Value of Information, Scoring Rules and Probabilistic Sensitivity Analysis

Emanuele Borgonovo* Gordon Hazen†
Victor Richmond R. Jose‡ Elmar Plischke§

January 19, 2018

We consider an analyst using a mathematical model of a scenario of interest, who seeks to provide probabilistic forecasts for key uncertain quantities, and some measure of sensitivity of the forecasts on parameters/exogenous variables. There are a variety of sensitivity measures available for the associated probabilistic sensitivity analysis, so many that an analyst may find it difficult to choose. We seek to provide some discipline in the choice by supposing an analyst anticipates evaluation of forecast quality by a scoring rule. In this situation, the natural measure of sensitivity is information value under this scoring rule. We show that in fact many established sensitivity measures are already information value under a suitable scoring rule, and develop new explicit formulas for information value under widely used scoring rules. We also examine the question of when a sensitivity measure is information value under some scoring rule. The analyst can then avoid sensitivity measures that cannot be information value, and thereby provide guidance and justification for the choice of the sensitivity measure. A numerical application in the context of a complex forecasting problem demonstrates the approach.

Keywords: Forecast Evaluation, Information Value, Sensitivity Measures, Decision Analysis Cycle

1 Introduction

Sensitivity analysis is recognized as an integral part of the decision analysis cycle (Howard, 1983; Clemen, 1997). In many applications, the decision-making process is informed by

*Bocconi University, Milan, Italy, emanuele.borgonovo@unibocconi.it
†Northwestern University, Evanston, IL, USA, gbh305@northwestern.edu
‡Georgetown University, Washington, DC, USA, vrj2@georgetown.edu
§Clausthal University of Technology, Clausthal-Zellerfeld, Germany, elmar.plischke@tu-clausthal.de

forecasts or predictions from quantitative models. For example, the Defense Advanced Research Project Agency has launched the world-modelers program aimed at developing models to help decision making concerning national and global security[1]. The Intergovernamental Panel for Climatic Change heavily relies on quantitative codes for the identification of climate mitigation strategies (Hu et al., 2012; Marangoni et al., 2017).

An analyst who develops or implements such a model is expected to provide a forecast and some description of the level of uncertainty in that forecast, but is not otherwise involved in comparing alternatives or eliciting preferences. The latter would typically be done informally by an institution via its representative panel or agents (French, 2017). The agents' required forecast might be a point estimate, a quantile, or the complete probability distribution for a key quantity Y of interest. The level of uncertainty in the forecast is commonly conveyed in a sensitivity analysis, in which the sensitivity of Y to uncertain model inputs (parameters or exogenous variables) is examined. The goal of a sensitivity analysis is to identify what collections of model inputs are the drivers of the forecast, and would therefore be candidates for additional information acquisition.

Our point of view, which has been articulated elsewhere (Felli and Hazen, 1998; Oakley, 2009), is that the preferred measure of sensitivity to a collection of $X \subset (X_1, X_2, ..., X_n)$ of model inputs is the information value of X, defined as the greatest reduction in objective value one would accept to permit the choice of alternative to depend on the hypothetically revealed value of X. Of course, this approach requires model inputs to be treated as uncertain quantities and prior distributions assigned. Moreover, an analyst whose forecast may be required in multiple or unanticipated decision contexts would have no access to any specified objective function or set of alternatives, and therefore no apparent way to compute information value. This limitation may also arise when an institution has already selected an alternative a maximizing the expectation of an objective Y_a. Before a is fully implemented, the analyst may be asked to examine the sensitivity of $Y = Y_a$ to model inputs for monitoring the important factors in post-optimality (Eschbach, 1992). In this case, there is again no access to other alternatives because the choice has already been made. Therefore no information value calculation can apparently be made.

This may explain why analysts resort to simple techniques such as tornado diagrams, one- and two-way analyses, that require the analyst to specify only a range of possible values for each parameter, and yield worst-case information. When the *likely* impact of parameter variations is desired rather than the worst-case scenario, the analyst must treat input parameters as random variables and supply probability distributions, often derived from the same data used to construct base estimates or from expert elicitation. Such approaches are known as *probabilistic sensitivity analysis* (Felli and Hazen, 1998). A number of sensitivity indicators have been introduced, which go under the name of probabilistic sensitivity measures. Most probabilistic sensitivity measures have been constructed to quantify the strength of the statistical dependence between Y and X. An analyst searching the literature would find measures based on variance, density

[1] `http://www.darpa.mil/program/world-modelers`

separation, or cumulative distribution function separation, and recently, on quantile separation, just to mention a few. The very variety of available sensitivity measures could be a stumbling block for the analyst: What is the right one to use?

The overarching theme of this paper is that if prior distributions are assigned, then information value can in fact be used as a sensitivity measure in spite of the absence of objective function or decision alternatives provided *(i)* the analyst treats the choice of forecast as the decision, and *(ii)* the analyst anticipates that her forecast will be evaluated by an appropriate scoring rule, and desires to maximize her expected score. A scoring rule is a function S that assigns a score $S(y, a)$ to each combination of forecast report a (usually one or more point estimates or distributions) and outcome y of Y. Scoring rules have been used on an ex-post basis to evaluate the quality of forecasts (Gneiting and Raftery, 2007).

By *appropriate* scoring rule, we mean a scoring rule whose maximizing report is the true characteristic of Y — be it mean, median, distribution or whatever — that the analyst desires to report (or if there are multiple optimal reports, the desired true characteristic should be one of them). This is another way of stating that the scoring rule should be a *proper* scoring rule.[2] Note that we do not ask the analyst to actually carry out this maximization. The analyst has a mathematical model that specifies the joint distribution of $Y, X_1, \ldots, X_n$, and the analyst can use it to compute any desired correct report about Y. Rather, by choosing a proper scoring rule, it is *as if* the analyst's computed report has been obtained by maximizing the expected score. The scoring rule would not actually be used to obtain an optimal report. The analyst would, however, use the scoring rule to compute the value of information of collections X of model inputs, and this value would serve as a measure of sensitivity of her forecast report to X.

Strategies *(i)*-*(ii)* have been suggested before in the context of Bayesian inference (e.g., Bernardo and Smith (1994)). What is new here is our suggestion that information value under S could be used as a sensitivity measure. The analyst would characterize the sensitivity of her forecast to model inputs X as the calculated information value of X under S — the greatest expected score she would sacrifice to allow her forecast a to depend on the hypothetically revealed value of X.

Perhaps surprisingly, a number of existing sensitivity measures are already information value under an appropriate scoring rule — for instance, the variance measure is information value under the quadratic scoring rule, as we are to see. In fact, as we will show, information value under any scoring rule qualifies as a probabilistic sensitivity measure under the common rationale definition of Borgonovo et al. (2016). Of equal importance, the analyst's choice of sensitivity measure now reduces to the choice of an appropriate scoring rule. The properness restriction — that the scoring rule, when (hypothetically) maximized, must produce the analyst's desired report — can be used to eliminate a number of candidate sensitivity measures from consideration. For example, using variance as a sensitivity measure would be inappropriate when reporting a median, because the

[2] Properness is usually defended based on an *ex ante* use of a scoring rule to motivate an expert to truthfully report his opinion. This perspective is irrelevant to our purpose here, in which we are merely supposing the *ex-post* use of a scoring rule to evaluate forecast quality.

