On Elicitation Complexity

Rafael Frongillo  
University of Colorado, Boulder  
raf@colorado.edu

Ian A. Kash  
Microsoft Research  
iankash@microsoft.com

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.

1 Introduction

Empirical risk minimization (ERM) is a dominant 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 “indirectly” 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 many parameters are 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 § 4.6), 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 § 4. We conclude with brief remarks and open questions.

2 Preliminaries and Foundation

Let \( \Omega \) 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 \( \Gamma(p) \) such as the mean or variance, by minimizing a loss function.

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

Given a property \( \Gamma \), 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 \( L : \mathbb{R}^k \times \Omega \rightarrow \mathbb{R} \) elicits a property \( \Gamma : \mathcal{P} \rightarrow \mathbb{R}^k \) if for all \( p \in \mathcal{P} \), \( \Gamma(p) = \text{argmin}_r L(r, p) \), where \( L(r, p) = \mathbb{E}_p[L(r, \cdot)] \). A property is elicitable if some loss elicits it.

For example, when \( \Omega = \mathbb{R} \), the mean \( \Gamma(p) = \mathbb{E}_p[\omega] \) is elicitable via squared loss \( L(r, \omega) = (r - \omega)^2 \).

A well-known necessary condition for elicitation is convexity of the level sets of \( \Gamma \).

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

One can easily check that the mean \( \Gamma(p) = \mathbb{E}_p[\omega] \) has convex level sets, yet the variance \( \Gamma(p) = \mathbb{E}_p[(\omega - \mathbb{E}_p[\omega])^2] \) does not, and hence is not elicitable [9].

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

**Definition 3.** A function \( V : \mathbb{R} \times \Omega \rightarrow \mathbb{R}^k \) is an identification function for \( \Gamma : \mathcal{P} \rightarrow \mathbb{R}^k \), 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 \( V(r, p) = \mathbb{E}_p[V(r, \omega)] \). \( \Gamma \) is identifiable if there exists a \( V \) identifying it.

One can check for example that \( V(r, \omega) = \omega - r \) identifies the mean.

We can now define the classes of identifiable and elicitable properties, along with the complexity of identifying or eliciting a given property. Naturally, a property is \( k \)-identifiable if it is the link of a \( k \)-dimensional identifiable property, and \( k \)-elicitable if it is the link of a \( k \)-dimensional elicitable property. The elicitation complexity of a property is then simply the minimum dimension \( k \) needed for it to be \( k \)-elicitable.

**Definition 4.** Let \( \mathcal{I}_k(\mathcal{P}) \) denote the class of all identifiable properties \( \Gamma : \mathcal{P} \rightarrow \mathbb{R}^k \), and \( \mathcal{E}_k(\mathcal{P}) \) denote the class of all elicitable properties \( \Gamma : \mathcal{P} \rightarrow \mathbb{R}^k \). 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 \( \Gamma \) 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 \( \Gamma \) is defined as \( \text{iden}(\Gamma) = \min\{ k : \Gamma \text{ is } k \text{-identifiable} \} \).

\(^1\)We will also consider \( \Gamma : \mathcal{P} \rightarrow \mathbb{R} \).
Definition 6. A property $\Gamma$ 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 $\Gamma$ is defined as $\text{elic}(\Gamma) = \min\{k : \Gamma \text{ is } k\text{-elicitable}\}$.

To make the above definitions concrete, recall that the variance $\sigma^2(p) = \mathbb{E}_p[(\mathbb{E}_p[\omega] - \omega)^2]$ is not elicitable, as its level sets are not convex, a necessity by Prop. 1. Note however that we may write $\sigma^2(p) = \mathbb{E}_p[\omega^2] - \mathbb{E}_p[\omega]^2$, which can be obtained from the property $\hat{\Gamma}(p) = (\mathbb{E}_p[\omega], \mathbb{E}_p[\omega^2])$. It is well-known [4, 7] that $\Gamma$ is both elicitable and identifiable as the expectation of a vector-valued random variable $X(\omega) = (\omega, \omega^2)$, using for example $L(r, \omega) = \|r - X(\omega)\|^2$ and $V(r, \omega) = r - X(\omega)$.

Thus, we can recover $\sigma^2$ as a link of the elicitable and identifiable $\hat{\Gamma} : \mathcal{P} \rightarrow \mathbb{R}^2$, and as no such $\Gamma : \mathcal{P} \rightarrow \mathbb{R}$ exists, we have $\text{iden}(\sigma^2) = \text{elic}(\sigma^2) = 2$.

