On Elicitation Complexity and Conditional Elicitation

Rafael Frongillo
Harvard University
raf@cs.berkeley.edu

Ian A. Kash
Microsoft Research
iankash@microsoft.com

June 25, 2015

Abstract

Elicitation is the study of statistics or *properties* which are computable via empirical risk minimization. While several recent papers have approached the general question of which properties are elicitable, we suggest that this is the wrong question—all properties are elicitable by first eliciting the entire distribution or data set, and thus the important question is *how* elicitable. Specifically, what is the minimum number of regression parameters needed to compute the property?

Building on previous work, we introduce a new notion of elicitation complexity and lay the foundations for a calculus of elicitation. We establish several general results and techniques for proving upper and lower bounds on elicitation complexity. These results provide tight bounds for eliciting the Bayes risk of any loss, a large class of properties which includes spectral risk measures and several new properties of interest. Finally, we extend our calculus to conditionally elicitable properties, which are elicitable conditioned on knowing the value of another property, giving a necessary condition for the elicitability of both properties together.

1 Introduction

Empirical risk minimization (ERM) is a domininant framework for supervised machine learning, and a key component of many learning algorithms. A statistic or *property* is simply a functional assigning a vector of values to each distribution. We say that such a property is *elicitable*, if for some loss function it can be represented as the unique minimizer of the expected loss under the distribution. Thus, the study of which properties are elicitable can be viewed as the study of which statistics are computable via ERM [1, 2, 3].

The study of property elicitation began in statistics [4, 5, 6, 7], and is gaining momentum in machine learning [8, 1, 2, 3], economics [9, 10], and most recently, finance [11, 12, 13, 14, 15]. A sequence of papers starting with Savage [4] has looked at the full characterization of losses which elicit the mean of a distribution, or more generally the expectation of a vector-valued random variable [16, 3]. The case of real-valued properties is also now well in hand [9, 1]. The general vector-valued case is still generally open, with recent progress in [3, 2, 15]. Recently, a parallel thread of research has been underway in finance, to understand which financial risk measures, among several in use or proposed to help regulate the risks of financial institutions, are computable via regression, i.e., elicitable (cf. references above). More often than not, these papers have concluded that most risk measures under consideration are not elicitable, notable exceptions being generalized quantiles (e.g. value-at-risk, expectiles) and expected utility [13, 12].

Throughout the growing momentum of the study of elicitation, one question has been central: which properties are elicitable? It is clear, however, that all properties are elicitable if one first elicits the distribution using a standard proper scoring rule. Therefore, in the present work, we suggest replacing this question with a more nuanced one: *how* elicitable are various properties? Specifically, heeding the suggestion of Gneiting [7], we adapt to our setting the notion of *elicitation complexity* introduced by Lambert et al. [17], which captures how many parameters one needs to maintain in an ERM procedure for the property in question. Indeed, if a real-valued property is found not to be elicitable, such as the variance, one should not abandon it, but rather ask how much effort is required to compute it via ERM.

Our work is heavily inspired by the recent progress along these lines of Fissler and Ziegel [15], who show that spectral risk measures of support k have elicitation complexity at most $k+1$. Spectral risk measures are among those under consideration in the finance community, and this result shows that while not elicitable in the classical sense, their elicitation complexity is still low, and hence one can develop reasonable regression procedures for them. Our results extend to these and many other risk measures (see § 3.1.6), often providing matching *lower bounds* on the complexity as well.

Our contributions are the following. We first introduce an adapted definition of elicitation complexity which we believe to be the right notion to focus on going forward. We establish a few simple but useful results which allow for a kind of calculus of elicitation; for example, conditions under which the complexity of eliciting two properties in tandem is the sum of their individual complexities. In § 3, we derive several techniques for proving both upper and lower bounds on elicitation complexity which apply primarily to the *Bayes risks* from decision theory, or optimal expected loss functions. The class includes spectral risk measures among several others; see § 3.1. Finally, in § 4 we turn to the case of *conditional elicitation*, where a property is elicitable if the correct value of another property is known. Complementing the case of the Bayes risk, which is conditionally elicitable, we give a necessary condition for the elicitability of a property together with its conditionee. We conclude with brief remarks and open questions.

2 Preliminaries and Foundation

Let Ω be a set of outcomes and $\mathcal{P} \subseteq \Delta(\Omega)$ be a convex set of probability measures. The goal of elicitation is to learn something about the distribution $p \in \mathcal{P}$, specifically some function $\Gamma(p)$, by minimizing a loss function.

Definition 1. *A* property *is a function $\Gamma : \mathcal{P} \to \mathbb{R}^k$, for some $k \in \mathbb{N}$, which associates a correct report value to each distribution.*[1] *We let $\Gamma_r \doteq \{p \in \mathcal{P} \,|\, r = \Gamma(p)\}$ denote the set of distributions p corresponding to report value r.*

Given a property Γ, we want to ensure that the best result is to reveal the value of the property using a *loss function* that evaluates the report using a sample from the distribution.

Definition 2. *A loss function $L : \mathbb{R}^k \times \Omega \to \mathbb{R}$* elicits *a property $\Gamma : \mathcal{P} \to \mathbb{R}^k$ if for all $p \in \mathcal{P}$, $\Gamma(p) = \operatorname{argsup}_r L(r, p)$, where $L(r, p) \doteq \mathbb{E}_p[L(r, \cdot)]$. A property is* elicitable *if some loss elicits it.*

A well-known necessary condition for elicitability is convexity of the level sets of Γ.

Proposition 1 (Osband [5])**.** *If Γ is elicitable, the level sets Γ_r are convex for all $r \in \Gamma(\mathcal{P})$.*

It is often useful to work with a stronger condition, that not only is Γ_r convex, but it is the intersection of a subspace with $\mathcal{P}$. This condition is equivalent the existence of an *identification function*, a functional describing the level sets of Γ [17, 1].

Definition 3. *A function $V : \mathcal{R} \times \Omega \to \mathbb{R}^k$ is an* identification function *for $\Gamma : \mathcal{P} \to \mathbb{R}^k$, or* identifies *$\Gamma$, if for all $r \in \Gamma(\mathcal{P})$ it holds that $p \in \Gamma_r \iff V(r, p) = 0 \in \mathbb{R}^k$, where as with $L(r, p)$ above we write $V(r, p) \doteq \mathbb{E}_p[V(r, \omega)]$. Γ is* identifiable *if there exists a V identifying it.*

We can now define the classes of identifiable and elicitable properties, along with the complexity of identifying or eliciting a given property.

Definition 4. *Let $\mathcal{I}_k(\mathcal{P})$ denote the class of all identifiable properties $\Gamma : \mathcal{P} \to \mathbb{R}^k$, and $\mathcal{E}_k(\mathcal{P})$ denote the class of all elicitable properties $\Gamma : \mathcal{P} \to \mathbb{R}^k$. We write $\mathcal{I}(\mathcal{P}) = \bigcup_{k \in \mathbb{N}} \mathcal{I}_k(\mathcal{P})$ and $\mathcal{E}(\mathcal{P}) = \bigcup_{k \in \mathbb{N}} \mathcal{E}_k(\mathcal{P})$.*

Definition 5. *A property Γ is k-*identifiable *if there exists $\hat{\Gamma} \in \mathcal{I}_k(\mathcal{P})$ and f such that $\Gamma = f \circ \hat{\Gamma}$. The* identification complexity *of Γ is defined as $\mathsf{iden}(\Gamma) = \min\{k : \Gamma \text{ is } k\text{-identifiable}\}$.*

Definition 6. *A property Γ is k-*elicitable *if there exists $\hat{\Gamma} \in \mathcal{E}_k(\mathcal{P})$ and f such that $\Gamma = f \circ \hat{\Gamma}$. The* elicitation complexity *of Γ is defined as $\mathsf{elic}(\Gamma) = \min\{k : \Gamma \text{ is } k\text{-elicitable}\}$.*
Similarly, Γ is k-elicitable with respect to a class of properties $\mathcal{C}$ if $\hat{\Gamma} \in \mathcal{E}_k(\mathcal{P}) \cap \mathcal{C}$ in the above, and $\mathsf{elic}_{\mathcal{C}}(\Gamma)$ is the corresponding elicitation complexity of Γ with respect to $\mathcal{C}$. In particular we will often use $\mathsf{elic}_{\mathcal{I}}(\Gamma)$, where $\hat{\Gamma}$ must be both elicitable and identifiable.

[1] We will also consider $\Gamma : \mathcal{P} \to \mathbb{R}^{\mathbb{N}}$.

Our definition of elicitation complexity differs from the notion proposed by Lambert et al. [17], in that the components of $\hat{\Gamma}$ above do not need to be individually elicitable. This turns out to have a large impact, as under their definition the property $\Gamma(p) = \max_{\omega\in\Omega} p(\{\omega\})$ for finite Ω has elicitation complexity $|\Omega| - 1$, whereas under our definition $\mathsf{elic}_{\mathcal{I}}(\Gamma) = 2$; see Example 3.1.3. Fissler and Ziegel [15] propose a closer but still different definition, with the complexity being the smallest k such that Γ is a component of a k-dimensional elicitable property. Again, this definition can lead to larger complexities than necessary; take for example the squared mean $\Gamma(p) = \mathbb{E}_p[\omega]^2$ when $\Omega = \mathbb{R}$, which has $\mathsf{elic}_{\mathcal{I}}(\Gamma) = 1$ with $\hat{\Gamma}(p) = \mathbb{E}_p[\omega]$ and $f(x) = x^2$, but is not elicitable and thus has complexity 2 under [15]. We believe that, modulo regularity assumptions on $\mathcal{E}_k(\mathcal{P})$, our definition is better suited to studying the difficulty of eliciting properties: viewing f as a (potentially dimension-reducing) link function, our definition captures the minimum number of parameters needed in an ERM computation of the property in question, followed by a simple one-time application of f.