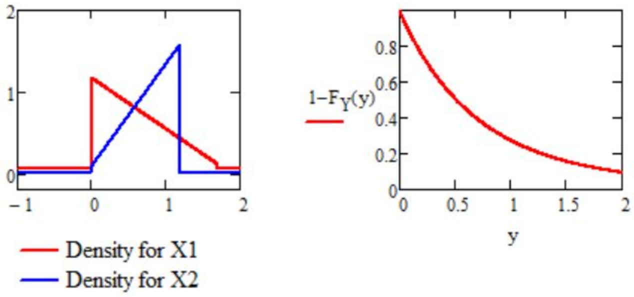

Figure 1: Densities for the contributing rates X_1, X_2 and the resulting complementary cumulative distribution $1 - F_Y(y)$ for the forecast Y. The rate X_1 has a trapezoidal distribution on $[0, 1.7]$ with mean 0.595, and X_2 has a trapezoidal distribution on $[0, 1.2]$ with mean 0.78.

variance measure is information value under quadratic scoring, under which the optimal report is the true mean.

For a hypothetical simplified example, suppose Y is the time until a critical event and it is exponentially distributed with rate parameter $\lambda = \lambda_0 + \lambda_1 + \lambda_2$, where λ_0 is the base rate, and $\lambda_1 = X_1$, $\lambda_2 = X_2$ are independent rate additions from two possible contributing factors. From data, the base rate estimate is $\lambda_0 = 0.05$, but the contributing rates X_1, X_2 are uncertain. Experts assess the trapezoidal distributions shown in Figure 1. The analyst wishes to report a distribution for Y, and based on this information, she derives the complementary cumulative distribution $1 - F_Y(y)$ through the above discussed model (Figure 1). Again, this is purely a calculation based on the conditional exponential distribution of Y given X_1, X_2 and their marginal trapezoidal distributions: It is not based on maximizing any scoring rule.

The analyst is concerned about sensitivity, and calculates the variance-based measure $\mathbb{V}\{\mathbb{E}[Y|X_i]\}$ as a measure of sensitivity to X_i, $i = 1, 2$. As shown in Table 1, this measure is 19.1% higher for X_2 than for X_1. However, as just mentioned, variance-based sensitivity is information value under quadratic scoring, under which the optimal report is the true mean. The analyst is reporting a distribution. If the analyst anticipates that her distribution report is to be evaluated using a scoring rule, a natural choice would be the Continuous Ranked Probability Score (CRPS), a proper scoring rule that we discuss below, under which it is optimal to report the true distribution of Figure 1. Information values for X_1, X_2 under CRPS scoring are also shown in Table 1. Here we see that sensitivity to X_2 is only 88.5% of sensitivity to X_1, contrary to the impression given by the variance-based sensitivity measure. The more appropriate choice of a CRPS scoring rule to compute information values gives, we argue, a truer picture of sensitivity to X_1, X_2 when a distribution is reported. The analyst could have realized without any computation that the properness restriction would rule out the variance-based measure, because it is information value under the quadratic scoring rule, whose optimal report is the mean of Y, not the distribution of Y that she desires to report.

Table 1: Sensitivity Measures for the Example

Sensitivity Measure	Sensitivity to X_1	Sensitivity to X_2	Ratio
Variance-based (quadratic scoring)	0.0854	0.1018	119.1%
Distribution-based (CRPS scoring)	0.0186	0.0164	88.5%

This paper makes several contributions to the literature. The first, as we have noted, is to suggest a means for analysts to choose from among the plethora of possible sensitivity measures for a forecast by relating potential measures to underlying proper scoring rules, and assessing the appropriateness of such rules for the desired forecast (§6). Towards this end, we show, as already mentioned, that information value does qualify as a sensitivity measure (§3.1), and that several popular sensitivity measures are already information value under known scoring rules (§4.1, §5.1). We also obtain new sensitivity measures for reporting density and cumulative distribution functions (§4.2, §5.1, §5.2) and in particular show that Szekely's energy statistic is information value under the CRPS score (§5.2).

Second, we show that other popular measures which are not information value under a scoring rule, are nevertheless information value under a utility function that is proper in the same sense as a scoring rule, but is not itself a scoring rule (§3.2). Third, we supply necessary conditions for a sensitivity measure to be information value under a given scoring rule (§3.3). Fourth, we address nullity implies independence, a desirable property of a sensitivity measure, and show that sensitivity measures based on strictly proper scoring rules achieve it for distribution reports (§3.1). We illustrate the findings through an application in the context of nuclear waste management decisions aided by a benchmark quantitative model developed by the OECD (§7).

2 Review and Basic Concepts

2.1 Related Literature

This paper connects three literature streams: the literature on information value, the literature on sensitivity analysis, and the literature on forecasting and scoring rules. Each of these research streams is vast in itself and a comprehensive review is forcedly out of reach. Therefore, we lay out a synthetic overview for positioning our work with respect to the extant literature.

The notion of information value has been introduced in Howard (1966). Intuitively, information value is the amount of money that a decision maker is willing to pay for gathering additional information about a given uncertainty. The properties of information value have been extensively studied, with initial attention devoted to the relationships among information value and its determinants (LaValle, 1968). Hilton (1981) summarizes the findings reporting that no definitive conclusion on these relationships can usually be drawn. Gilboa and Lehrer (1991) provide an axiomatization of information value and identify the features that make a set function an information value function. The work

Table 2: Examples of probabilistic sensitivity measures encompassed by the rationale in eq. (1). In terms of notation, μ represents the mean, $\mathbb{V}$ the variance, f the density, F the cumulative distribution function of random variable Y or distribution $\mathbb{P}$, possibly conditional on X.

Name		Definition	Inner Operator
Variance-based	η_X	$\mathbb{E}\{\mathbb{V}[Y] - \mathbb{V}[Y\|X]\}$	$(\mu_{\mathbb{P}} - \mu_{\mathbb{Q}})^2$
δ-importance	δ_X	$\frac{1}{2}\mathbb{E}[\int_{\mathbb{R}} \left\| f_Y(y) - f_{Y\|X}(y)\right\| dy]$	$\frac{1}{2}\int_{\mathbb{R}} \|f_{\mathbb{P}}(y) - f_{\mathbb{Q}}(y)\| dy$
Kullback-Leibler	θ_X	$\mathbb{E}[\int_{\mathbb{R}} f_{Y\|X}(y) \log \frac{f_{Y\|X}(y)}{f_Y(y)} dy]$	$\int_{\mathbb{R}} f_{\mathbb{Q}}(y) \left(\log f_{\mathbb{Q}}(y) - \log f_{\mathbb{P}}(y)\right) dy$
Beta-Kuiper	β_X^{Ku}	$\mathbb{E}[\sup_{y\in\mathbb{R}}\{F_Y(y) - F_{Y\|X}(y)\}$ $+ \sup_{y\in\mathbb{R}}\{F_{Y\|X}(y) - F_Y(y)\}]$	$\sup_{y\in\mathbb{R}}\{F_{\mathbb{P}}(y) - F_{\mathbb{Q}}(y)\}$ $+ \sup_{y\in\mathbb{R}}\{F_{\mathbb{Q}}(y) - F_{\mathbb{P}}(y)\}$

of Gilboa and Lehrer provides the starting point for our discussion about whether a probabilistic sensitivity measure can be interpreted as information value. Hazen and Sounderpandian (1999) compare alternative formulations and interpretations of information value, distinguishing willingness to buy from willingness to sell information and characterizing expected utility increase. Pflug (2006) studies the link between information value and risk measures, showing that information value expressed as expected utility increase satisfies the axioms of coherent risk measures of Artzner et al. (1999). Bratvold et al. (2009) and Keisler et al. (2014) illustrate information value applications in a variety of decision problems. Among these, of interest to this work is the application of information value as a probabilistic sensitivity measure. This suggestion comes from the works of Felli and Hazen (1998, 1999), who apply information value in the context of medical decision making. Oakley (2009) extends the intuition to the sensitivity analysis complex quantitative models. This line of research continues in Strong and Oakley (2013) and Strong et al. (2014) who obtain results that improve the efficiency in the estimation of information value as a sensitivity measure.