In this example, the variance has a stronger property than merely being 2-identifiable and 2-elicitable, namely that there is a single $\hat{\Gamma}$ that satisfies both of these simultaneously. In fact this is quite common, and identifiability provides geometric structure that we make use of in our lower bounds. Thus, most of our results use this refined notion of elicitation complexity.

Definition 7. A property $\Gamma$ has (identifiable) elicitation complexity $\text{elic}_\mathcal{E}(\Gamma) = \min\{k : \exists \Gamma, f \text{ such that } \Gamma \in \mathcal{E}_k(\mathcal{P}) \cap \mathcal{I}_k(\mathcal{P}) \text{ and } \Gamma = f \circ \hat{\Gamma}\}$.

Note that restricting our attention to $\text{elic}_\mathcal{E}$ effectively requires $\text{elic}_\mathcal{E}(\Gamma) \geq \text{iden}(\Gamma)$; specifically, if $\Gamma$ 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.\footnote{Our main lower bound (Thm 2) merely requires $\Gamma$ to have convex level sets, which is necessary by Prop. 1.} 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. [11]) while others like we do require identifiability (e.g. [15]), which can be motivated by the fact that for any differentiable loss $L$ for $\Gamma$, $V(r, \omega) = \nabla_r L(r, \omega)$ will identify $\Gamma$ 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 $\text{iden} > \text{elic}$\footnote{One may take for example $\hat{\Gamma}(p) = \arg\max p(A_i)$ for a finite measurable partition $A_1, \ldots, A_n$ of $\Omega$.}.

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 $\Omega$ has elicitation complexity $|\Omega| - 1$, whereas under our definition $\text{elic}_\mathcal{E}(\Gamma) = 2$; see Example 4.3. Fissler and Ziegel [15] propose a closer but still different definition, with the complexity being the smallest $k$ such that $\Gamma$ 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 $\text{elic}_\mathcal{E}(\Gamma) = 1$ 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 $f$.

### 2.1 Foundations of Elicitation Complexity

In the remainder of this section, we make some simple, but useful, observations about $\text{iden}(\Gamma)$ and $\text{elic}_\mathcal{E}(\Gamma)$. We have already discussed one such observation after Definition 7: $\text{elic}_\mathcal{E}(\Gamma) \geq \text{iden}(\Gamma)$.

It is natural to start with some trivial upper bounds. 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.\footnote{Note that these restrictions on $\Omega$ may easily be placed on $\mathcal{P}$ instead; e.g. finite $\Omega$ is equivalent to $\mathcal{P}$ having support on a finite subset of $\Omega$, or even being piecewise constant on some disjoint events.}

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

**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. \(\square\)
It is easy to create redundant properties in various ways. For example, given elicitable properties
Γ = \{\text{mean squared, variance, 2nd moment}\}, which, as we have seen, has elic_{2}(Γ) = 2. The
following definitions and lemma capture various aspects of a lack of such redundancy.

Definition 8. Property Γ : \mathcal{P} \to \mathbb{R}^k in \mathcal{I}(\mathcal{P}) is of full rank if \text{id}(\text{iden}(Γ)) = k.

Note that there are two ways for a property to fail to be full rank. First, as the examples above
suggest, Γ can be “redundant” so that it is a link of a lower-dimensional identifiable property. Full
rank can also be violated if more dimensions are needed to identify the property than to specify it.
This is the case with, e.g., the variance which is a 1 dimensional property but has \text{id}(\text{variance}) = 2.

Definition 9. Properties Γ, Γ′ ∈ \mathcal{I}(\mathcal{P}) are independent if \text{id}(\text{id}(Γ) + \text{id}(Γ′)) = \text{id}(Γ) + \text{id}(Γ′).

Lemma 2. If Γ, Γ′ ∈ \mathcal{E}(\mathcal{P}) are full rank and independent, thenelic_{2}(\{Γ, Γ′\}) = elic_{2}(Γ) + elic_{2}(Γ′).

To illustrate the lemma, elic_{2}(\text{variance}) = 2, yet Γ = \{\text{mean, variance}\} has elic_{2}(Γ) = 2, so clearly
the mean and variance are not both independent and full rank. (As we have seen, variance is not full
rank.) However, the mean and second moment satisfy both by Lemma 1.

Another important case is 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 Ω = \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}, span({F^{-1}(x_1), \ldots, F^{-1}(x_k)}), F \in \mathcal{P} = \mathbb{R}^k. Let q_\alpha, denote the \alpha-quantile function. Then if \alpha_1, \ldots, \alpha_k are all distinct, \Gamma = \{q_{\alpha_1}, \ldots, q_{\alpha_k}\} has elic_{2}(Γ) = k.