As noted, we will restrict our attention to $\mathsf{elic}_{\mathcal{I}}$ for much of the paper, which effectively requires $\mathsf{elic}_{\mathcal{I}}(\Gamma) \geq \mathsf{iden}(\Gamma)$; specifically, if Γ is derived from some elicitable $\hat{\Gamma}$, then $\hat{\Gamma}$ must be identifiable as well. This restriction is only relevant for our lower bounds, as our upper bounds give losses explicitly.[2] Note however that *some* restriction on $\mathcal{E}_k(\mathcal{P})$ is necessary, as otherwise pathological constructions giving injective mappings from $\mathbb{R}$ to $\mathbb{R}^k$ would render all properties 1-elicitable. To alleviate this issue, some authors require continuity (e.g. [1]) while others like we do require identifiability (e.g. [15]), which can be motivated by the fact that for any differentiable loss L for Γ, $V(r,\omega) = \nabla_r L(\cdot,\omega)$ will identify Γ provided $\mathbb{E}_p[L]$ has no inflection points or local minima. An important future direction is to relax this identifiability assumption, as there are very natural (set-valued) properties with $\mathsf{iden} > \mathsf{elic}$.[3]

2.1 Foundations of Elicitation Complexity

In the remainder of this section, we make some simple, but useful, observations about $\mathsf{iden}(\Gamma)$ and $\mathsf{elic}_{\mathcal{I}}(\Gamma)$. We have in fact already discussed one such observation: $\mathsf{elic}_{\mathcal{I}}(\Gamma) \geq \mathsf{iden}(\Gamma)$.

It is easy to create redundant properties in various ways. For example, given elicitable properties Γ_1 and Γ_2 the property $\Gamma \doteq \{\Gamma_1, \Gamma_2, \Gamma_1 + \Gamma_2\}$ clearly contains redundant information. A concrete case is $\Gamma = \{$mean squared, variance, 2nd moment$\}$, which has $\mathsf{elic}_{\mathcal{I}}(\Gamma) = 2$; see Example 3.1.1. The following definitions and lemma capture various aspects of a lack of such redundancy. All omitted proofs may be found in the appendix.

Definition 7. *Property $\Gamma : \mathcal{P} \to \mathbb{R}^k$ in $\mathcal{I}(\mathcal{P})$ is of* full rank *if* $\mathsf{iden}(\Gamma) = k$.

Definition 8. *Properties $\Gamma, \Gamma' \in \mathcal{I}(\mathcal{P})$ are* independent *if* $\mathsf{iden}(\{\Gamma, \Gamma'\}) = \mathsf{iden}(\Gamma) + \mathsf{iden}(\Gamma')$.

Lemma 1. *If $\Gamma, \Gamma' \in \mathcal{E}(\mathcal{P})$ are full rank and independent, then* $\mathsf{elic}_{\mathcal{I}}(\{\Gamma, \Gamma'\}) = \mathsf{elic}_{\mathcal{I}}(\Gamma) + \mathsf{elic}_{\mathcal{I}}(\Gamma')$.

Returning to the discussion above, it is well-known that $\mathsf{elic}_{\mathcal{I}}(\text{variance}) = 2$, yet $\Gamma = \{$mean,variance$\}$ has $\mathsf{elic}_{\mathcal{I}}(\Gamma) = 2$, so clearly the mean and variance are not both independent and full rank. (In fact, variance is not full rank.) However, the mean and second moment are.

Clearly, whenever $p \in \mathcal{P}$ can be uniquely determined by some number of elicitable parameters then the elicitation complexity of every property is at most that number. The following propositions give two notable applications of this observation.[4]

Proposition 2. *When $|\Omega| = n$, every property Γ has $\mathsf{elic}_{\mathcal{I}}(\Gamma) \leq n - 1$.*

Proof. The probability distribution is determined by the probability of any $n - 1$ outcomes, and the probability associated with a given outcome is both elicitable and identifiable. □

Proposition 3. *When $\Omega = \mathbb{R}$,[5] every property Γ has $\mathsf{elic}_{\mathcal{I}}(\Gamma) \leq \boldsymbol{\omega}$ (countable).*

One well-studied class of properties are those where Γ is linear, i.e., the expectation of some vector-valued random variable. All such properties are elicitable and identifiable (cf. [4, 8, 3]), with $\mathsf{elic}_{\mathcal{I}}(\Gamma) \leq k$, but of course the complexity can be lower if Γ is not full rank.

[2] Our main lower bound (Thm 2) merely requires Γ to have convex level sets, which is necessary by Prop. 1.

[3] One may take for example $\Gamma(p) = \operatorname{argmax}_i p(A_i)$ for a finite measurable partition $A_1, \ldots, A_n$ of Ω.

[4] Note that these restrictions on Ω may easily be placed on $\mathcal{P}$ instead; e.g. finite Ω is equivalent to $\mathcal{P}$ having support on a finite subset of Ω, or even being piecewise constant on some disjoint events.

[5] Here and throughout, when $\Omega = \mathbb{R}^k$ we assume the Borel σ-algebra.

Lemma 2. *Let $X : \Omega \to \mathbb{R}^k$ be $\mathcal{P}$-integrable and $\Gamma(p) = \mathbb{E}_p[X]$. Then* $\mathsf{elic}_\mathcal{I}(\Gamma) = \dim(\mathrm{affhull}(\Gamma(\mathcal{P})))$*, the dimension of the affine hull of the range of* Γ.

Another important case is the when Γ consists of some number of distinct quantiles. Osband [5] essentially showed that quantiles are independent and of full rank, so their elicitation complexity is the number of quantiles being elicited.

Lemma 3. *Let $\Omega = \mathbb{R}$ and $\mathcal{P}$ be a class of probability measures with continuously differentiable and invertible CDFs F, which is sufficiently rich in the sense that for all $x_1, \ldots, x_k \in \mathbb{R}$, $\mathrm{span}(\{F^{-1}(x_1), \ldots, F^{-1}(x_k)\}, F \in \mathcal{P}) = \mathbb{R}^k$. Let q_α, denote the α-quantile function. Then if $\alpha_1, \ldots, \alpha_k$ are all distinct, $\Gamma = \{q_{\alpha_1}, \ldots, q_{\alpha_k}\}$ has* $\mathsf{elic}_\mathcal{I}(\Gamma) = k$.

The quantile example in particular allows us to see that all complexity classes, including $\boldsymbol{\omega}$, are occupied. In fact, our results to follow will show something stronger: even for *real-valued* properties $\Gamma : \mathcal{P} \to \mathbb{R}$, all classes are occupied; we give here the result that follows from our bounds on spectral risk measures in Example 3.1.4, but this holds for many other $\mathcal{P}$; see e.g. Example 3.1.2.

Proposition 4. *Let $\mathcal{P}$ as in Lemma 3. Then for all $k \in \mathbb{N}$ there exists $\gamma : \mathcal{P} \to \mathbb{R}$ with* $\mathsf{elic}_\mathcal{I}(\gamma) = k$.

3 Eliciting the Bayes Risk

In this section we prove two theorems that provide our main tools for proving upper and lower bounds respectively on elicitation complexity. Of course many properties are known to be elicitable, and the losses that elicit them provide such an upper bound for that case. We provide such a construction for properties that can be expressed as the pointwise minimum of an indexed set of functions. Interestingly, our construction does not elicit the minimum directly, but as a joint elicitation of the value and the function that realizes this value. The form (1) is that of a scoring rule for the linear property $p \mapsto \mathbb{E}_p[X_a]$, except that here the index a itself is also elicited[6].

Theorem 1. *Let $\{X_a : \Omega \to \mathbb{R}\}_{a \in \mathcal{A}}$ be a set of $\mathcal{P}$-integrable functions indexed by $\mathcal{A} \subseteq \mathbb{R}^k$. Then if $\inf_a \mathbb{E}_p[X_a]$ is attained, the property $\gamma(p) = \min_a \mathbb{E}_p[X_a]$ is $(k+1)$-elicitable. In particular,*

$$L((r,a),\omega) = H(r) + h(r)(X_a - r) \tag{1}$$

elicits $p \mapsto \{(\gamma(p), a) : \mathbb{E}_p[X_a] = \gamma(p)\}$ for any strictly decreasing $h : \mathbb{R} \to \mathbb{R}_+$ with $\frac{d}{dr}H = h$.