More in general, the works of Rabitz and Alis (1999); Saltelli et al. (2000); Saltelli and Tarantola (2002) and Oakley and O'Hagan (2004) have contributed in establishing sensitivity analysis as an integral part of modelling and, over the years, a variety of probabilistic sensitivity measures has been introduced. Table 2 reports a sample of the available sensitivity measures, comprising variance-based sensitivity measures (Saltelli and Tarantola, 2002), distribution-based measures using the L^1-norm (Borgonovo, 2007), the Kullback-Leibler divergence between densities (Critchfield and Willard, 1986) and the Kuiper distance between cumulative distribution functions (Baucells and Borgonovo, 2013). Recent works study sensitivity measures based on the family of f-divergences (Rahman, 2016), on a transformation invariant version of the Cramér-von Mises distance (Gamboa et al., 2015) and on quantiles. In fact, quantile-based sensitivity measures are a topical research subject – (Fort et al., 2014; Browne et al., 2017).

Due to space limitations, we cannot enter into a detailed description of all methods, but we refer to the recent reviews of Ferretti et al. (2016); Borgonovo and Plischke (2016) and to the handbook of Ghanem et al. (2016) for additional details. However, it is relevant to note that probabilistic sensitivity measures have been studied mostly from a computational viewpoint. This was in fact required to make them available in realistic applications (Saltelli et al., 2008). They remain much less understood from a decision-analytic viewpoint. Research questions such as whether probabilistic sensitivity measures are information value, or whether a null value of a sensitivity measure reassures the analyst that the report is independent of the exogenous variable have not been systematically addressed to date.

In our analysis, we find guidance from the following facts appearing in the literature. Bernardo and Smith (1994) observe that variance reduction can be reinterpreted as information value under a quadratic scoring rule and the Kullback-Leibler divergence as information value under a logarithmic score on densities. We then argue that scoring rules can be the missing link between information value and probabilistic sensitivity measures. Scoring rules have been introduced and have become over the years an essential a tool for guiding and assessing forecasts. The works of Gneiting and Raftery (2007), Winkler and Jose (2011), and, more recently, Gneiting and Katzfuss (2014) provide comprehensive reviews on this topic. However, the connection between scoring rules and probability sensitivity analysis has not been fully examined in the literature and is one of the purposes of this work.

2.2 Basic Definitions and Relevant Properties

To support the decision-making process at hand, the analyst creates or relies on a model that forecasts the uncertain quantity of interest Y whose value depends (perhaps probabilistically) on a set of uncertain exogenous variables $\mathbf{X} = (X_1, X_2, ..., X_n)$. Let $\mathbb{P}_Y$ denote the distribution of Y. The analyst has the possibility of gathering information about one or more exogenous variables $X \subset \mathbf{X}$. After information has been gathered, $\mathbb{P}_{Y|X}$ denotes the conditional distribution of Y. In preparation for the definition below, note that $\mathbb{P}_{Y|X}$ is absolutely continuous with respect to $\mathbb{P}_Y$. In the remainder, f_Y, $f_{Y|X}$, F_Y, $F_{Y|X}$, μ_Y and $\mu_{Y|X}$ are respectively the density, the cumulative distribution functions and the mean values of $\mathbb{P}_Y$ and $\mathbb{P}_{Y|X}$.

Let $\zeta(\mathbb{P}, \mathbb{Q})$ be a generic operator between probability measures $\mathbb{P}$ and $\mathbb{Q}$.

Definition 1. We call an operator $\zeta(\cdot, \cdot)$ defined over pairs $\mathbb{P}, \mathbb{Q}$ with $\mathbb{Q}$ absolutely continuous with respect to $\mathbb{P}$ an *inner operator*, if $\zeta(\mathbb{P}, \mathbb{Q}) \geq 0$ for all such $\mathbb{P}, \mathbb{Q}$ and also $\zeta(\mathbb{P}, \mathbb{P}) = 0$ for all $\mathbb{P}$. We say that ξ_X is the *probabilistic sensitivity measure with inner operator* ζ if for uncertain input parameters $X \subset \mathbf{X}$,

$$\xi_X = \mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X})]. \tag{1}$$

Even though Y is a random variable, the distribution $\mathbb{P}_Y$ is a non-random quantity, whereas the conditional distribution $\mathbb{P}_{Y|X}$ depends on the random variable X, and is therefore random. So the expectation $\mathbb{E}$ in this expression is over the random variable X. We continue this notational convention through this paper, although in places we use the notation $\mathbb{E}_X$ to emphasize this point.

Definition 1 extends to models with stochastic output the common rationale of global sensitivity measure of Borgonovo et al. (2016) and encompasses several probabilistic sensitivity measures, a sample of which are reported in Table 2. To illustrate, let us consider variance-based sensitivity measures (first entry in Table 2). The inner operator is

$$\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X}) = \mathbb{V}[Y] - \mathbb{V}[Y|X] = (\mu_Y - \mu_{Y|X})^2,$$

and its expectation leads to the well-known expression[3]

$$\eta_X = \mathbb{E}\left\{\mathbb{V}[Y] - \mathbb{V}[Y|X]\right\}. \tag{2}$$

One important property of a probabilistic sensitivity measure is whether a null value of a sensitivity measure reassures the analyst that Y is independent of X.

Definition 2. A probabilistic sensitivity measure ξ_X possesses the *nullity-implies-independence* property if $\xi_X = 0$ implies that X and Y are probabilistically independent.

Historically, nullity-implies-independence is a property characterizing measures of statistical dependence since the seminal work of Rényi (1959). However, this property has a relevant decision-analytic implication. Only when nullity-implies-independence holds is the analyst reassured that a null value of the sensitivity measure guarantees that Y is independent of X. When nullity-implies independence fails, an indication of zero sensitivity can be misleading, as potentially X could still influence Y. An example of such effect in a business context is offered by Baucells and Borgonovo (2013) in which Y is an investment net present value (NPV) and X is a set of cash flows.
Monotonic transformation invariance has emerged as a convenient property in estimation. When the model output Y spans several orders of magnitudes, accurate estimation of a probabilistic sensitivity measure may require a long computation and become out of reach. This problem is often overcome using a monotonic transformation of the output, such as a logarithmic transformation. Then, analysts face the problem that *the log-based uncertainty importance calculations do not readily translate back to a linear scale* (Iman and Hora, 1990, p. 402). However, if the probabilistic sensitivity measure is transformation invariant, results on the transformed and on the original scales coincide. This eliminates the interpretation problem, while allowing analysts to fully exploit the accelerated numerical convergence — see (Borgonovo et al., 2014) among others.