The quantile example in particular allows us to see that all complexity classes, including \omega, are
occupied. In fact, our results to follow will show something stronger: even for \text{real-valued} properties
Γ : \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 4.4, but this holds for many other \mathcal{P}; see e.g. Example 4.2.

Proposition 4. Let \mathcal{P} as in Lemma 3. Then for all k ∈ \mathbb{N} there exists γ : \mathcal{P} \to \mathbb{R} with elic_{2}(γ) = 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
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.7

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} ⊆ \mathbb{R}^k. Then if
\inf_a \mathbb{E}_p[X_a] is attained, the property γ(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)
(1)
elicits p \mapsto \{(γ(p), a) : \mathbb{E}_p[X_a] = γ(p)\} for any strictly decreasing h : \mathbb{R} \to \mathbb{R}_+ with \frac{d}{dr}H = h.

7Here and throughout, when Ω = \mathbb{R}^k we assume the Borel \sigma-algebra.
8Omitted proofs can be found in the appendix of the full version of this paper.
One illustration of the power of Theorem 1 is that the Bayes risk of a loss eliciting a potentially amenable to elicitation. In fact, \( \hat{\Gamma}(p) \) is any surrogate loss eliciting \( \Gamma(p) \). We will use this additional power in Example 4.4.

Let then we have giving \( L \) two distributions, a minimizer. Crucial to our lower bound is the fact that whenever the minimizer of \( L \) is essentially strictly concave between them. In particular, while \( L \) is strictly convex, and thus it is unlikely to be elicitable directly, as the level sets of \( L \) are not convex, we will need much stronger techniques. To show a lower bound greater than 1, however, we will need much stronger techniques.

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 \( p \).

**Definition 10.** Given loss function \( L : A \times \Omega \rightarrow \mathbb{R} \) on some prediction set \( A \), the Bayes risk of \( L \) is defined as \( L(p) := \inf_{a \in A} L(a, p) \).

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 \( L : \mathbb{R}^k \times \Omega \rightarrow \mathbb{R} \) is a loss function eliciting \( \Gamma : \mathcal{P} \rightarrow \mathbb{R}^k \), then the loss \( L((r, a), \omega) = L'(a, \omega) + H(r) + h(r)(L(a, \omega) - r) \) elicits \( \{L, \Gamma\} \), where \( h : \mathbb{R} \rightarrow \mathbb{R}_+ \) is any positive strictly decreasing function, \( H(r) = \int_0^r h((x)dx \), and \( L' \) is any surrogate loss eliciting \( \Gamma \).\(^8\) If \( \Gamma \in \mathcal{E}_k(\mathcal{P}) \), \( \text{elicit}^\ell(L) \leq k + 1 \).

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 \( g \) is strictly increasing. A first observation, which follows from standard convex analysis, is that \( g \) is strictly increasing. Because \( g \) is strictly increasing, we have \( \hat{\Gamma}(p) \) is any surrogate loss eliciting \( \Gamma(p) \).

\[ \text{elicit}^\ell(L) \leq k + 1. \]

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. Corollary 1 tells us that \( k + 1 \) parameters suffice for the Bayes risk of a \( k \)-dimensional property, and our lower bound shows this is often necessary. Only \( k \) parameters suffice, however, when the property value itself provides all the information required to compute the Bayes risk; for example, dropping the \( y^2 \) term from squared loss gives \( L(x, y) = x^2 - 2xy \) and \( L(p) = \mathbb{E}_p[y^2] \), giving \( \text{elicit}^\ell(L) = 1 \). Thus the theorem splits the lower bound into two cases.

**Theorem 2.** If a loss \( L \) elicits some \( \Gamma \in \mathcal{E}_k(\mathcal{P}) \) with elicitation complexity \( \text{elicit}^\ell(\Gamma) = k \), then its Bayes risk \( L \) has \( \text{elicit}^\ell(L) \geq k \). Moreover, if we can write \( L = f \circ \Gamma \) for some function \( f : \mathbb{R}^k \rightarrow \mathbb{R} \), then we have \( \text{elicit}^\ell(L) = k \); otherwise, \( \text{elicit}^\ell(L) = k + 1 \).