Proof. We will work with gains instead of losses, and show that $S((r,a),\omega) = g(r) + dg_r(X_a - r)$ elicits $p \mapsto \{(\gamma(p), a) : \mathbb{E}_p[X_a] = \gamma(p)\}$ for $\gamma(p) = \max_a \mathbb{E}_p[X_a]$. Here g is convex with strictly increasing and positive subgradient dg. For any fixed a, we have by the subgradient inequality,

$$S((r,a),p) = g(r) + dg_r(\mathbb{E}_p[X_a] - r) \le g(\mathbb{E}_p[X_a]) = S((\mathbb{E}_p[X_a], a), p) \ ,$$

and as dg is strictly increasing, g is strictly convex, so $r = \mathbb{E}_p[X_a]$ is the unique maximizer. Now letting $\tilde{S}(a,p) = S((\mathbb{E}_p[X_a], a), p)$, we have

$$\operatorname*{argmax}_{a \in \mathcal{A}} \tilde{S}(a,p) = \operatorname*{argmax}_{a \in \mathcal{A}} g(\mathbb{E}_p[X_a]) = \operatorname*{argmax}_{a \in \mathcal{A}} \mathbb{E}_p[X_a] \ ,$$

because g is strictly increasing. We now have

$$\operatorname*{argmax}_{a \in \mathcal{A}, r \in \mathbb{R}} S((r,a),p) = \left\{ (\mathbb{E}_p[X_a], a) : a \in \operatorname*{argmax}_{a \in \mathcal{A}} \mathbb{E}_p[X_a] \right\} \ . \qquad \square$$

One natural way to get such an indexed set of functions is to take an arbitrary loss function $L(r,\omega)$, in which case this pointwise minimum corresponds to the *Bayes risk*, which is simply the minimum possible expected loss under some distribution p.

[6] As we are focused on the complexity of elicitation, we have not tried to fully characterize all ways to elicit this joint property (or other properties we give explicit losses for). See Section 3.1.1 for an example where additional losses are possible.

Definition 9. *Given loss function* $L : \mathcal{A} \times \Omega \to \mathbb{R}$ *on some prediction set* $\mathcal{A}$*, the* Bayes risk *of* L *is defined as* $\underline{L}(p) := \inf_{a \in \mathcal{A}} L(a, p)$.

One illustration of the power of Theorem 1 is that the Bayes risk of a loss eliciting a k-dimensional property is itself $(k+1)$-elicitable.

Corollary 1. *If* $L : \mathbb{R}^k \times \Omega \to \mathbb{R}$ *is a loss function eliciting* $\Gamma : \mathcal{P} \to \mathbb{R}^k$*, then the loss*

$$L((r, a), \omega) = L'(a, \omega) + H(r) + h(r)(L(a, \omega) - r) \tag{2}$$

elicits $\{\underline{L}, \Gamma\}$*, where* $h : \mathbb{R} \to \mathbb{R}_+$ *is any positive strictly decreasing function,* $H(r) = \int_0^r h(x)dx$*, and* L' *is any surrogate loss eliciting* Γ.[7] *If* $\Gamma \in \mathcal{I}_k(\mathcal{P})$*,* $\mathsf{elic}_{\mathcal{I}}(\underline{L}) \le k + 1$.

We now turn to our second theorem which provides lower bounds for the elicitation complexity of the Bayes risk. A first observation, which follows from standard convex analysis, is that $\underline{L}$ is concave, and thus it is unlikely to be elicitable directly, as the level sets of $\underline{L}$ are likely to be non-convex. To show a lower bound greater than 1, however, we will need much stronger techniques. In particular, while $\underline{L}$ must be concave, it may not be strictly so, thus enabling level sets which are potentially amenable to elicitation. In fact, $\underline{L}$ *must* be flat between any two distributions which share a minimizer. Crucial to our lower bound is the fact that whenever the minimizer of L *differs* between two distributions, $\underline{L}$ is essentially strictly concave between them.

Lemma 4. *Suppose loss* L *with Bayes risk* $\underline{L}$ *elicits* $\Gamma : \mathcal{P} \to \mathbb{R}^k$*. Then for any* $p, p' \in \mathcal{P}$ *with* $\Gamma(p) \neq \Gamma(p')$*, we have* $\underline{L}(\lambda p + (1-\lambda)p') > \lambda \underline{L}(p) + (1-\lambda)\underline{L}(p')$ *for all* $\lambda \in (0, 1)$.

With this lemma in hand we can prove our lower bound. The crucial insight is that an identification function for the Bayes Risk of a loss eliciting a property can, through a link, be used to identify that property. The construction from Corollary 1 increases the dimension of the elicitation by 1, and our lower bound shows this is often necessary. However, it is not always, as in the case of linear properties the property value provides all the information required to compute the Bayes risk for some choices of proper loss; for example, dropping the y^2 term from squared loss gives $L(x, y) = x^2 - 2xy$ and $\underline{L}(p) = -\mathbb{E}_p[y]^2$. Thus the theorem splits the lower bound into two cases.

Theorem 2. *If* $\Gamma \in \mathcal{E}_k(\mathcal{P})$ *is elicited by loss* L *and has* $\mathsf{elic}_{\mathcal{I}}(\Gamma) = k$*, then the expected loss* $\underline{L} : p \mapsto L(\Gamma(p), p)$ *has* $\mathsf{elic}_{\mathcal{I}}(\underline{L}) \ge k$*. Moreover, if there is a function* $f : \mathbb{R}^k \to \mathbb{R}$ *such that* $\underline{L} = f \circ \Gamma$*, then* $\mathsf{elic}_{\mathcal{I}}(\underline{L}) = k$*; otherwise,* $\mathsf{elic}_{\mathcal{I}}(\underline{L}) = k + 1$.

Proof. Let $\hat{\Gamma} \in \mathcal{E}_\ell$ such that $\underline{L} = g \circ \hat{\Gamma}$ for some $g : \mathbb{R}^\ell \to \mathbb{R}$.

We show by contradiction that for all $p, p' \in \mathcal{P}$, $\hat{\Gamma}(p) = \hat{\Gamma}(p')$ implies $\Gamma(p) = \Gamma(p')$. Otherwise, we have p, p' with $\hat{\Gamma}(p) = \hat{\Gamma}(p')$, and thus $\underline{L}(p) = \underline{L}(p')$, but $\Gamma(p) \neq \Gamma(p')$. Lemma 4 would then give us some $p_\lambda = \lambda p + (1-\lambda)p'$ with $\underline{L}(p_\lambda) > \underline{L}(p)$. But as the level sets $\hat{\Gamma}_{\hat{r}}$ are convex by Prop. 1, we would have $\hat{\Gamma}(p_\lambda) = \hat{\Gamma}(p)$, which would imply $\underline{L}(p_\lambda) = \underline{L}(p)$.

We now can conclude that there exists $h : \mathbb{R}^\ell \to \mathbb{R}^k$ such that $\Gamma = h \circ \hat{\Gamma}$. But as $\hat{\Gamma} \in \mathcal{E}_\ell$, this implies $\mathsf{elic}_{\mathcal{I}}(\Gamma) \le \ell$, so clearly we need $\ell \ge k$. Finally, if $\ell = k$ we have $\underline{L} = g \circ \hat{\Gamma} = g \circ h^{-1} \circ \Gamma$. The upper bounds follow from Corollary 1. ∎

3.1 Examples and Applications

We now give several applications of our results. Several upper bounds are novel, as well as all lower bounds greater than 2. In the examples, unless we refer to Ω explicitly we will assume $\Omega = \mathbb{R}$ and write $y \in \Omega$ so that $y \sim p$. In each setting, we also make several standard regularity assumptions which we suppress for ease of exposition — for example, for the variance and variantile we assume finite first and second moments (which span $\mathbb{R}^2$), and whenever we discuss quantiles we will assume that $\mathcal{P}$ is as in Lemma 3, though we will not require as much regularity for our upper bounds.

[7] Note that one could easily lift the requirement that Γ be a function, and allow $\Gamma(p)$ to be the set of minimizers of the loss (cf. [18]). We will use this additional power in Example 3.1.4.

3.1.1 Variance

It is well-known that the variance $\sigma^2(p) = \mathbb{E}_p[(\mathbb{E}_p[y] - y)^2]$ is not elicitable, as its level sets are not convex, a necessary condition by Prop. 1. Of course, one can recover σ^2 as a link of the linear (and thus elicitable) property $(\mathbb{E}_p[y], \mathbb{E}_p[y^2])$, and hence $\mathsf{elic}_{\mathcal{I}}(\sigma^2) = 2$. To warm up, we show how to recover this result using our results on the Bayes risk. We can view σ^2 as the Bayes risk of squared loss $L(x, y) = (x - y)^2$, which of course elicits the mean: $\underline{L}(p) = \min_{x\in\mathbb{R}} \mathbb{E}_p[(x - y)^2] = \mathbb{E}_p[(\mathbb{E}_p[y] - y)^2] = \sigma^2(p)$. This gives us $\mathsf{elic}_{\mathcal{I}}(\sigma^2) \leq 2$ by Corollary 1, with a matching lower bound by Theorem 2, as the variance is not simply a function of the mean. Corollary 1 gives losses such as $L(x, v, y) = e^{-v}((x - y)^2 - v) - e^{-v}$ which elict $\{\mathbb{E}_p[y], \sigma^2(p)\}$, but in fact there are losses which cannot be represented by the form (2), showing that we do not have a full characterization; for example, $\hat{L}(x, v, y) = v^2 + v(x - y)(2(x + y) + 1) + (x - y)^2 \left((x + y)^2 + x + y + 1\right)$. This $\hat{L}$ was generated via squared loss $\left\| z - \begin{bmatrix} y \\ y^2 \end{bmatrix} \right\|^2$ with respect to the norm $\|z\|^2 = z^\top \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix} z$, which elicits the first two moments, and link function $(z_1, z_2) \mapsto (z_1, z_2 - z_1^2)$.

