann. Statist.* **27** 65–81. MR1701101

DELBAEN, F. (2012). *Monetary Utility Functions*. Osaka Univ. Press, Osaka.

DELBAEN, F., BELLINI, F., BIGNOZZI, V. and ZIEGEL, J. F. (2016). Risk measures with the CxLS property. *Finance Stoch.* **20** 433–453. MR3479327

EMBRECHTS, P. and HOFERT, M. (2014). Statistics and quantitative risk management for banking and insurance. *Ann. Rev. Stat. Appl.* **1**.

EMMER, S., KRATZ, M. and TASCHE, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *J. Risk* **18** 31–60.

ENGELBERG, J., MANSKI, C. F. and WILLIAMS, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *J. Bus. Econom. Statist.* **27** 30–41. MR2484982

FISSLER, T. and ZIEGEL, J. F. (2016). Supplement to "Higher order elicitability and Osband's principle." DOI:10.1214/16-AOS1439SUPP.

FRONGILLO, R. and KASH, I. (2015). Vector-valued property elicitation. *J. Mach. Learn. Res. Workshop Conf. Proc.* **40** 1–18.

GNEITING, T. (2011). Making and evaluating point forecasts. *J. Amer. Statist. Assoc.* **106** 746–762. MR2847988

GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548

GRAUERT, H. and FISCHER, W. (1978). *Differential- und Integralrechnung. II*: *Differentialrechnung in Mehreren Veränderlichen Differentialgleichungen*, corrected ed. *Heidelberger Taschenbücher* [*Heidelberg Paperbacks*] **36**. Springer, Berlin. MR0514400

HEINRICH, C. (2014). The mode functional is not elicitable. *Biometrika* **101** 245–251. MR3180670

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101. MR0161415

JOUINI, E., SCHACHERMAYER, W. and TOUZI, N. (2006). Law invariant risk measures have the Fatou property. In *Advances in Mathematical Economics*. *Vol.* 9. *Adv. Math. Econ.* **9** 49–71. Springer, Tokyo. MR2277714

KOENKER, R. (2005). *Quantile Regression*. *Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. MR2268657

KOU, S. and PENG, X. (2014). On the measurement of economic tail risk. Preprint. Available at http://arxiv.org/pdf/1401.4787v2.pdf.

KOU, S., PENG, X. and HEYDE, C. C. (2013). External risk measures and Basel accords. *Math. Oper. Res.* **38** 393–417. MR3092538

KRÄTSCHMER, V., SCHIED, A. and ZÄHLE, H. (2012). Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *J. Multivariate Anal.* **103** 35–47. MR2823706

KRÄTSCHMER, V., SCHIED, A. and ZÄHLE, H. (2015). Quasi-Hadamard differentiability of general risk functionals and its applications. *Stat. Risk Model.* **32** 25–47. MR3331780

KRÄTSCHMER, V., SCHIED, A. and ZÄHLE, H. (2014). Comparative and qualitative robustness for law-invariant risk measures. *Finance Stoch.* **18** 271–295. MR3177407

KUSUOKA, S. (2001). On law invariant coherent risk measures. In *Advances in Mathematical Economics*, *Vol.* 3. *Adv. Math. Econ.* **3** 83–95. Springer, Tokyo. MR1886557

LAMBERT, N. (2013). Elicitation and evaluation of statistical functionals. Preprint. Available at http://web.stanford.edu/~nlambert/papers/elicitation.pdf.

LAMBERT, N., PENNOCK, D. M. and SHOHAM, Y. (2008). Eliciting properties of probability distributions. In *Proceedings of the* 9*th ACM Conference on Electronic Commerce* 129–138. ACM, Chicago, IL.

MURPHY, A. H. and DAAN, H. (1985). Forecast evaluation. In *Probability, Statistics and Decision Making in the Atmospheric Sciences* (A. H. Murphy and R. W. Katz, eds.) 379–437. Westview Press, Boulder, CO.

NEWEY, W. K. and POWELL, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55** 819–847. MR0906565

OSBAND, K. H. (1985). Providing incentives for better cost forecasting. Ph.D. thesis, Univ. California, Berkeley.

OSBAND, K. and REICHELSTEIN, S. (1985). Information-eliciting compensation schemes. *J. Public Econ.* **27** 107–115.

SAVAGE, L. J. (1971). Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* **66** 783–801. MR0331571

STEINWART, I., PASIN, C., WILLIAMSON, R. and ZHANG, S. (2014). Elicitation and identification of properties. In *Proceedings of the 27th Conference on Learning Theory. Journal of Machine Learning Research*: *Workshop and Conference Proceedings*, *Vol*. 35 (M. F. Balcan and C. Szepesvari, eds.) 482–526.

WANG, R. and ZIEGEL, J. F. (2015). Elicitable distortion risk measures: A concise proof. *Statist. Probab. Lett*. **100** 172–175. MR3324090

WEBER, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Math. Finance* **16** 419–441. MR2212272

ZIEGEL, J. F. (2014). Coherence and elicitability. *Math. Finance*. To appear.

INSTITUTE OF MATHEMATICAL STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF BERN
SIDLERSTRASSE 5, CH-3012 BERN
SWITZERLAND
E-MAIL: tobias.fissler@stat.unibe.ch
johanna.ziegel@stat.unibe.ch