**Proof.** Let \( \hat{\Gamma} \in \mathcal{E}_k \) such that \( L = g \circ \hat{\Gamma} \) for some \( g : \mathbb{R}^k \rightarrow \mathbb{R} \).

\[^8\text{Note that one could easily lift the requirement that } \Gamma \text{ be a function, and allow } \Gamma(p) \text{ to be the set of minimizers of the loss (cf. [18]). We will use this additional power in Example 4.4.}\]
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 $L(p) = L(p')$, but $\Gamma(p) \neq \Gamma(p')$. Lemma 4 would then give us some $p_\lambda = \lambda p + (1 - \lambda)p'$ with $L(p_\lambda) > L(p)$. But as the level sets $\Gamma_x$ are convex by Prop. 1, we would have $\hat{\Gamma}(p_\lambda) = \hat{\Gamma}(p)$, which would imply $L(p_\lambda) = L(p)$. 

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

4 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 $\Omega$ 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 quantile we assume finite first and second moments (which must 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.

4.1 Variance

In Section 2 we showed that $\text{elic}_\ell(\sigma^2) = 2$. As a warm up, let us see how to recover this statement using our results on the Bayes risk. We can view $\sigma^2$ as the Bayes risk of squared loss $L(x, y) = (x - y)^2$, which of course elicits the mean: $L(p) = \min_{x \in \mathbb{R}} \mathbb{E}_p[(x - y)^2] = \mathbb{E}_p[(\mathbb{E}_p[y] - y)^2] = \sigma^2(p)$. This gives us $\text{elic}_\ell(\sigma^2) \leq 2$ 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 $L(x, y) = e^{-r((x - y)^2 - v)} - e^{-v}$ which elicit $\{\mathbb{E}_p[y], \sigma^2(p)\}$, 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, $L((x, v), y) = v^2 + v(x - y)(2(x + y) + 1) + (x - y)^2((x + y)^2 + x + y + 1)$. This $L$ was generated via squared loss $\|z - \left[\begin{array}{c} y \\ y' \end{array}\right]\|^2$ with respect to the norm $\|z\|^2 = z^T \left[\begin{array}{cc} 1 & -1/2 \\ -1/2 & 1 \end{array}\right] z$, which elicits the first two moments, and link function $(z_1, z_2) \mapsto (z_1, z_2 - z_1^2)$.

4.2 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 $\dim \text{aff hull}\{\mathbb{E}_p[X] : p \in \mathcal{P}\} = k$, i.e. $\Gamma$ 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 $\Gamma : p \mapsto \mathbb{E}_p[X]$ (cf. [3]), and moreover we have $\gamma(p) = -L(p)$. By Lemma 1, $\text{elic}_\ell(\Gamma) = k$. One easily checks that $L = G \circ \gamma$, so now by Theorem 2, $\text{elic}_\ell(\gamma) = k$ 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 $\gamma_k(p) = ||\mathbb{E}_p[X_k]||^2$ with $\text{elic}_\ell(\gamma_k) = k$, proving Proposition 4.

4.3 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 $2\beta$. Theorem 1 easily shows $\text{elic}_\ell(\gamma_\beta) \leq 2$, as $\gamma_\beta(p) = \max_{x \in \mathbb{R}} p([x - \beta, x + \beta])$ is elicited by $L(x, y) = \mathbb{1}_{|x - y| > \beta}$, and $\gamma_\beta(p) = 1 - L(p)$. Similarly, in the case of finite $\Omega$, $\gamma(p) = \max_{\omega \in \Omega} p(\{\omega\})$ is simply the expected score (gain rather than loss) of the mode $\gamma(p) = \max_{\omega \in \Omega} p(\{\omega\})$, which is elicitable for finite $\Omega$ (but not otherwise; see Heinrich [19]).

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

6
4.4 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 ES_α : \mathcal{P} \to \mathbb{R}, also called conditional value at risk (CVaR) or average value at risk (AVaR), which we define as follows (cf. [20, eq.(18)], [21, eq.(3.21)]):

\[ \text{ES}_\alpha(p) = \inf_{z \in \mathbb{R}} \left\{ \mathbb{E}_p \left[ \mathbb{I}_d(z-y) \mathbb{I}_{z \geq y} - z \right] \right\} = \inf_{z \in \mathbb{R}} \left\{ \mathbb{E}_p \left[ \frac{1}{\alpha}(z-y)(\mathbb{I}_{z \geq y} - \alpha) - y \right] \right\} . \]  

(3)

Despite the importance of elicitability to financial regulation [11, 22], ES_α is not elicitable [7]. It was recently shown by Fissler and Ziegel [15], however, that elic_\tau(ES_\alpha) = 2. They also consider the broader class of spectral risk measures, which can be represented as \( \rho_{\mu}(p) = \int_{[0,1]} \text{ES}_\alpha(p)d\mu(z) \), where \( \mu \) is a probability measure on \([0,1]\) (cf. [20, eq. (36)]). In the case where \( \mu \) has finite support \( \mu = \sum_{i=1}^k \beta_i \delta_{\alpha_i} \), for point distributions \( \delta, \beta_i > 0 \), we can rewrite \( \rho_\mu \) using the above as:

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

(4)

They conclude elic_\tau(\rho_\mu) \leq k + 1 unless \( \mu(\{1\}) = 1 \) in which case elic_\tau(\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 elic_\tau(\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 \S D.1). For a lower bound, as elic_\tau(\{q_{\alpha_1}, \ldots, q_{\alpha_k}\}) = k whenever the \( \alpha_i \) are distinct by Lemma 3, Theorem 2 gives us elic_\tau(\rho_\mu) = k + 1 whenever \( \mu(\{1\}) < 1 \), and of course elic_\tau(\rho_\mu) = 1 if \( \mu(\{1\}) = 1 \).

4.5 Variantile

The \( \tau \)-quantile, a type of generalized quantile introduced by Newey and Powell [23], is defined as the solution \( x = \mu_\tau \) to the equation \( \mathbb{E}_p [\mathbb{I}_{x \geq y} \tau (x-y)] = 0 \). (This also shows \( \mu_\tau \in \mathcal{I}_\tau \).) Here we propose the \( \tau \)-variantile, an asymmetric variance-like measure with respect to the \( \tau \)-quantile: just as the mean is the solution \( x = \mu \) to the equation \( \mathbb{E}_p [x-y] = 0 \), and the variance is \( \sigma^2(p) = \mathbb{E}_p [(x-y)^2] \), we define the \( \tau \)-variantile \( \sigma^2_\tau \) by \( \sigma^2_\tau(p) = \mathbb{E}_p [\mathbb{I}_{\mu_\tau \geq y} \tau (\mu_\tau - y)^2] \).

It is well-known that \( \mu_\tau \) can be expressed as the minimizer of an asymmetric least squares problem: the loss \( L(x,y) = \mathbb{I}_{x \geq y} \tau (x-y)^2 \) elicits \( \mu_\tau \) [23, 7]. Hence, just as the variance turned out to be a Bayes risk for the mean, so is the \( \tau \)-quantile for the \( \tau \)-quantile:

\[ \mu_\tau = \arg \min_{x \in \mathbb{R}} \mathbb{E}_p [\mathbb{I}_{x \geq y} \tau (x-y)^2] \implies \sigma^2_\tau = \min_{\tau \in \mathbb{R}} \mathbb{E}_p [\mathbb{I}_{x \geq y} \tau (x-y)^2] . \]

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

4.6 Deviation and Risk Measures

Rockafellar and Uryasev [21] introduce “risk quadrangles” in which they relate a risk \( R \), deviation \( D \), error \( E \), and a statistic \( S \), all functions from random variables to the reals, as follows:

\[ \mathcal{R}(X) = \min_C \{ C + \mathcal{E}(X-C) \}, \quad \mathcal{D}(X) = \min_C \{ \mathcal{E}(X-C) \}, \quad \mathcal{S}(X) = \arg \min_C \{ \mathcal{E}(X-C) \} . \]

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 \( \mathcal{E}(X) = \mathbb{E}[\epsilon(X)] \) for some \( \epsilon : \mathbb{R} \to \mathbb{R} \).

In this case, if \( \mathcal{S}(X) \in \mathcal{I}_\tau(P) \) Theorem 2 implies elic_\tau(\mathcal{R}) = elic_\tau(\mathcal{D}) = 2 provided \( \mathcal{S} \) is non-constant and \( \epsilon \) 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

\[ \mathcal{D}(X) = \min_{B_1, \ldots, B_k} \left\{ \sum_{i=1}^k \lambda_i \mathcal{E}_i(X-C-B_i) \left| \sum_{i=1}^k \lambda_i B_i = 0 \right. \right\} = \min_{B_1, \ldots, B_k} \left\{ \sum_{i=1}^k \lambda_i \mathcal{E}_i(X-B_i) \right\} . \]

Once again, if the \( \mathcal{E}_i \) are all of expectation type and \( S_i \in \mathcal{I}_\tau \), Theorem 1 gives elic_\tau(D) = elic_\tau(R) \leq k + 1, with a matching lower bound from Theorem 2 provided the \( S_i \) 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 $E_2(X)$ by $E_2(-X)$. 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 [24].

Our results offer an explanation for the existence of regression procedures for some of these risk/deviation measures. For example, a procedure called superquantile regression was introduced in Rockafellar et al. [25], 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 alongside the measure of interest. Even more broadly, such regression procedures exist for any Bayes risk.