3.1.2 Convex Functions of Means

Another simple example is $\gamma(p) = G(\mathbb{E}_p[X])$ for some strictly convex function $G : \mathbb{R}^k \to \mathbb{R}$ and $\mathcal{P}$-integrable $X : \Omega \to \mathbb{R}^k$. To avoid degeneracies, we assume $\dim \operatorname{affhull}\{\mathbb{E}_p[X] : p \in \mathcal{P}\} = k$, i.e. Γ is full rank. Letting $\{dG_p\}_{p\in\mathcal{P}}$ be a selection of subgradients of G, the loss $L(r, \omega) = -(G(r) + dG_r(X(\omega) - r))$ elicits $\Gamma : p \mapsto \mathbb{E}_p[X]$, and moreover we have $\gamma(p) = -\underline{L}(p)$. By Lemma 2, $\mathsf{elic}_{\mathcal{I}}(\Gamma) = k$. One easily checks that $\underline{L} = G \circ \Gamma$, so now by Theorem 2, $\mathsf{elic}_{\mathcal{I}}(\gamma) = k$ as well. Letting $\{X_k\}_{k\in\mathbb{N}}$ be a family of such "full rank" random variables, this gives us a sequence of real-valued properties $\gamma_k(p) = \|\mathbb{E}_p[X]\|^2$ with $\mathsf{elic}_{\mathcal{I}}(\gamma_k) = k$, proving Proposition 4.

3.1.3 Modal Mass

With $\Omega = \mathbb{R}$ consider the property $\gamma_\beta(p) = \max_{x\in\mathbb{R}} p([x - \beta, x + \beta])$, namely, the maximum probability mass contained in an interval of width 2β. Theorem 1 easily shows $\mathsf{elic}_{\mathcal{I}}(\gamma_\beta) \leq 2$, as $\hat{\gamma}_\beta(p) = \operatorname{argmax}_{x\in\mathbb{R}} p([x - \beta, x + \beta])$ is elicited by $L(x, y) = \mathbb{1}_{|x-y|>\beta}$, and $\gamma_\beta(p) = 1 - \underline{L}(p)$. Similarly, in the case of finite Ω, $\gamma(p) = \max_{\omega\in\Omega} p(\{\omega\})$ is simply the expected score (gain rather than loss) of the mode $\gamma(p) = \operatorname{argmax}_{\omega\in\Omega} p(\{\omega\})$, which is elicitable for finite Ω.

In both cases, one can easily check that the level sets of γ are not convex, so $\mathsf{elic}_{\mathcal{I}}(\gamma) = 2$; alternatively Theorem 2 applies in the first case. As mentioned following Definition 6, the result for finite Ω differs from the definitions of Lambert et al. [17], where the elicitation complexity of γ is $|\Omega| - 1$.

3.1.4 Expected Shortfall and Other Spectral Risk Measures

One important application of our results on the elicitation complexity of the Bayes risk is the elicitability of various financial risk measures. One of the most popular financial risk measures is *expected shortfall* $\mathrm{ES}_\alpha : \mathcal{P} \to \mathbb{R}$, also called *conditional value at risk (CVaR)* or *average value at risk (AVaR)*, which we define as follows (cf. [19, eq.(18)], [20, eq.(3.21)]):

$$\mathrm{ES}_\alpha(p) = \inf_{z\in\mathbb{R}} \left\{\mathbb{E}_p\left[\tfrac{1}{\alpha}(z - y)\mathbb{1}_{z\geq y} - z\right]\right\} = \inf_{z\in\mathbb{R}} \left\{\mathbb{E}_p\left[\tfrac{1}{\alpha}(z - y)(\mathbb{1}_{z\geq y} - \alpha) - y\right]\right\} . \tag{3}$$

It was recently shown by Fissler and Ziegel [15] that $\mathsf{elic}_{\mathcal{I}}(\mathrm{ES}_\alpha) \leq 2$. They also consider the broader class of *spectral risk measures*, which can be represented as $\rho_\mu(p) = \int_{[0,1]} \mathrm{ES}_\alpha(p) d\mu(\alpha)$, where μ is a probability measure on $[0, 1]$ (cf. [19, eq. (36)]). In the case where μ has finite support $\mu = \sum_{i=1}^k \beta_i \delta_{\alpha_i}$ for point distributions δ, $\beta_i > 0$, we can rewrite ρ_μ using the above as:

$$\rho_\mu(p) = \sum_{i=1}^k \beta_i \mathrm{ES}_{\alpha_i}(p) = \inf_{z\in\mathbb{R}^k} \left\{\mathbb{E}_p\left[\sum_{i=1}^k \frac{\beta_i}{\alpha_i}(z_i - y)(\mathbb{1}_{z_i\geq y} - \alpha_i) - y\right]\right\} . \tag{4}$$

They conclude $\mathsf{elic}_{\mathcal{I}}(\rho_\mu) \leq k + 1$ unless $\mu(\{1\}) = 1$ in which case $\mathsf{elic}_{\mathcal{I}}(\rho_\mu) = 1$. We show how to recover these results together with matching lower bounds. It is well-known that the infimum in eq. (4) is attained by any of the k quantiles in $q_{\alpha_1}(p), \ldots, q_{\alpha_k}(p)$, so we conclude $\mathsf{elic}_{\mathcal{I}}(\rho_\mu) \leq k+1$ by Theorem 1, and in particular the property $\{\rho_\mu, q_{\alpha_1}, \ldots, q_{\alpha_k}\}$ is elicitable. The family of losses from Corollary 1 coincide with the characterization of Fissler and Ziegel [15] (see § E.1). For a lower bound, as $\mathsf{elic}_{\mathcal{I}}(\{q_{\alpha_1}, \ldots, q_{\alpha_k}\}) = k$ whenever the α_i are distinct by Lemma 3, Theorem 2 gives us $\mathsf{elic}_{\mathcal{I}}(\rho_\mu) = k+1$ whenever $\mu(\{1\}) < 1$, and of course $\mathsf{elic}_{\mathcal{I}}(\rho_\mu) = 1$ if $\mu(\{1\}) = 1$.

3.1.5 Variantile

The τ-expectile, a type of generalized quantile introduced by Newey and Powell [21], is defined as the solution $x = \mu_\tau$ to the equation $\mathbb{E}_p[|\mathbb{1}_{x \geq y} - \tau|(x-y)] = 0$. (This also shows $\mu_\tau \in \mathcal{I}_1$.) Here we propose the *$\tau$-variantile*, an asymmetric variance-like measure with respect to the τ-expectile: just as the mean is the solution $x = \mu$ to the equation $\mathbb{E}_p[x-y] = 0$, and the variance is $\sigma^2(p) = \mathbb{E}_p[(\mu - y)^2]$, we define the τ-variantile σ^2_τ by $\sigma^2_\tau(p) = \mathbb{E}_p\left[|\mathbb{1}_{\mu_\tau \geq y} - \tau|(\mu_\tau - y)^2\right]$.

It is well-known that μ_τ can be expressed as the minimizer of a *asymmetric least squares* problem: the loss $L(x,y) = |\mathbb{1}_{x \geq y} - \tau|(x-y)^2$ elicits μ_τ [21, 7]. Hence, just as the variance turned out to be a Bayes risk for the mean, so is the τ-variantile for the τ-expectile:

$$\mu_\tau = \operatorname*{argmin}_{x \in \mathbb{R}} \mathbb{E}_p\left[|\mathbb{1}_{x \geq y} - \tau|(x-y)^2\right] \implies \sigma^2_\tau = \min_{x \in \mathbb{R}} \mathbb{E}_p\left[|\mathbb{1}_{x \geq y} - \tau|(x-y)^2\right] .$$

We now see the pair $\{\mu_\tau, \sigma^2_\tau\}$ is elicitable by Corollary 1, and by Theorem 2 we have $\mathsf{elic}_{\mathcal{I}}(\sigma^2_\tau) = 2$.

3.1.6 Deviation and Risk Measures

Rockafellar and Uryasev [20] introduce "risk quadrangles" in which they relate a risk $\mathcal{R}$, deviation $\mathcal{D}$, error $\mathcal{E}$, and a statistic $\mathcal{S}$, all functions from random variables to the reals, as follows:

$$\mathcal{R}(X) = \min_C \{C + \mathcal{E}(X - C)\}, \quad \mathcal{D}(X) = \min_C \{\mathcal{E}(X-C)\}, \quad \mathcal{S}(X) = \operatorname*{argmin}_C \{\mathcal{E}(X-C)\} .$$

Our results provide tight bounds for many of the risk and deviation measures in their paper. The most immediate case is the *expectation quadrangle* case, where $\mathcal{E}(X) = \mathbb{E}[e(X)]$ for some $e : \mathbb{R} \to \mathbb{R}$. In this case, if $\mathcal{S}(X) \in \mathcal{I}_1(\mathcal{P})$ Theorem 2 implies $\mathsf{elic}_{\mathcal{I}}(\mathcal{R}) = \mathsf{elic}_{\mathcal{I}}(\mathcal{D}) = 2$ provided $\mathcal{S}$ is non-constant and e non-linear. This includes several of their examples, e.g. truncated mean, log-exp, and rate-based. Beyond the expectation case, the authors show a Mixing Theorem, where they consider

$$\mathcal{D}(X) = \min_C \min_{B_1,..,B_k} \left\{ \sum_{i=1}^k \lambda_i \mathcal{E}_i(X - C - B_i) \,\Big|\, \sum_i \lambda_i B_i = 0 \right\} = \min_{B'_1,..,B'_k} \left\{ \sum_{i=1}^k \lambda_i \mathcal{E}_i(X - B'_i) \right\} .$$

Once again, if the $\mathcal{E}_i$ are all of expectation type and $\mathcal{S}_i \in \mathcal{I}_1$, Theorem 1 gives $\mathsf{elic}_{\mathcal{I}}(\mathcal{D}) = \mathsf{elic}_{\mathcal{I}}(\mathcal{R}) \leq k+1$, with a matching lower bound from Theorem 2 provided the $\mathcal{S}_i$ are all independent. The Reverting Theorem for a pair $\mathcal{E}_1, \mathcal{E}_2$ can be seen as a special case of the above where one replaces $\mathcal{E}_2(X)$ by $\mathcal{E}_2(-X)$. Consequently, we have tight bounds for the elicitation complexity of several other examples, including superquantiles (the same as spectral risk measures), the quantile-radius quadrangle, and optimized certainty equivalents of Ben-Tal and Teboulle [22].