3 Probabilistic Sensitivity Measures and Information Value

This section is divided into three parts. After showing that information value under a scoring rule S is a probabilistic sensitivity measure according to Definition 1, we derive the conditions under which a probabilistic sensitivity measure can be interpreted as information value under a proper utility function. We also examine the conditions that are necessary for a sensitivity measure to be information value under a given scoring rule.

[3] When X is an individual variable, the probabilistic sensitivity measure in (2), η_X, is the well known first order variance-based sensitivity measure (Saltelli and Tarantola, 2002), and coincides with Pearson's correlation ratio Pearson (1905). Usually, one considers the normalized version dividing the right hand side in (2) by $\mathbb{V}[Y]$.

3.1 Information Value for Scoring Rules

A scoring rule S assigns a score $S(y, a)$ to a forecast report $a \in \mathcal{A}$ when the outcome is $y \in \mathbb{R}$. The quality of a report a can be evaluated by the expected score $\mathbb{E}[S(Y, a)]$. The set $\mathcal{A}$ of possible reports could be a subset of the reals (for point estimates such as the mean or median), a subset of n-dimensional Euclidean space (for multiple quantile reports), or a space of probability measures (for density or distribution reporting). We let the context dictate what set $\mathcal{A}$ is appropriate.

For a proper scoring rule S, we define information value in the standard way as

$$\epsilon_X^S = \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \mathbb{E}[S(Y, a)|X] \right] - \max_{a \in \mathcal{A}} \mathbb{E}[S(Y, a)] \tag{3}$$

representing the greatest score improvement one would achieve by learning the value of X.[4] Here we assume the indicated maxima exist. Let $a_{\mathbb{P}}^S$ denote the set of reports a that maximize the expected score[5] $\mathbb{E}_{\mathbb{P}}[S(Y, a)]$, where the notation $\mathbb{E}_{\mathbb{P}}$ denotes that the expectation is taken with Y having distribution $\mathbb{P}$. Then we can write

$$\epsilon_X^S = \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \mathbb{E}\left[S(Y, a) - S(Y, a_{\mathbb{P}}^S) \middle| X \right] \right] = \mathbb{E}_X \left[\max_{a \in \mathcal{A}} \mathbb{E}_{\mathbb{Q}} \left[S(Y, a) - S(Y, a_{\mathbb{P}}^S) \right] \right],$$

where $\mathbb{Q} = \mathbb{P}_{Y|X}$. The quantity inside the last expectation is nonnegative, and equals zero when $\mathbb{P} = \mathbb{Q}$. It is therefore an inner operator by Definition 1. This demonstrates the following result.

Proposition 1. *Information value under a scoring rule ϵ_X^S is a probabilistic sensitivity measure* (1) *with inner operator*

$$\zeta^S(\mathbb{P}, \mathbb{Q}) = \max_{a \in \mathcal{A}} \mathbb{E}_{\mathbb{Q}} \left[S(Y, a) - S(Y, a_{\mathbb{P}}^S) \right]. \tag{4}$$

In fact, by writing $\zeta^S(\mathbb{P}, \mathbb{Q}) = \mathbb{E}_{\mathbb{Q}} \left[S(Y, a_{\mathbb{Q}}^S) - S(Y, a_{\mathbb{P}}^S) \right]$, it follows that ζ^S is the so-called *divergence* function for scoring rule S (Dawid, 2007).

It is easy to see that for X probabilistically independent of Y any probabilistic sensitivity measure ξ_X, including ϵ_X^S, is null. In fact, when Y is independent of X, $\mathbb{P}_{Y|X} = \mathbb{P}_Y$ and $\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X}) = \zeta(\mathbb{P}_Y, \mathbb{P}_Y) = 0$ for all values of X. However, as it is well known, the converse is not true, i.e., a null information value does not guarantee that Y is independent of X. For this, however, the following holds (see Appendix B for the proof).

Proposition 2. *Consider a strictly proper scoring rule $S(y, a)$ in which the report a is a probability distribution, and let ϵ_X^S be the information value of X under scoring rule S. Then $\epsilon_X^S = 0$ if and only if Y is independent of X.*

[4] This definition of information value as score improvement may differ from the definition of information value in monetary terms, and is only equal if S is a monetary amount and there is risk neutrality or constant risk attitude. See Hazen and Sounderpandian (1999) for a discussion.

[5] We write the maximum expected score as $\mathbb{E}_{\mathbb{P}} \left[S(Y, a_{\mathbb{P}}^S) \right]$ even though $a_{\mathbb{P}}^S$ may be a *collection* of reports. $\mathbb{E}_{\mathbb{P}} \left[S(Y, a_{\mathbb{P}}^S) \right]$ is defined to be the common value $\mathbb{E}_{\mathbb{P}}[S(Y, a)]$ for $a \in a_{\mathbb{P}}^S$.

Therefore, if the analyst is reporting the entire distribution and anticipating evaluation under a strictly proper scoring rule S, then she is reassured that a null information value under S implies that Y is independent of X. Thus, one way to guarantee the equivalence of nullity and independence under a strictly proper scoring rule is to report the entire distribution. Moreover, the necessary and sufficient condition in Proposition 2 makes a probabilistic sensitivity measure ξ_X comply with the fifth requirement of Rényi (1959)'s axioms on measures of statistical dependence.

3.2 Probabilistic Sensitivity Measures as Information Value

Our suggestion in this paper is that in problems in which a key variable Y is to be predicted through a computer program based on input parameters/exogenous variables $X_1, ..., X_n$, the most appropriate sensitivity measures are measures ξ_X that arise as information value ϵ_X^S from a scoring rule $S(Y, a)$ in the manner indicated in Proposition 1. In other words, ξ_X should be the information value of X derived from the family of reporting problems (one for each event B)

$$\begin{array}{cc} \text{maximize} & \mathbb{E}[S(Y, a_{\mathbb{P}})|B] \\ \mathbb{P} \in \mathcal{P} & \end{array} \tag{5}$$

where $a_{\mathbb{P}}$ is some report about distribution $\mathbb{P}$ (possibly the full distribution, but possibly something less, such as the mean or a quantile), and $\mathcal{P}$ is a suitably broad set of distributions. Here B is some event involving $X_1, ..., X_n$. Technically, the last statement means that there is an underlying σ-algebra $\mathcal{B}$ over a probability space Ω with respect to which $X_1, ..., X_n$ and X are measurable[6], and that B is a non-null member of $\mathcal{B}$.
A corollary of this point of view is that putative sensitivity measures ξ_X that do *not* arise as information value under some scoring rule should be avoided. One way to rule out a putative measure ξ_X being information value under a particular scoring rule S is provided by Proposition 3 in the next section. But how is one to know that a sensitivity measure ξ_X can *never* be information value under any possible scoring rule S? Currently, this remains an open research question.
The difficulty of deriving a general negative result of this type is emphasized by the following fact: There is a broad class of sensitivity measures ξ_X which are in fact information value under some utility function that is proper in the same sense as a scoring rule, but which may or may not actually be a scoring rule