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 many natural open problems in elicitation complexity. Perhaps the most apparent are the characterizations of the complexity classes $\{\Gamma : \text{elic}(\Gamma) = k\}$, and in particular, determining the elicitation complexity of properties which are known to be non-elicitabile, such as the mode [19] and smallest confidence interval [18].

In this paper we have focused on elicitation complexity with respect to the class of identifiable properties $\mathcal{I}$, which we denoted $\text{elic}_\mathcal{I}$. This choice of notation was deliberate; one may define $\text{elic}_\mathcal{C} := \min\{k : \exists \Gamma \in \mathcal{E}_k \cap \mathcal{C}, \exists f, \Gamma = f \circ \hat{\Gamma}\}$ to be the complexity with respect to some arbitrary class of properties $\mathcal{C}$. Some examples of interest might be $\text{elic}_\mathcal{E}$ for expected values, of interest to the prediction market literature [8], and $\text{elic}_{\text{cvx}}$ for properties elicitable by a loss which is convex in $r$, of interest for efficiently performing ERM.

Another interesting line of questioning follows from the notion of conditional elicitation, properties which are elicitable as long as the value of some other elicitable property is known. This notion was introduced by Emmer et al. [11], who showed that the variance and expected shortfall are both conditionally elicitable, on $E_p[y]$ and $q_\alpha(p)$ respectively. Intuitively, knowing that $\Gamma$ is elicitable conditional on an elicitable $\Gamma'$ would suggest that perhaps the pair $\{\Gamma, \Gamma'\}$ is elicitable; Fissler and Ziegel [15] note that it is an open question whether this joint elicibility holds in general. The Bayes risk $L$ for $\Gamma$ is elicitable conditioned on $\Gamma$, and as we saw above, the pair $\{\Gamma, L\}$ is jointly elicitable as well. We give a counter-example in Figure 1, however, which also illustrates the subtlety of characterizing all elicitable properties.

![Figure 1: Depictions of the level sets of two properties, one elicitable and the other not. The left is a Bayes risk together with its property, and thus elicitable, while the right is shown in [3] not to be elicitable. Here the planes are shown to illustrate the fact that these are both conditionally elicitable: the height of the plane (the intercept $(p_3, 0, 0)$ for example) is elicitable from the characterizations for scalar properties [9, 1], and conditioned on the plane, the properties are both linear and thus links of expected values, which are also elicitable.](image-url)
References


[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.


A  Short Proofs

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

Proof of Lemma 2. Let \( \Gamma : P \rightarrow \mathbb{R}^k \) and \( \Gamma' : P \rightarrow \mathbb{R}^k' \). Unfolding our definitions, we have \( \text{elic}_{\mathbb{R}}(\{\Gamma, \Gamma'\}) \geq \text{idem}(\Gamma) + \text{idem}(\Gamma') = k + k' \). For the upper bound, we simply take losses \( L \) and \( L' \) for \( \Gamma \) and \( \Gamma' \), respectively, and elicit \( \{\Gamma, \Gamma'\} \) via \( L(r, r', \omega) = L(r, \omega) + L'(r, \omega) \).

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 \( \Gamma(P)_i = F(q_i) \). We can elicit \( \hat{\Gamma} \) with the loss \( L((r_i), \omega) = \sum_{i \in \mathbb{N}} \beta^i (r_i - 1_{y \leq q_i})^2 \) 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 1. Let \( \ell = \dim \text{aff hull}(\Gamma(P)) \) and \( r_0 \in \text{relint}(\Gamma(P)) \). Then \( V = \text{span}(\Gamma(p) - r_0 : p \in P) \) is a vector space of dimension \( \ell \) and basis \( v_1, \ldots, v_\ell \). Let \( M = \{v_1, \ldots, v_\ell\} \in \mathbb{R}^{k \times \ell} \). Now define \( V : \Gamma(P) \times \Omega \rightarrow \mathbb{R}^\ell \) by \( V(r, \omega) = M^+(X(\omega) - r) \). Clearly \( E_{P}[X] = r \implies V(r, p) = 0 \), and by properties of the pseudoinverse \( F^+ \), as \( E_{P}[X] - r \in \text{im} M, M^+(E_{P}[X] - r) = 0 = E_{P}[X] - r = 0 \). Thus \( \text{idem}(\Gamma) \leq \ell \). As \( \dim \text{span} \{\{V(r, p) : p \in P\}\} = \dim V = \ell \), by Lemma 7, \( \text{idem}(\Gamma) = \ell \).