Our results offer an explaination for the existence of regression procedures for some of these risk/deviation measures. For example, a proceedure called *superquantile regression* was introduced in Rockafellar et al. [23], which computes spectral risk measures. In light of Theorem 1, one could interpret their procedure as simply performing regression on the k different quantiles as well as the Bayes risk. In fact, our results show that any risk/deviation generated by mixing several expectation quadrangles will have a similar procedure, in which the B'_i variables are simply computed along side the measure of interest. Even more broadly, such regression procedures exist for *any* Bayes risk.

4 Conditional Elicitation

When considering a non-elicitable property Γ, it is sometimes the case that Γ would be elicitable if only the value of some other elicitable property Γ' were known. This is the notion of *conditional elicitability*, introduced by Emmer et al. [11], who showed that the variance and expected shortfall are both conditionally elicitable, on $\mathbb{E}_p[y]$ and $q_\alpha(p)$ respectively. Intuitively, knowing that Γ is elicitable conditional on an elicitable Γ' would suggest that perhaps the pair $\{\Gamma, \Gamma'\}$ is elicitable; Fissler and Ziegel [15] note that it is an open question whether this joint elicitability holds in general. We now extend our definitions to this setting, and provide insights as to when conditionally elicitable properties are jointly elicitable in the sense above, with both positive and negative examples.

Definition 10. *Given $\Gamma' \in \mathcal{E}_{k'}(\mathcal{P})$, we define $\mathcal{E}_k(\mathcal{P}|\Gamma')$ to be the set of properties $\Gamma : \mathcal{P} \to \mathbb{R}^k$ such that $\Gamma|_{\Gamma'_r} \in \mathcal{E}_k(\Gamma'_r)$ for all $r \in \Gamma'(\mathcal{P})$. That is, for any r, Γ restricted to the set Γ'_r is elicitable. As before we define $\mathcal{I}_k(\mathcal{P}|\Gamma')$ similarly and $\mathcal{E}(\mathcal{P}|\Gamma') = \bigcup_{k\in\mathbb{N}} \mathcal{E}_k(\mathcal{P}|\Gamma')$, $\mathcal{I}(\mathcal{P}|\Gamma') = \bigcup_{k\in\mathbb{N}} \mathcal{I}_k(\mathcal{P}|\Gamma')$.*

Definition 11. *We say Γ is k-*elicitable conditioned on Γ' *if $\Gamma|_{\Gamma'_{r'}}$ is k-elicitable for all $r' \in \Gamma'(\mathcal{P})$, and* $\mathsf{elic}(\Gamma|\Gamma') = \min\{k : \Gamma$ *is k-elicitable conditioned on Γ'*$\}$. *We define* $\mathsf{elic}_{\mathcal{I}}(\Gamma|\Gamma')$ *similarly.*

A first observation, also made by Fissler and Ziegel [15], is that any conditionally identifiable property is jointly identifiable by simply combining the identification functions.

Lemma 5. *Let $\Gamma' \in \mathcal{I}_{k'}(\mathcal{P})$ and $\Gamma \in \mathcal{I}_k(\mathcal{P}|\Gamma')$. Then $\{\Gamma, \Gamma'\} \in \mathcal{I}_{k+k'}(\mathcal{P})$.*

A natural question is then, are conditionally elicitable properties jointly elicitable? Our main theorem in this section is the following result, which gives a necessary condition for identifiability to imply joint elicitability. We apply techniques from previous work, and as such need two regularity/smoothness assumptions which we detail in the appendix.

Theorem 3. *Let $\Gamma' \in \mathcal{E}(\mathcal{P})$ identified by differentiable $V'(r', \omega)$ and $\Gamma \in \mathcal{E}(\mathcal{P}|\Gamma')$ identified by differentiable $V(r, r', \omega)$ conditioned on Γ'. If $W(r, r', \omega) = \{V'(r', \omega), V(r, r', \omega)\}$ satisfies Assumption 1 and $\partial_{r'} V(r, r', p)$ is not constant on $\Gamma_r \cap \Gamma'_{r'}$, but both $\partial_{r'} V'(r', p)$ and $\partial_r V(r, r', p)$ are, the pair $\{\Gamma', \Gamma\}$ is not elicitable by any twice differentiable loss L which satisfies Assumption 2.*

This result confirms the non-elicitability of Example 1 of [3], which upon examination is conditionally elicitable on a linear property. Theorem 3 also allows us to show in Corollary 2 that the central moment $\mu_n(p) = \mathbb{E}_p[(y - \mathbb{E}_p[y])^n]$, while elicitable conditioned on the mean $\mathbb{E}_p[y]$, is not jointly elicitable. Of course, properties *are* jointly elicitable with their Bayes risk, and one easily checks that ∇V is constant within level sets in that case. Theorem 3 suggests that in fact these may be the only such properties, which would be an intuitive result: Bayes risks for Γ are "incentive-aligned" with Γ, allowing us to elicit them together, but this may be necessary as well.

Corollary 2. *Let $n \geq 3$, and let $\Omega = \mathbb{R}$ and $\mathcal{P}$ such that the first n moments are finite and span $\mathbb{R}^n$. Then the property $\{\mu, \mu_{n_1}, \mu_{n_2}, \ldots, \mu_{n_k}\}$ is not elicitable by any twice differentiable loss which satisfies Assumption 2 for $n_i \in \{2, \ldots, n\}$ and $k < n - 1$. In particular, this applies to $\{\mu, \mu_n\}$.*

By analogy to Lemma 5, one might expect elicitability to be subadditive in the sense that $\mathsf{elic}_{\mathcal{I}}(\Gamma) \leq \mathsf{elic}_{\mathcal{I}}(\Gamma|\Gamma') + \mathsf{elic}_{\mathcal{I}}(\Gamma')$, and in fact this holds with equality for any Bayes risk from Theorem 2: if L elicits Γ' then $\mathsf{elic}_{\mathcal{I}}(\underline{L}) = \mathsf{elic}_{\mathcal{I}}(\underline{L}|\Gamma') + \mathsf{elic}_{\mathcal{I}}(\Gamma')$, as we actually have $\mathsf{elic}_{\mathcal{I}}(\underline{L}|\Gamma') = 0$ when $\underline{L} = f \circ \Gamma'$. In light of Corollary 2 and the following Proposition, however, we conjecture $\mathsf{elic}_{\mathcal{I}}(\mu_n) = n$ for all $n \geq 0$, which would provide a counter-example as $\mathsf{elic}_{\mathcal{I}}(\mu_n|\mu) = 1$ and $\mathsf{elic}_{\mathcal{I}}(\mu) = 1$. If this conjecture were the case, we would also have an unbounded gap between iden and $\mathsf{elic}_{\mathcal{I}}$.

Proposition 5. *If the central moment is elicitable via $\Gamma \in \mathcal{I}_2(\mathcal{P})$, then $\Gamma = g \circ \{\mu, \mu_n\}$.*

5 Discussion

We have outlined a theory of elicitation complexity which we believe is the right notion of complexity for ERM, and provided techniques and results for upper and lower bounds. In particular, we now have tight bounds for the large class of Bayes risks, including several applications of note such as spectral risk measures. Our results also offer an explanation for why procedures like superquantile regression are possible, and extend this logic to all Bayes risks. There are also a number of natural open problem in elicitation complexity. For example, is the elicitation complexity of the nth central moment equal to n? Are there conditionally elicitable properties other than Bayes risks which are jointly elicitable? What is the elicitation complexity of the mode and other properties whose non-elicitability is known? Finally, the most general open question remains a full characterization of elicitable vector-valued properties and the losses eliciting them.

References

[1] Ingo Steinwart, Robert Williamson, Siyu Zhang, and others. Elicitation and Identification of Properties. In *Proceedings of The 27th Conference on Learning Theory*, pages 482–526, 2014.

[2] A. Agarwal and S. Agrawal. On Consistent Surrogate Risk Minimization and Property Elicitation. In *COLT*, 2015.

[3] Rafael Frongillo and Ian Kash. Vector-Valued Property Elicitation. In *Proceedings of the 28th Conference on Learning Theory*, 2015.

[4] L.J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, pages 783–801, 1971.

[5] Kent Harold Osband. *Providing Incentives for Better Cost Forecasting*. University of California, Berkeley, 1985.

[6] T. Gneiting and A.E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[7] T. Gneiting. Making and Evaluating Point Forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

[8] J. Abernethy and R. Frongillo. A characterization of scoring rules for linear properties. In *Proceedings of the 25th Conference on Learning Theory*, 2012.

[9] N.S. Lambert. Elicitation and Evaluation of Statistical Forecasts. *Preprint*, 2011.

[10] N.S. Lambert and Y. Shoham. Eliciting truthful answers to multiple-choice questions. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 109–118, 2009.