Comment 1. Maybe we can say that it is model dependent? The Editor is sensitive to how we state sentences about scoring rules

. We prove this result below and in Appendix B, and follow with a discussion of why such sensitivity measures may still be less desirable.
In contrast with (5), consider a reporting problem in which (i) the report $a_{\mathbb{P}}$ is the full distribution $\mathbb{P}$; (ii) instead of a scoring rule $S(Y, a_{\mathbb{P}})$, we allow a general $\mathcal{B}$-measurable utility function $U(\omega, \mathbb{P})$ for $\omega \in \Omega$; and (iii) the set $\mathcal{P}$ of possible reports is replaced by

[6] $\mathcal{B}$ may be only a sub-algebra of the full σ-algebra defining the probability space Ω.

the set of conditional distributions

$$\mathcal{P}_{\mathcal{B}} = \left\{ \mathbb{P}_{Y|A} \middle| A \in \mathcal{B}, A \neq \emptyset \right\},$$

where $\mathbb{P}_{Y|A}(dy) = \mathbb{P}(Y \in dy|A)$. In other words, we consider the family of reporting problems

$$\begin{array}{ll} \text{maximize} & \mathbb{E}[U(\mathbb{P})|B]. \\ \mathbb{P} \in \mathcal{P}_{\mathcal{B}} & \end{array} \tag{6}$$

where, as is conventional notation, $U(\mathbb{P})$ is the random variable that takes on value $U(\omega, \mathbb{P})$ should outcome $\omega \in \Omega$ occur.

In direct analogy to the notion of proper scoring rule, say that utility function U is *proper* if an optimal solution to this problem is $\mathbb{P} = \mathbb{P}_{Y|B}$, that is, for all non-null $A, B \in \mathcal{B}$,

$$\mathbb{E}[U(\mathbb{P}_{Y|A})|B] \leq \mathbb{E}[U(\mathbb{P}_{Y|B})|B].$$

Let $\zeta_B = \zeta(\mathbb{P}_Y, \mathbb{P}_{Y|B})$, and say that utility function U is *consistent* with ζ if there is a constant u_0 such that for all $B \in \mathcal{B}$

$$\mathbb{E}[U(\mathbb{P}_{Y|B})|B] = \zeta_B + u_0. \tag{7}$$

Whether there exist proper utility functions U consistent with ζ is a question that we treat shortly. But if such U do exist, then it follows that the information value of X under U must be $\xi_X = \mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X})]$. This can be seen as follows[7]. Consider first the reporting problem (6) given $B = \Omega$. Because U is proper, the optimal report is $\mathbb{P} = \mathbb{P}_{Y|\Omega} = \mathbb{P}_Y$, and because $\zeta_\Omega = \zeta(\mathbb{P}_Y, \mathbb{P}_Y) = 0$, and U is consistent with ξ, we have $\mathbb{E}[U(\mathbb{P}_Y)] = u_0$. Second, consider the reporting problem (6) given $B = \{X = x\} \in \mathcal{B} \setminus \{\emptyset\}$. Because U is proper, the optimal report is $\mathbb{P}_{Y|\{X=x\}}$ with conditional expected utility $\mathbb{E}[U(\mathbb{P}_{Y|\{X=x\}})|X = x] = \zeta_{\{X=x\}} + u_0 = \zeta(\mathbb{P}_Y, \mathbb{P}_{Y|\{X=x\}}) + u_0$. The expected utility in anticipation of learning X is therefore $\mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X})] + u_0 = \xi_X + u_0$. Information value is the excess of this over the unconditional expected utility u_0, which excess is ξ_X. Therefore we have shown that if there is a proper utility function U consistent with ζ, then the information value under U of any $\mathcal{B}$-measurable X is ξ_X.

Concerning the existence of such utility functions U, we have the following result.

Theorem 1. *Consider the case in which the σ-algebra $\mathcal{B}$ contains finitely many sets B. Suppose the inner operator $\zeta(\mathbb{P}, \mathbb{Q})$ of ξ is convex in $\mathbb{Q}$. Then*

1. *there is a proper utility function U consistent with ζ, and consequently*
2. *for any such U, and any $\mathcal{B}$-measurable X, the value of information X in the family (6) of reporting problems is equal to the sensitivity measure $\xi_X = \mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X})]$.*

We suspect that the restriction that the σ-algebra $\mathcal{B}$ contains finitely many sets B is not crucial to this result, because the infinite-$\mathcal{B}$ case is in some sense the limit of the finite-$\mathcal{B}$ case. Unfortunately, at this time we do not have a proof for $\mathcal{B}$ infinite[8].

[7] See Gilboa and Lehrer (1991) for the origin of this argument.

[8] Note because $\mathcal{B}$ may be only a sub-algebra of the full σ-algebra defining the probability space Ω, the finite-$\mathcal{B}$ restriction does *not* restrict $X_1, ..., X_n, Y$ to be discrete random variables.

The inner operators of the sensitivity measures η_X, θ_X, δ_X and β_X^{Ku} (Table 2) are all convex in their second argument $\mathbb{Q}$, as we show in Appendix A. So we have the following.

Corollary 1. *Under the conditions of Theorem 1, the sensitivity measures η_X, θ_X, δ_X and β_X^{Ku} are information value under the corresponding reporting problem (6).*

As we show in Section 4, the sensitivity measure η_X is also information value in problem (5) under a quadratic scoring rule. Moreover, θ_X is information value under a log scoring rule, as we note in Section 5.1. However, we are unaware of any scoring rules under which δ_X and β_X^{Ku} are always information value for all Y, X, and conjecture there are none. These sensitivity measures are, however information value in problem (6) under some proper utility function, as the corollary states.

Theorem 1 unfortunately guarantees only the existence of a proper utility function $U(\omega, \mathbb{P})$ consistent with ξ, with no reassurance that U derives from a scoring rule $S(Y, \mathbb{P})$ that depends on the key quantity Y. In fact, the utility function derived in the proof of Theorem 1 has no apparent closed form. Not only that, but the utility function U *depends on the model*, that is, on the joint distribution of $Y, X_1, \ldots, X_n$, through the quantities $\zeta_B = \zeta(\mathbb{P}_Y, \mathbb{P}_{Y|B})$ that appear as inputs into the consistency conditions (7). In contrast, a scoring rule S is model independent. Sensitivity measures such as δ_X and β_X^{Ku} may have computational or other desirable properties. Theorem 1 then provides the additional reassurance that these measures do behave as information value under some U. However, the reassurance is limited because U is model dependent, and not necessarily equal to any model-independent scoring rule S.

3.3 Relationship Between Scoring Rules and Inner Operators

Previous literature has shown that a probabilistic sensitivity measure is uniquely determined by its inner operator. However the converse is not true. If any linear real-valued function of $\mathbb{P} - \mathbb{Q}$ is added to an inner operator $\zeta(\mathbb{P}, \mathbb{Q})$, the resulting inner operator generates the same sensitivity measure. For example the inner operator

$$\zeta'(\mathbb{P}, \mathbb{Q}) = (\mu_{\mathbb{P}} - \mu_{\mathbb{Q}} + k)^2 - k^2, \tag{8}$$

differs from the inner operator for η_X in Table 2 by a multiple of $(\mu_{\mathbb{P}} - \mu_{\mathbb{P}})$, and also generates the probabilistic sensitivity measure η_X. And, letting $z^+ = \frac{1}{2}(z + |z|)$ denote the positive part of z, the inner operator

$$\zeta^+(\mathbb{P}, \mathbb{Q}) = \int_{\mathbb{R}} (f_{\mathbb{P}}(y) - f_{\mathbb{Q}}(y))^+ \, dy, \tag{9}$$

differs from the inner operator of δ_X in Table 2 by half the integral of $f_{\mathbb{P}}(y) - f_{\mathbb{Q}}(y)$, but it still yields δ_X as sensitivity measure — see Appendix B for details.