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

B  Proof of Lemma 4

Lemma 5 ([18]). Let \( G : X \rightarrow \mathbb{R} \) convex for some convex subset \( X \) of a vector space \( V \), 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 6. Let \( G : X \rightarrow \mathbb{R} \) convex for some convex subset \( X \) of a vector space \( V \). 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 \( x_\lambda \) we have \( G(x) - G(x_\lambda) \geq d(x - x_\lambda) \), and furthermore Lemma 5 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 \( \lambda \) of the first inequality to \( (1 - \lambda) \) of the second gives
\[
\lambda G(x) + (1 - \lambda)G(x') - G(x_\lambda) > \lambda d(x - x_\lambda) + (1 - \lambda)d(x' - x_\lambda) = \lambda(1 - \lambda)d(x - x') + (1 - \lambda)\lambda d(x' - x) = 0 ,
\]
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 \( L \) elicits \( \Gamma : P \rightarrow \mathbb{R}^k \). Then for any \( p, p' \in P \) with \( \Gamma(p) \neq \Gamma(p') \), we have \( L(\lambda p + (1 - \lambda)p') > \lambda L(p) + (1 - \lambda)L(p') \) for all \( \lambda \in (0, 1) \).

Proof. Let \( G = -L \), 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 \( D \subseteq \partial G \) and function \( \varphi : \Gamma(P) \rightarrow \mathcal{D} \) such that \( \Gamma(p) = \varphi^{-1}(D \cap \partial G_p) \). In other words, as our \( \Gamma \) is a function, there is a subgradient \( d_r \in \varphi(r) \) associated to each report value \( r \in \Gamma(P) \), and \( d_r \in \partial G_p \iff r = \Gamma(p) \). Thus, as we have \( p, p' \in P \) 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 6, 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 \( \varphi \). Now
assume for a contradiction that \( G(p) = \lambda G(p) + (1 - \lambda)G(p') \). By Lemma 5 for \( d_r \) we have
\[
G(p) - G(p) = d_r(p - p) = \frac{(1 - \lambda)}{\lambda}d_r(p - p')\]
Solving for \( G(p) \) and substituting into the previous equation gives \( (1 - \lambda) \) times the equation \( G(p) = d_r(p - p') + G(p') \), and applying Lemma 5 one more gives \( d_r \in \partial G(p) \), a contradiction. \( \square \)

**C Identification Lower Bounds**

**Lemma 7.** Let \( \Gamma \in \mathcal{I}(P) \) be given, and suppose for some \( r \in \Gamma(P) \) there exists \( V : \Omega \to \mathbb{R}^k \) with \( \mathbb{E}_p[V] = 0 \) for all \( p \in \Gamma_r \). If \( \text{span}(\{\mathbb{E}_p[V] : p \in P\}) = \mathbb{R}^k \) and some \( p \in \Gamma_r \) can be written \( p = \lambda p' + (1 - \lambda)p'' \) where \( p', p'' \notin \Gamma_r \), then \( \text{iden}(\Gamma) \geq k \).

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

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

Now assume for a contradiction that \( \Gamma = f \circ \hat{\Gamma} \), with \( \hat{\Gamma} \in \mathcal{I}_k(P) \), with \( \ell < k \). Let \( r' \) denote the level set such that \( p \in \hat{\Gamma}_r \). Since \( \hat{\Gamma}_r \subseteq \Gamma_r \), \( \text{codim}(\hat{\Gamma}_r) \geq \text{codim}(\Gamma_r) \geq k \). Let \( W : \text{span}(P) \to \mathbb{R}^\ell \) identify \( \hat{\Gamma}_r \). As before, there is a unique \( T_W : \text{span}(P) / \text{span}(\hat{\Gamma}_r) \to \mathbb{R}^\ell \) such that \( W = T_W \circ \pi_{\hat{\Gamma}_r} \).

By the rank nullity theorem, \( \text{dim}(\text{span}(P) / \text{span}(\hat{\Gamma}_r)) = \text{dim}(\ker(T_W)) + \text{dim}(\text{im}(T_W)) \). Thus \( \text{dim}(\ker(T_W)) = k - \ell > 0 \). To complete the proof we need to show that this means there is a \( q \in P - \hat{\Gamma}_r \) such that \( \pi_{\hat{\Gamma}_r}(q) \in \ker(T_W) \).