[11] Susanne Emmer, Marie Kratz, and Dirk Tasche. What is the best risk measure in practice? A comparison of standard measures. 2013.

[12] Fabio Bellini and Valeria Bignozzi. Elicitable risk measures. *This is a preprint of an article accepted for publication in Quantitative Finance (doi 10.1080/14697688.2014. 946955)*, 2013.

[13] Johanna F. Ziegel. Coherence and elicitability. *Mathematical Finance*, 2014. arXiv: 1303.1690.

[14] Ruodu Wang and Johanna F. Ziegel. Elicitable distortion risk measures: A concise proof. *Statistics & Probability Letters*, 100:172–175, May 2015.

[15] Tobias Fissler and Johanna F. Ziegel. Higher order elicitability and Osband's principle. *arXiv:1503.08123 [math, q-fin, stat]*, March 2015. arXiv: 1503.08123.

[16] A. Banerjee, X. Guo, and H. Wang. On the optimality of conditional expectation as a Bregman predictor. *IEEE Transactions on Information Theory*, 51(7):2664–2669, July 2005.

[17] N.S. 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, 2008.

[18] Rafael Frongillo and Ian Kash. General truthfulness characterizations via convex analysis. In *Web and Internet Economics*, pages 354–370. Springer, 2014.

[19] Hans Fllmer and Stefan Weber. The Axiomatic Approach to Risk Measures for Capital Determination. 2015.

[20] R. Tyrrell Rockafellar and Stan Uryasev. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, 18(1):33–53, 2013.

[21] Whitney K. Newey and James L. Powell. Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, pages 819–847, 1987.

[22] Aharon Ben-Tal and Marc Teboulle. AN OLD-NEW CONCEPT OF CONVEX RISK MEASURES: THE OPTIMIZED CERTAINTY EQUIVALENT. *Mathematical Finance*, 17(3):449–476, 2007.

[23] R. T. Rockafellar, J. O. Royset, and S. I. Miranda. Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. *European Journal of Operational Research*, 234:140–154, 2014.

A Short Proofs

Proof of Corollary 1. The only nontrivial part is showing $\{\underline{L}, \Gamma\} \in \mathcal{I}(\mathcal{P})$. Let $V(a,\omega)$ identify Γ. Then $V'((r,a),\omega) = \{V(a,\omega), L(a,\omega) - r\}$ identifies $\{\underline{L}, \Gamma\}$. □

Proof of Lemma 1. Let $\Gamma : \mathcal{P} \to \mathbb{R}^k$ and $\Gamma' : \mathcal{P} \to \mathbb{R}^{k'}$. Unfolding our definitions, we have $\mathsf{elic}_{\mathcal{I}}(\{\Gamma, \Gamma'\}) \geq \mathsf{iden}(\{\Gamma, \Gamma'\}) = \mathsf{iden}(\Gamma) + \mathsf{iden}(\Gamma') = k + k'$. For the upper bound, we simply take losses L and L' for Γ and Γ', respectively, and elicit $\{\Gamma, \Gamma'\}$ via $\hat{L}(r, r', \omega) = L(r,\omega) + L'(r,\omega)$. □

Proof of Proposition 3. We will simply show how to compute the CDF F of p, using only countably many parameters. Let $\{q_i\}_{i\in\mathbb{N}}$ be an enumeration of the rational numbers, and $\hat{\Gamma}(F)_i = F(q_i)$. We can elicit $\hat{\Gamma}$ with the loss $L(\{r_i\}_{i\in\mathbb{N}}, y) = \sum_{i\in\mathbb{N}} \beta^i (r_i - \mathbb{1}_{y \leq q_i})^2$ for $0 < \beta < 1$. We now have F at every rational number, and by right-continuity of F we can compute F at irrationals. Thus, we can compute F, and then $\Gamma(F)$. □

Proof of Lemma 2. Let $\ell = \dim \operatorname{affhull}(\Gamma(\mathcal{P}))$ and $r_0 \in \operatorname{relint}(\Gamma(\mathcal{P}))$. Then $\mathcal{V} = \operatorname{span}\{\Gamma(p) - r_0 : p \in \mathcal{P}\}$ is a vector space of dimension ℓ and basis $v_1, \ldots, v_\ell$. Let $M = [v_1 \ldots v_\ell] \in \mathbb{R}^{k\times\ell}$. Now define $V : \Gamma(\mathcal{P}) \times \Omega \to \mathbb{R}^\ell$ by $V(r,\omega) = M^+(X(\omega) - r)$. Clearly $\mathbb{E}_p[X] = r \implies V(r,p) = 0$, and by properties of the pseudoinverse M^+, as $\mathbb{E}_p[X] - r \in \operatorname{im} M$, $M^+(\mathbb{E}_p[X] - r) = 0 \implies \mathbb{E}_p[X] - r = 0$. Thus $\mathsf{iden}(\Gamma) \leq \ell$. As $\dim \operatorname{span}(\{V(r,p) : p \in \mathcal{P}\}) = \dim \mathcal{V} = \ell$, by Lemma 8, $\mathsf{iden}(\Gamma) = \ell$.

Elicitability follows by letting $\Gamma'(p) = M^+(\mathbb{E}_p[X] - r_0) = \mathbb{E}_p[M^+(X - r_0)] \in \mathbb{R}^\ell$ with link $f(r') = Mr' + r_0$; Γ' is of course elicitable as a linear property. □

Proof of Lemma 5. Let $V'(r',\omega)$ identify Γ' and $V(r,r',\omega)$ identify Γ conditioned on Γ'. Let $W(r,r',\omega) = \{V'(r',\omega), V(r,r',\omega)\}$. Then $W(r,r',p) = 0 \iff V'(r',p) = 0 \wedge V(r,r',p) = 0 \iff \Gamma'(p) = r' \wedge V(r,r',p) = 0 \iff \Gamma'(p) = r' \wedge \Gamma(p) = r$. □

Proof of Corollary 2. We label the components of r from 0 for convenience. Take $V(r_0, y) = y - r_0$ and $V'(r,y) = \{(y - r_0)^{n_i} - r_i\}_{i=1}^k$. One easily checks that these functions satisfy the conditions of Theorem 3 for $\Gamma = \mu$ and $\Gamma' = \{\mu_{n_1}, \mu_{n_2}, \ldots, \mu_{n_k}\}$; in particular $\partial_{r_0} V = -1$ and $\nabla_{r_{1..k}} V' = -I_k$, the $k \times k$ identity matrix. Thus we need only compute $\partial_{r_0} V'(r,p) = \{\mathbb{E}_p[-n_i(y - r_0)^{n_i - 1}]\}_{i=1}^k = \{-n_i \mu_{n_i - 1}(p)\}_{i=1}^k$. By assumption, the first n moments span $\mathbb{R}^n$ on $\mathcal{P}$, and one can check that $\{\mu, \mu_2, \ldots, \mu_n\}$ constitute a change of basis and hence also span $\mathbb{R}^n$. Thus, letting $N = \{1, n_1, \ldots, n_k\}$ be the indices for the moments covered by V, the only way for $\partial_{r_0} V'(r,p)$ to be constant for all $p \in \Gamma_r$ is if $n_i - 1 \in N$ for all i. But the only way for this to hold is $N = \{1, \ldots, n\}$, contradicting $k < n - 1$. □

B Proof of Lemma 4

Lemma 6 ([18])**.** *Let $G : X \to \mathbb{R}$ convex for some convex subset X of a vector space $\mathcal{V}$, and let $d \in \partial G_x$ be a subgradient of G at x. Then for all $x' \in X$ we have*

$$d \in \partial G_{x'} \iff G(x) - G(x') = d(x - x') \ .$$

Lemma 7. *Let $G : X \to \mathbb{R}$ convex for some convex subset X of a vector space $\mathcal{V}$. Let $x, x' \in X$ and $x_\lambda = \lambda x + (1-\lambda)x'$ for some $\lambda \in (0,1)$. If there exists some $d \in \partial G_{x_\lambda} \setminus (\partial G_x \cup \partial G_{x'})$, then $G(x_\lambda) < \lambda G(x) + (1-\lambda)G(x')$.*

Proof. By the subgradient inequality for d at x_λ we have $G(x) - G(x_\lambda) \geq d(x - x_\lambda)$, and furthermore Lemma 6 gives us $G(x) - G(x_\lambda) > d(x - x_\lambda)$ since otherwise we would have $d \in \partial G_x$. Similarly for x', we have $G(x') - G(x_\lambda) > d(x' - x_\lambda)$.

Adding λ of the first inequality to $(1-\lambda)$ of the second gives

$$\begin{aligned} \lambda G(x) + (1-\lambda)G(x') - G(x_\lambda) &> \lambda d(x - x_\lambda) + (1-\lambda) d(x' - x_\lambda) \\ &= \lambda(1-\lambda) d(x - x') + (1-\lambda)\lambda d(x' - x) = 0 \ , \end{aligned}$$

where we used linearity of d and the identity $x_\lambda = x' + \lambda(x - x')$. □

Restatement of Lemma 4: Suppose loss L with Bayes risk $\underline{L}$ elicits $\Gamma : \mathcal{P} \to \mathbb{R}^k$. Then for any $p, p' \in \mathcal{P}$ with $\Gamma(p) \neq \Gamma(p')$, we have $\underline{L}(\lambda p + (1-\lambda)p') > \lambda \underline{L}(p) + (1-\lambda)\underline{L}(p')$ for all $\lambda \in (0,1)$.