Suppose sensitivity measure ξ_X has inner operator ζ and is always equal to the information value of X under some scoring rule $S(y, a)$, that is $\xi_X = \epsilon_X^S$ for all X. Because a sensitivity measure does not determine its inner operator, it does not follow that $\zeta = \zeta^S$ uniquely. If not, then what kind of relationship must exist between ζ and S? The following result provides a partial answer.

Say that an inner operator ζ *distinguishes between* two probability measures $\mathbb{P}$, $\mathbb{Q}$ if either $\zeta(\mathbb{P},\mathbb{Q}) > 0$ or $\zeta(\mathbb{Q},\mathbb{P}) > 0$. Analogously, say that a scoring rule $S(y,a)$ distinguishes between $\mathbb{P}$, $\mathbb{Q}$ if optimal action $a_{\mathbb{P}}$ when Y has distribution $\mathbb{P}$ differs from optimal action $a_{\mathbb{Q}}$ when Y has distribution $\mathbb{Q}$. Say that a sensitivity measure ξ_X is information value under scoring rule S if, for all X, the information value of X under S is equal to ξ_X. Then we have the following result.

Proposition 3. *Suppose the map $\alpha \mapsto \zeta\left((1-\alpha)\mathbb{Q}_0 + \alpha\mathbb{Q}_1, \mathbb{Q}_0\right)$ is a continuous function over $\alpha \in [0,1]$ for all distributions $\mathbb{Q}_1, \mathbb{Q}_0$. A necessary condition for the sensitivity measure ξ_X with inner operator ζ to be information value under scoring rule $S(y,a)$ is that for all $\mathbb{P}$, $\mathbb{Q}$, the scoring rule S distinguishes between $\mathbb{P}$, $\mathbb{Q}$ whenever the inner operator ζ does.*

As a simple example consider the sensitivity measures δ_X and β_X^{Ku}. They fall under Proposition 3 because their inner operators satisfy its continuity requirement. If ζ distinguishes between $\mathbb{P}$, $\mathbb{Q}$, then $\zeta(\mathbb{P},\mathbb{Q}) > 0$, and therefore $F_{\mathbb{P}} \neq F_{\mathbb{Q}}$. But note, for example, that the quadratic scoring rule

$$S^{Quad}(y,a) = -(y-a)^2 \tag{10}$$

has optimal report a^* equal to the mean of Y (see Section 4). Therefore, S^{Quad} does not distinguish two distributions with the same mean. It follows that δ_X and β_X^{Ku} cannot be information value under S^{Quad}, or for that matter, under any scoring rule that reports only summary distribution statistics. What is needed is a scoring rule whose optimal report is a distribution (or from which a distribution can be derived).

A related argument shows that the condition of Proposition 3 cannot be sufficient. Consider any strictly proper scoring rule S whose optimal report is a distribution (e.g. the CRPS score we discuss in Section 5.2). Because it is strictly proper, S must distinguish any two distinct distributions. Therefore S distinguishes any two distributions that ζ does, and this for arbitrary inner operators ζ. If this condition were sufficient, it would follow that an arbitrary sensitivity measure ξ_X is information value under S, a clear impossibility because different ζ can give non-equivalent values of ξ_X.

4 Sensitivity Measures and Point Estimate Reports

4.1 Bregman Scores: Reporting Mean Values

A scoring rule S^B over numerical reports $a \in \mathbb{R}$ is said to be a *Bregman function* if there exists a differentiable strictly convex function $\psi(y)$ over $\mathbb{R}$ such that

$$S^B(y,a) = \psi(a) + \psi'(a)(y-a) - \psi(y). \tag{11}$$

The report a that maximizes $\mathbb{E}\left[S^B(Y,a)\right]$ is μ_Y, the expected value of Y, i.e. $a_Y^{S^B} = \mathbb{E}[Y] = \mu_Y$. It is also the case, under some weak regularity conditions, that any scoring function that is maximized by the mean must be a Bregman function (Savage (1971), Schervish (1989), and Theorem 7 of Gneiting (2011)). Information value ϵ_X^B under the

Bregman score is by Proposition 1 a probabilistic sensitivity measure with inner operator given by (4). The following proposition provides the specifics.

Proposition 4. *The information value of X under a Bregman scoring rule is given by*

$$\epsilon_X^B = \mathbb{E}\psi(\mu_{Y|X}) - \psi(\mu_Y) \tag{12}$$

with inner statistic

$$\zeta^B(\mathbb{P}_Y, \mathbb{P}_{Y|X}) = \psi(\mu_{Y|X}) - (\psi(\mu_Y) + \psi'(\mu_Y)(\mu_{Y|X} - \mu_Y)). \tag{13}$$

Because of the strict convexity of ψ, we have $\epsilon_X^B > \psi(\mathbb{E}\mu_{Y|X}) - \psi(\mu_Y) = \psi(\mu_Y) - \psi(\mu_Y) = 0$ as long as $\mu_{Y|X} \neq \mu_Y$ for some values of X. The only way ϵ_X^B can be zero is when $\mu_{Y|X} = \mu_Y$ for all values of X. It follows that ϵ_X^B need not possess the nullity-implies-independence property, because nullity only entails $\mu_{Y|X} = \mu_Y$ for all values X, which can occur even when X and Y are dependent random variables. The literature offers several examples of models in which $\mu_{Y|X} = \mu_Y$ for all values X, even though Y is dependent on X (see Ishigami and Homma (1990) and Plischke et al. (2013)). It is however true that nullity implies independence for binary random variables, or when the joint distribution of Y, X is multivariate normal. In these settings, when $\mu_{Y|X} = \mu_Y$ for all values X, then Y must be independent of X. The binary case when Y is an indicator for some event E is indeed the most common case to which the quadratic score (discussed next) is applied.

The most popular example of a Bregman scoring function is the quadratic score. Setting $\psi(y) = y^2$, we obtain the quadratic or squared error score in (10) (Gneiting, 2011). Following Proposition 4, information value under this scoring rule is given by

$$\epsilon_X^{Quad} = \mathbb{E}\left[\mu_{Y|X}^2 - \mu_Y^2\right] = \mathbb{V}[\mu_{Y|X}] = \mathbb{V}[Y] - \mathbb{E}[\mathbb{V}[Y|X]], \tag{14}$$

so ϵ_X^{Quad} is equal to the sensitivity measure η_X in Table 2 and is equal to the amount of variance reduction in Y due to information X — see also Bernardo and Smith (1994, p. 300).