To this end, let \( q'' \in \text{span}(P) \). Then \( q'' = \sum_i \lambda_i q_i \), with \( q_i \in P \) for all \( i \). Thus \( \pi_{\hat{\Gamma}_r}(q'') = \pi_{\hat{\Gamma}_r}(\sum \lambda_i q_i) = \sum \lambda_i \pi_{\hat{\Gamma}_r}(q_i) \) and so \( \text{span}(P) / \text{span}(\hat{\Gamma}_r) = \text{span}(\{\pi_{\hat{\Gamma}_r}(q') : q' \in P\}) \). 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 \( \{\pi_{\hat{\Gamma}_r}(q') : q' \in P\} \). Since \( \text{dim}(\ker(T_W)) > 0 \), this means there exists \( q \in P - \hat{\Gamma}_r \) such that \( \pi_{\hat{\Gamma}_r}(q) \in \ker(T_W) \). This contradicts the assumption that \( W \) identifies \( \hat{\Gamma}_r \), completing the proof. \( \square \)

**Lemma 8.** Let \( V : \Gamma(P) \times \Omega \to \mathbb{R}^k \) identify \( \Gamma \), and suppose for all \( r \in \text{relint}(\Gamma(P)) \) there exists \( p, p' \in P \) and \( \lambda \in (0, 1) \) such that \( r = \Gamma((\lambda p + (1 - \lambda)p')) \neq \Gamma(p) \) and \( \text{span}(\{\mathbb{E}_p[V(r, \omega)] : p \in P\}) = \mathbb{R}^k \). Let \( u : \Gamma(P) \times \Omega \to \mathbb{R} \) be given. If for all \( r \in \Gamma(P) \) we have \( \Gamma(p) = r \implies \mathbb{E}_p[u(r, \omega)] = 0 \) and \( \mathbb{E}_p[u(r, \omega)] \neq 0 \) for some \( p \in P \), then there exists \( V' : \Gamma(P) \times \Omega \to \mathbb{R}^k \) identifying \( \Gamma \) with \( \Gamma' \) completing the proof.

**Proof.** Fix \( r \in \text{relint}(\Gamma(P)) \). As in Lemma 7 we will treat functions \( f : \Omega \to \mathbb{R}^\ell \) as linear maps from \( \text{span}(P) \) to \( \mathbb{R}^\ell \), so that \( \text{im} f = \{\mathbb{E}_p[f] : p \in \text{span}(P)\} \).

Let \( U : \Omega \to \mathbb{R}^{k+1} \) be given by \( U(\omega) = \{V(r, \omega), u(r, \omega)\} \). If we have \( \text{im} U = \mathbb{R}^{k+1} \), then Lemma 7 gives us a contradiction with \( V(r, \cdot) : \Omega \to \mathbb{R}^k \). Thus \( \text{dim} \text{im} U = k \), and there exists some \( \alpha \in \mathbb{R}^{k+1}, \alpha \neq 0 \), such that \( \alpha U = 0 \) on \( \text{span}(P) \). As \( \text{dim} V(r, \cdot) = k \), we cannot have \( \alpha \cdot 1_{k+1} = 0 \), and as \( u \neq 0 \) on \( \text{span}(P) \), we must have some \( i \neq k + 1 \) with \( \alpha_i \neq 0 \). Taking \( \alpha_i = -1 \) with loss of generality, we have \( V_i = U_i = \sum_{j \neq i} \alpha_j U_j \) on \( \text{span}(P) \). Taking \( V'(r, \cdot) = \{u(r, \cdot)\} \cup \{V_j(r, \cdot)\}_{j \neq i} \), we have for all \( p \in P \), \( \mathbb{E}_p[V'(r, \omega)] = 0 \iff \forall j \neq i \mathbb{E}_p[U_j] = 0 \iff \mathbb{E}_p[U] = 0 \iff \mathbb{E}_p[V(\omega)] = 0 \). \( \square \)
D Other Omitted Material

D.1 Losses for Expected Shortfall

Corollary 1 gives us a large family of losses eliciting \( \{ \text{ES}_\alpha, q_\alpha \} \) (see footnote 8). Letting \( L(a, y) = \frac{1}{\alpha} (a - y) \mathbb{1}_{a \geq y} - a \), we have \( \text{ES}_\alpha(p) = \inf_{a \in \mathbb{R}} L(a, p) = L(p) \). Thus may take

\[
L((r, a), y) = L'(a, y) + H(r) + h(r)(L(a, y) - r),
\]

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

\[
L'(a, y) = (\mathbb{1}_{a \geq y} - \alpha)(f(a) - f(y)) + g(y),
\]

where is \( f : \mathbb{R} \to \mathbb{R} \) is nondecreasing and \( g \) is an arbitrary \( \mathcal{P} \)-integrable function.\(^9\) 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 the assumptions stated in their Theorem 5.2(iii). Note however that due to a differing convention in the sign of \( \text{ES}_\alpha \), their loss is given by \( L((-x_1, x_2), y) \).

\(^9\)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.