Proof. Let $G = -\underline{L}$, which is the expected score function for the (positively-oriented) scoring rule $S = -L$. By Theorem 3.5 Frongillo and Kash [18] , we have some $\mathcal{D} \subseteq \partial G$ and function $\varphi : \Gamma(\mathcal{P}) \to \mathcal{D}$ such that $\Gamma(p) = \varphi^{-1}(\mathcal{D} \cap \partial G_p)$. In other words, as our Γ is a function, there is a subgradient $d_r = \varphi(r)$ associated to each report value $r \in \Gamma(\mathcal{P})$, and $d_r \in \partial G_p \iff r = \Gamma(p)$. Thus, as we have $p, p' \in \mathcal{P}$ with $r = \Gamma(p) \neq \Gamma(p') = r'$, we also have $d_r \in \partial G_p \setminus \partial G_{p'}$ and $d_{r'} \in \partial G_{p'} \setminus \partial G_p$.

By Lemma 7, if $\Gamma(p_\lambda)$, $\Gamma(p)$, and $\Gamma(p')$ are all distinct, then we are done. Otherwise, we have $\Gamma(p_\lambda) = \Gamma(p)$ without loss of generality, which implies $d_r \in \partial G_{p_\lambda}$ by definition of φ. Now assume for a contradiction that $G(p_\lambda) = \lambda G(p) + (1-\lambda)G(p')$. By Lemma 6 for d_r we have $G(p) - G(p_\lambda) = d_r(p - p_\lambda) = \frac{(1-\lambda)}{\lambda} d_r(p_\lambda - p')$. Solving for $G(p)$ and substituting into the previous equation gives $(1-\lambda)$ times the equation $G(p_\lambda) = d_r(p_\lambda - p') + G(p')$, and applying Lemma 6 one more gives $d_r \in \partial G_{p'}$, a contradiction. □

C Proofs for Conditional Elicitation

We will need some assumptions about some of the identification functions and losses we will use to guarantee they are applicable. The first assumption essentially says that $\mathcal{P}$ is sufficiently rich, and is required to reply a result due to Fissler and Ziegel [15]. The second assumption is that we can work directly with expected losses, and has previously been used by Frongillo and Kash [3].

Assumption 1. *Let V identify Γ. For all $r \in int(\Gamma(\mathcal{P})) \subseteq \mathbb{R}^k$ there exist $p_1, \ldots, p_{k+1}$ such that $0 \in int(\mathsf{Conv}(\{V(r,p_1), \ldots, V(r,p_{k+1})\}))$.*

Assumption 2. *Twice differentiable loss L satisfies $\mathbb{E}_p[\nabla_r L(r,\omega)] = \nabla_r \mathbb{E}_p[L(r,\omega)]$ and similarly for second derivatives.*

Proof of Theorem 3. Let $\Gamma : \mathcal{P} \to \mathbb{R}^k$, $\Gamma' : \mathcal{P} \to \mathbb{R}^{k'}$. Then by the proof of Lemma 5, $W(r, r', \omega) = \{V'(r', \omega), V(r, r', \omega)\}$ identifies $\{\Gamma, \Gamma'\}$. Letting L be a twice differentiable loss eliciting $\{\Gamma, \Gamma'\}$. By assumption 1, the conditions of [15, Theorem 3.2] (commonly known as Osband's principle [5]) apply, and we may write the Jacobian of L as $\nabla L(r, r', p) = H(r, r')W(r, r', p)$ where $H(r, r') \in \mathbb{R}^{(k+k') \times (k+k')}$ and $W(r, r', p) \in \mathbb{R}^{(k+k')}$. Differentiating L once more (using assumption 2), and letting $p \in \Gamma_r \cap \Gamma'_{r'}$, we have $\nabla^2 L(r, r', p) = \nabla H(r, r')W(r, r', p) + H(r, r')\nabla W(r, r', p) = H(r, r')\nabla W(r, r', p)$, as the first term is zero due to identifiability. Now note that we may write

$$\nabla W(r, r', p) = \begin{bmatrix} \partial_r V(r,p) & 0 \\ \partial_{r'} V(r, r', p) & \partial_r V(r, r', p) \end{bmatrix} = \begin{bmatrix} X(r,p) & 0 \\ Y(r, r', p) & Z(r, r', p) \end{bmatrix} .$$

Writing H in block form $[A, B; C, D]$ we have

$$\begin{aligned} \nabla^2 L(r, r', p) &= \begin{bmatrix} A(r,r') & B(r,r') \\ C(r,r') & D(r,r') \end{bmatrix} \begin{bmatrix} X(r,p) & 0 \\ Y(r,r',p) & Z(r,r',p) \end{bmatrix} \\ &= \begin{bmatrix} A(r,r')X(r,p) + B(r,r')Y(r,r',p) & B(r,r')Z(r,r',p) \\ C(r,r')X(r,p) + D(r,r')Y(r,r',p) & D(r,r')Z(r,r',p) \end{bmatrix} . \end{aligned}$$

By assumption, we have some $p, p' \in \Gamma_r \cap \Gamma'_{r'}$ such that $Y(r, r', p) \neq Y(r, r', p')$, but $X(r,p) = X(r,p')$ and $Z(r, r', p) = Z(r, r', p')$. As $\nabla^2 L$ must be symmetric for both p, p', we must have $CX + DY = Z^\top B^\top$ for both as well. By strict optimality of (r, r'), we also know that $\nabla^2 L$ must be positive definite, and thus the block diagonal elements are also positive definite, and both D and Z, being square, are of full rank. This tells us $Y = D^{-1}(Z^\top B^\top - CX)$, which cannot hold for both p and p' as all terms are fixed except Y. □

C.1 Proof of Proposition 5

Proof. Let $m_k(p) = \mathbb{E}_p[y^k]$ denote the kth raw moment of p. Then we may write $\mu_n(p) = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} m_k(p) m_1(p)^{n-k}$. Suppose $\mu_n = f \circ \Gamma$ for $\Gamma : \mathcal{P} \to \mathbb{R}^2$ identified by V and elicited by a twice differentiable loss L.

Fix $p' \in \mathcal{P}$, $r = \Gamma(p')$, and consider $p \in \Gamma_r$. By assumption we must have $\mu_n(p) = \mu_n(p') = f(r)$, which by convexity of Γ_r implies $\mu_n(\lambda p + (1-\lambda)p') = f(r)$ for all $\lambda \in [0,1]$. Expanding this equation out yields a polynomial in λ, which by linearity of the m_k becomes,

$$f(r) = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} (m'_k + \lambda(m_k - m'_k))(m'_1 + \lambda(m_1 - m'_1))^{n-k} \ , \tag{5}$$

where we write $m_k = m_k(p)$ and $m'_k = m_k(p')$. For this polynomial to be constant in λ on $[0,1]$, the coefficients of nonzero powers of λ must be zero. Expanding the first few terms of eq. (5) gives,

$$\begin{aligned} f(r) &= (-1)^n \binom{n}{0} (m'_0 + \lambda(m_0 - m'_0))(m'_1 + \lambda(m_1 - m'_1))^n \\ &\quad + (-1)^{n-1} \binom{n}{1} (m'_1 + \lambda(m_1 - m'_1))(m'_1 + \lambda(m_1 - m'_1))^{n-1} + O(\lambda^{n-1}) \\ &= (-1)^{n-1}(n-1)(m'_1 + \lambda(m_1 - m'_1))^n + O(\lambda^{n-1}) \ , \end{aligned}$$

where we used $m_0 = m'_0 = 1$ and $O(\lambda^{n-1})$ encompasses the remaining terms. Now, in particular, the coefficient of λ^n must be zero, which gives us $(-1)^{n-1}(n-1)(m_1 - m'_1) = 0$, yielding $m_1 = m'_1$. As $m_1(p)$ is constant for all $p \in \Gamma_r$, we have some $c_1(r)$ such that

$$\mathbb{E}_p[y - c_1(r)] = 0 \ . \tag{6}$$

Now the constraint that $\mu_n(p) = f(r)$ may be simplified to

$$f(r) = \mu_n(p) = \mathbb{E}_p[(y - \mathbb{E}_p[y])^n] = \mathbb{E}_p[(y - c_1(r))^n] \ . \tag{7}$$

Thus, by Lemma 9, $\Gamma \in \mathcal{I}_2$ is identified by $V(r, y) = \{y - c_1(r), (y - c_1(r))^n - f(r)\}$. Letting $g(r) = \{c_1(r), f(r)\}$, we can see that g must be invertible: if $g(r) = g(r')$ for $r \neq r'$, then as V identifies Γ but depends on r only through g, we have $\Gamma_r = \{p \in \mathcal{P} : V(r, p) = 0\} = \Gamma_{r'}$, a contradiction. We conclude $\Gamma = g^{-1} \circ \Gamma'$ for $\Gamma'(p) = \{m_1(p), \mu_n(p)\}$. □

D Identification Lower Bounds

Lemma 8. *Let $\Gamma \in \mathcal{I}(\mathcal{P})$ be given, and suppose for some $r \in \Gamma(\mathcal{P})$ there exists $V : \Omega \to \mathbb{R}^k$ with $\mathbb{E}_p[V] = 0$ for all $p \in \Gamma_r$. If $\mathrm{span}(\{\mathbb{E}_p[V] : p \in \mathcal{P}\}) = \mathbb{R}^k$ and some $p \in \Gamma_r$ can be written $p = \lambda p' + (1-\lambda)p''$ where $p', p'' \notin \Gamma_r$, then $\mathsf{iden}(\Gamma) \geq k$.*

Proof. The proof proceeds in two parts. First, we show that the conditions regarding V suffice to show that $\mathrm{codim}(\mathrm{span}(\Gamma_r)) \geq k$ in $\mathrm{span}(\mathcal{P})$. Second, we show that this means (any flat subset of) Γ_r cannot be identified by a $W : \mathrm{span}(\mathcal{P}) \to \mathbb{R}^\ell$ for $\ell < k$.