4.2 Piecewise Linear Score Functions and Quantile Reports

Consider a situation in which the analyst must report a particular quantile of the distribution of Y. A notable example is in finance, where the so-called value at risk (VaR), has become a popular risk measure (see Gourieroux et al. (2000); Fermanian and Scaillet (2005)). VaR is typically the 95^{th} or 99^{th} quantile of an investment risk profile. Suppose the analyst anticipates her report of the p-quantile will be evaluated by a scoring rule. As Gneiting (2011) generalizing Thomson (1979) shows, the only possible proper rules lie in the family of generalized piece-wise linear loss functions associated with the p-quantile (also known as the check function in quantile regression and the newsvendor function in operations management):

$$S_p^Q(y, a) = k + h \cdot \left(p(t(y) - t(a))^+ + (1 - p)(t(a) - t(y))^+\right), \tag{15}$$

where $p \in (0,1)$, $t : \mathbb{R} \mapsto \mathbb{R}$ is a non-decreasing function, $k \in \mathbb{R}$ and $h \in \mathbb{R}^+$ are constants. The optimal report under $S_p^Q(y,a)$ is the p-quantile[9] of Y (Cervera and Muñoz, 1996; Gneiting and Raftery, 2007; Jose et al., 2009), which we denote by $Q_Y(p)$.
Let $\mathbb{E}^{a,b}[Z] = \int_a^b z dF_Z(z)$ denote the partial expectation of a random variable over interval (a,b).

Proposition 5. *The information value of X when the choice problem is to report the p-quantile is given by:*

$$\epsilon_X^Q = h \cdot \mathbb{E}_X \Big\{ \mathbb{E}^{-\infty, Q_{Y|X}(p)} \left[t(Y)|X \right] + t(Q_{Y|X}(p)) \left(p - F_{Y|X}(Q_{Y|X}(p)) \right) \\ - \mathbb{E}^{-\infty, Q_Y(p)} \left[t(Y) \right] + t(Q_Y(p)) \left(p - F_Y(Q_Y(p)) \right) \Big\}. \tag{16}$$

When Y is a continuous random variable, the quantile information value becomes

$$\epsilon_X^Q = h \cdot \left(\mathbb{E}_X \left\{ \mathbb{E}^{-\infty, Q_{Y|X}(p)} \left[t(Y)|X \right] \right\} - \mathbb{E}^{-\infty, Q_Y(p)}[t(Y)] \right). \tag{17}$$

If we take a linear function $t(Y)$,the previous equation simplifies further into

$$\epsilon_X^Q = h \cdot \left(\mathbb{E}_X \left\{ \mathbb{E}^{-\infty, Q_{Y|X}(p)}[Y|X] \right\} - \mathbb{E}^{-\infty, Q_Y(p)}[Y] \right). \tag{18}$$

Example 1. [Quantiles of Linear Models] In many applications, Y a linear combination of n exogenous X_i's, i.e.,

$$Y = \sum_{i=1}^{n} a_i X_i. \tag{19}$$

Analysts deal frequently with linear models, either because the structure of the problem is linear (e.g., when Y is a portfolio return and the X_i's are the returns of the assets in the portfolio or when Y is the net present value for an investment and the X_i's are the potential cash flows) or because a linear response surface is used to approximate the input-output mapping. If normal distributions are appropriate to characterize beliefs about the X_i's, then the quantile information value can be expressed in closed form as:

$$\epsilon_X^Q = h\varphi(\Phi^{-1}(p)) \left\{ \sigma_Y - \mathbb{E}\left[\sigma_{Y|X} \right] \right\}, \tag{20}$$

where $\varphi(\cdot)$ and Φ are, respectively, the standard normal density and cumulative distribution function and σ_Y and $\sigma_{Y|X}$ are, respectively, the portolio standard deviation and conditional standard deviation — see Appendix B for the analytical expressions and derivation of (20). Thus, the quantile information value of X is proportional to the decrease in standard deviation associated with the additional information provided by the variable X under a normality assumption. Of course, this relationship may not hold for more general distributions.

[9]Formally, the quantile function is defined as $Q_Y(p) = \inf\{y \in \mathbb{R} : p \leq F(y)\}$. However, when Y is absolutely continuous and strictly increasing, then $Q_Y(p) = F_Y^{-1}(p)$.

Proposition 5 has analogs for interval estimates. To illustrate, consider a symmetric $(1-\alpha) \times 100\%$ interval estimate for Y. The lower bound is the $\alpha/2$-quantile and the upper bound is the $1-(\alpha/2)$-quantile. With $a = (a_L, a_U)$, the scoring function then is simply the sum $\mathcal{S}^Q_{\alpha/2}(y, a_L) + \mathcal{S}^Q_{1-(\alpha/2)}(y, a_U)$. Because the overall score is additively separable in a_L and a_U, score maximization can be performed separately for a_L and a_U. If the random variables are absolutely continuous, we can express information value for a confidence interval as

$$\begin{aligned}\epsilon_X^{Interval} &= h\left(\mathbb{E}_X\left\{\mathbb{E}^{-\infty,Q_{Y|X}(\alpha/2)}\left[t(Y)|X\right] + \mathbb{E}^{-\infty,Q_{Y|X}(1-(\alpha/2))}\left[t(Y)|X\right]\right\}\right. \\ &\qquad \left. -\mathbb{E}^{-\infty,Q_Y(\alpha/2)}[t(Y)] - \mathbb{E}^{-\infty,Q_Y(1-(\alpha/2))}[t(Y)]\right).\end{aligned}$$

Because the quantile scores are additive, the interval information value is additive over the two endpoints. To illustrate, let us refer back to Example 1. Interval information value still remains proportional to the decrease in standard deviation $\sigma_Y - \mathbb{E}[\sigma_{Y|X}]$. We note that when we generalize this concept to a set of quantiles for the set of probabilities $\mathbf{p} = (p_1, \ldots, , p_n)$, the scoring function given by

$$\mathcal{S}^Q_{\mathbf{p}}(y, a) = \sum_{p \in \mathbf{p}} \mathcal{S}^Q_p(y, a).$$

remains additive in each component. We can easily show for normal variates that the property of proportionality to the decrease in standard deviation remains intact.

5 Reporting Distributions

Often an analyst is interested in the entire distribution (or density) of Y. One general result that marks a departure between this case and the case of point estimates has been illustrated in Section 3.1. By Proposition 2, all the sensitivity measures derived in this section possess the nullity-implies-independence property.

5.1 Density Forecasts

Suppose the density q_Y of Y is the quantity of interest to the analyst (Y is assumed absolutely continuous in this section). Then the choice set A is represented by some space of probability densities and a possible report is $a = q_Y$. Dawid (2007) generalizes the Bregman function to the case of density function reporting, introducing the generalized Bregman score

$$S(y, q) = \psi'(q(y)) + \int_{\mathbb{R}} [\psi(q(s)) - q(s)\psi'(q(s))]ds. \tag{21}$$

This family of scores is theoretically appealing and contains many well-known examples of existing proper scoring rules. Information value under the generalized score (21) encompasses several probabilistic sensitivity measures. First, letting $\psi(q) = q - q\log q$, we obtain the log scoring rule

$$S(y, q) = \log q(y). \tag{22}$$

Applying the definition of information value and taking the expectation of $\zeta^{\mathcal{D}}(\mathbb{P}_Y, \mathbb{P}_{Y|X})$, we obtain (Bernardo and Smith, 1994) :

$$\epsilon_X^{KL} = \mathbb{E}_X\Big[\int_{\mathbb{R}} f_{Y|X}(y)(\log f_{Y|X}(y) - \log f_Y(y))dy\Big], \tag{23}$$

which is the sensitivity measure θ_X introduced in Table 2, and the inner operator $\zeta^{KL}(\mathbb{P}, \mathbb{Q})$ is the well known Kullback-Leibler divergence between f_Y and $f_{Y|X}$ (Kullback and Leibler, 1951). Bernardo (1979b) discusses how the logarithmic function plays an important role in deriving and reporting reference posteriors in general inferential problems.