Let V and r as in the statment of the lemma be given. By definition, $\mathrm{codim}(\mathrm{span}(\Gamma_r)) = \dim(\mathrm{span}(\mathcal{P})/\mathrm{span}(\Gamma_r))$, where S_1/S_2 is the quotient space of S_1 by S_2. Let $\pi_{\Gamma_r} : \mathrm{span}(\mathcal{P}) \to \mathrm{span}(\mathcal{P})/\mathrm{span}(\Gamma_r)$ denote the projection from $\mathrm{span}(\mathcal{P})$ to its quotient by $\mathrm{span}(\Gamma_r)$. By the universal property of quotient spaces, there is a unique $T_V : \mathrm{span}(\mathcal{P})/\mathrm{span}(\Gamma_r) \to \mathbb{R}^k$ such that $V = T_V \circ \pi_{\Gamma_r}$. By the rank nullity theorem, $\dim(\mathrm{span}(\mathcal{P})/\mathrm{span}(\Gamma_r)) = \dim(\ker(T_V)) + \dim(\mathrm{im}(T_V))$. By assumption $\dim(\mathrm{im}(T_V)) = \dim(\mathrm{im}(V)) = k$, so $\mathrm{codim}(\mathrm{span}(\Gamma_r)) \geq k$.

Now assume for contradictiont that $\Gamma = f \circ \hat{\Gamma}$, with $\hat{\Gamma} \in \mathcal{I}_\ell(\mathcal{P})$, with $\ell < k$. Let r' denote the level set such that $p \in \hat{\Gamma}_{r'}$. Since $\hat{\Gamma}_{r'} \subseteq \Gamma_r$, $\mathrm{codim}(\hat{\Gamma}_{r'}) \geq \mathrm{codim}(\Gamma_r) \geq k$. Let $W : \mathrm{span}(\mathcal{P}) \to \mathbb{R}^\ell$ identify $\hat{\Gamma}_{r'}$. As before, there is a unique $T_W : \mathrm{span}(\mathcal{P})/\mathrm{span}(\hat{\Gamma}_{r'}) \to \mathbb{R}^\ell$ such that $W = T_W \circ \pi_{\hat{\Gamma}_{r'}}$. By the rank nullity theorem, $\dim(\mathrm{span}(\mathcal{P})/\mathrm{span}(\hat{\Gamma}_{r'})) = \dim(\ker(T_W)) + \dim(\mathrm{im}(T_W))$. Thus $\dim(\ker(T_W)) \geq k - \ell > 0$. To complete the proof we need to show that this means there is a $q \in \mathcal{P} - \hat{\Gamma}_{r'}$ such that $\pi_{\hat{\Gamma}_{r'}}(q) \in \ker(T_W)$.

To this end, let $q'' \in \mathrm{span}(\mathcal{P})$. Then $q'' = \sum_i \lambda_i q_i$, with $q_i \in \mathcal{P}$ for all i. Thus $\pi_{\hat{\Gamma}_{r'}}(q'') = \pi_{\hat{\Gamma}_{r'}}(\sum_i \lambda_i q_i) = \sum_i \lambda_i \pi_{\hat{\Gamma}_{r'}}(q_i)$ and so $\mathrm{span}(\mathcal{P})/\mathrm{span}(\hat{\Gamma}_{r'}) = \mathrm{span}(\{\pi_{\hat{\Gamma}_{r'}}(q') \mid q' \in \mathcal{P}\})$. Since $p = \lambda p' + (1-\lambda)p''$ where $p', p'' \notin \Gamma_r$, $\pi_{\hat{\Gamma}_{r'}}(p) = 0$ is not an extreme point of the convex set $\{\pi_{\hat{\Gamma}_{r'}}(q') \mid q' \in \mathcal{P}\}$. Since $\dim(\ker(T_W)) > 0$, this means there exists $q \in \mathcal{P} - \hat{\Gamma}_{r'}$ such that $\pi_{\hat{\Gamma}_{r'}}(q) \in \ker(T_W)$. This contradicts the assumption that W identifies $\hat{\Gamma}_{r'}$, completing the proof. □

Lemma 9. *Let $V : \Gamma(\mathcal{P}) \times \Omega \to \mathbb{R}^k$ identify Γ, and suppose for all $r \in \operatorname{relint}(\Gamma(\mathcal{P}))$ there exists $p, p' \in \mathcal{P}$ and $\lambda \in (0,1)$ such that $r = \Gamma(\lambda p + (1-\lambda)p') \neq \Gamma(p)$ and $\operatorname{span}(\{\mathbb{E}_p[V(r,\omega)] : p \in \mathcal{P}\}) = \mathbb{R}^k$. Let $u : \Gamma(\mathcal{P}) \times \Omega \to \mathbb{R}$ be given. If for all $r \in \Gamma(\mathcal{P})$ we have $\Gamma(p) = r \implies \mathbb{E}_p[u(r,\omega)] = 0$ and $\mathbb{E}_p[u(r,\omega)] \neq 0$ for some $p \in \mathcal{P}$, then there exists $V' : \Gamma(\mathcal{P}) \times \Omega \to \mathbb{R}^k$ identifying Γ with $V'_1 = u$.*

Proof. Fix $r \in \operatorname{relint}(\Gamma(\mathcal{P}))$. As in Lemma 8 we will treat functions $f : \Omega \to \mathbb{R}^\ell$ as linear maps from $\operatorname{span}(\mathcal{P})$ to $\mathbb{R}^\ell$, so that $\operatorname{im} f = \{\mathbb{E}_p[f] : p \in \operatorname{span}(\mathcal{P})\}$.

Let $U : \Omega \to \mathbb{R}^{k+1}$ be given by $U(\omega) = \{V(r,\omega), u(r,\omega)\}$. If we have $\operatorname{im} U = \mathbb{R}^{k+1}$, then Lemma 8 gives us a contradiction with $V(r,\cdot) : \Omega \to \mathbb{R}^k$. Thus $\dim \operatorname{im} U = k$, and there exists some $\alpha \in \mathbb{R}^{k+1}, \alpha \neq 0$, such that $\alpha^\top U = 0$ on $\operatorname{span}(\mathcal{P})$. As $\dim \operatorname{im} V(r,\cdot) = k$, we cannot have $\alpha_{k+1} = 0$, and as $u \neq 0$ on $\operatorname{span}(\mathcal{P})$, we must have some $i \neq k+1$ with $\alpha_i \neq 0$. Taking $\alpha_i = -1$ without loss of generality, we have $V_i = U_i = \sum_{j \neq i} \alpha_j U_j$ on $\operatorname{span}(\mathcal{P})$. Taking $V'(r,\cdot) = \{u(r,\cdot)\} \cup \{V_j(r,\cdot)\}_{j \neq i}$, we have for all $p \in \mathcal{P}$, $\mathbb{E}_p[V'(r,\omega)] = 0 \iff \forall j \neq i \; \mathbb{E}_p[U_j] = 0 \iff \mathbb{E}_p[U] = 0 \iff \mathbb{E}_p[V(r,\omega)] = 0$. □

E Other Omitted Material

E.1 Losses for Expected Shortfall

Corollary 1 gives us a large family of losses eliciting $\{\mathrm{ES}_\alpha, q_\alpha\}$ (see footnote 7). Letting $L(a,y) = \frac{1}{\alpha}(a - y)\mathbb{1}_{a \geq y} - a$, we have $\mathrm{ES}_\alpha(p) = \inf_{a \in \mathbb{R}} L(a,p) = \underline{L}(p)$. Thus may take

$$L((r,a),y) = L'(a,y) + H(r) + h(r)(L(a,y) - r) \;, \tag{8}$$

where $h(r)$ is positive and decreasing, $H(r) = \int_0^r h(x)dx$, and $L'(a,y)$ is any other loss for q_α, the full characterization of which is given in Gneiting [7, Theorem 9]:

$$L'(a,y) = (\mathbb{1}_{a \geq y} - \alpha)(f(a) - f(y)) + g(y) \;, \tag{9}$$

where is $f : \mathbb{R} \to \mathbb{R}$ is nondecreasing and g is an arbitrary $\mathcal{P}$-integrable function.[8] Hence, losses of the following form suffice:

$$L((r,a),y) = (\mathbb{1}_{a \geq y} - \alpha)(f(a) - f(y)) + \frac{1}{\alpha}h(r)\mathbb{1}_{a \geq y}(a - y) - h(r)(a + r) + H(r) + g(y) \;.$$

Comparing our $L((r,a),y)$ to the characterization given by Fissler and Ziegel [15, Cor. 5.5], we see that we recover all possible scores for this case (at least when restricting to $\mathcal{P}$ which). Note however that due to a differing convention in the sign of ES_α, their loss is given by $L((-x_1, x_2), y)$.

[8] Note that Gneiting [7] assumes $L(x,y) \geq 0$, $L(x,x) = 0$, L is continuous in x, dL/dx exists and is continuous in x when $y \neq x$; we add g because we do not normalize.