Consider then an analyst who selects a power-based Bregman function for values $s \notin \{0, 1\}$

$$\psi^{Power}(q) = \omega q^s, \tag{24}$$

where $\omega = 1$ for $0 < s < 1$ and $\omega = -1$ for other permissible values of s. Substituting (24) into (21), we obtain the power scoring function (Dawid, 2007):

$$S^{Power}(y, q) = \omega\left[s q_Y^{s-1}(y) - (s-1)\int_{\mathbb{R}} q_Y^s(t)dt\right]. \tag{25}$$

We can then determine the corresponding information value for an analyst reporting the density of Y as

$$\epsilon_X^{Power} = \omega \mathbb{E}_X\left[\int_{\mathbb{R}} \left(q_{Y|X}^s(y) - s q_{Y|X}(y) q_Y^{s-1}(y) + (s-1) q_Y^s(y)\right) dy\right]. \tag{26}$$

To illustrate, when $s = 2$, we obtain

$$\epsilon_X^{Power2} = \mathbb{E}_X\left[\int_{\mathbb{R}} \left(q_{Y|X}(y) - q_Y(y)\right)^2 dy\right]. \tag{27}$$

That is, ϵ_X^{Power2} is a probabilistic sensitivity measure whose inner operator is the L^2-norm between densities, $\int_{\mathbb{R}} \left(q_{Y|X}(y) - q_Y(y)\right)^2 dy$.

While not part of the family of generalized Bregman scores, pseudospherical scoring rules are a further well known family of scores (Dawid, 2007). The score function is given by:

$$S^{Spherical}(y, q; s) = \frac{q_Y^{s-1}(y)}{\left(\int_{\mathbb{R}} q_Y^s(t)dt\right)^{1-\frac{1}{s}}}.$$

By applying the definition of sensitivity measure in (1), we obtain the information value:

$$\epsilon_X^{Spherical} = \mathbb{E}_X \int_{\mathbb{R}} \left(\frac{q_{Y|X}^s(y)}{(\int_{\mathbb{R}} q_{Y|X}^s(t)dt)^{1-\frac{1}{s}}} - \frac{q_{Y|X}(y) q_Y^{s-1}(y)}{(\int_{\mathbb{R}} q_Y^s(t)dt)^{1-\frac{1}{s}}}\right) dy. \tag{28}$$

5.2 Distribution Forecasts

Analysts may also be interested in the entire cdf of the key variable Y. A clear example is business planning, where decision makers consider the so-called risk-profile, namely, the cdf of a project net present value. In well-known test cases, (e.g., Genzyme-Geltech case study in Baucells and Borgonovo (2013)), the investment NPV is not an absolutely continuous random variable. In this case, we need to remove the absolute continuity assumption needed to obtain a value-of-information measure stated in the previous section.

Scoring rules defined for distributional forecasts are not as common since elicited distributions which do not have well-defined density functions are often cognitively difficult or uncommon. However, some scoring functions that depend on distribution functions exist and are used. A popular score used in practice is CRPS:

$$S^{CRPS}(y, F) = -\int_{\mathbb{R}} (F(z) - \mathbf{1}\{z \geq y\})^2 dz, \tag{29}$$

where $\mathbf{1}\{z \geq y\}$ is the indicator variable of $z \geq y$. This scoring rule has several interesting properties such as strict properness and sensitivity to distance (see Jose et al. (2009) also for a generalization).[10]

Proposition 6. *Information value for X under CRPS is given by:*

$$\epsilon_X^{CRPS} = \mathbb{E}\left[\int_{\mathbb{R}} \left(F_Y(y) - F_{Y|X}(y)\right)^2 dy\right]. \tag{30}$$

Thus, information value in this case is a probabilistic sensitivity measure based on the Cramér-von Mises divergence (Hoeffding, 1948; Anderson, 1962). The corresponding inner operator

$$\zeta^{CRPS}(\mathbb{P}_Y, \mathbb{P}_{Y|X}) = \int_{\mathbb{R}} \left(F_Y(y) - F_{Y|X}(y)\right)^2 dy \tag{31}$$

is equal to one-half of the energy statistic of Szekely (1989) applied to comparing F_Y with $F_{Y|X}$. Szekely and Rizzo (2013, 2017) highlight that the energy statistic is gaining increasing interest in applied statistics and machine learning as a measure of dependence.

Example 2. [Example 1 Continued] Consider an analyst who is interested in learning which of the asset returns is more informative when the entire distribution of the portfolio return in Example 1 is of interest and scored with CRPS. Because Y is normally distributed, we can write (Gneiting and Raftery, 2007):

$$S^{CRPS}(y; \mu_Y, \sigma_Y) = \sigma_Y \left[\frac{1}{\sqrt{\pi}} - 2\varphi\left(\frac{y-\mu_Y}{\sigma_Y}\right) - \frac{y-\mu_Y}{\sigma_Y}\left(2\Phi\left(\frac{y-\mu_Y}{\sigma_Y}\right) - 1\right)\right]. \tag{32}$$

[10]Sensitivity to distance roughly implies that forecasts that place more weight to states that are closer to the outcome that materializes receive a higher score.

Table 3: Summary of the sensitivity measures analyzed in this work, TI: Transformation-Invariance, NIIP: Nullity-Implies-Independence Property

Sensitivity Measure	Report	Score	TI	NIIP
η_X (14)	Mean	Quadratic (14)	No	No
ϵ_X^Q (16)	Quantile	Piecewise Linear (15)	No	No
δ_X (Table 2)	Density	N/A	Yes	Yes
$\epsilon_X^{KL} = \theta_X$ (23)	Density	Log (22)	Yes	Yes
ϵ_X^{Power} (26)	Density	Power (24)	No	Yes
$\epsilon_X^{Spherical}$ (28)	Density	Pseudospherical (28)	No	Yes
β_X^{Ku} (Table 2)	cdf	N/A	Yes	Yes
ϵ_X^{CRPS} (30)	cdf	CRPS (29)	No	Yes

Then, the information value of asset return X_i is given by:

$$\begin{aligned}\epsilon_{X_i}^{CRPS} &= \int_{\mathbb{R}} S^{CRPS}(y;\mu_Y,\sigma_Y)\nu(y,\mu_Y,\sigma_Y)dy - \\ &\iint_{\mathbb{R}^2} S^{CRPS}(y;\mu_{Y|X_i},\sigma_{Y|X_i})\nu(y-a_ix_i,\mu_{Y|X_i},\sigma_{Y|X_i})\nu(x_i,\mu_{X_i},\sigma_{X_i})dydx_i,\end{aligned}$$

where $\nu(y;\mu;\sigma)$ denotes the normal density with parameters μ and σ. To illustrate, for a portfolio of three standard normally distributed asset returns with relative weights $a_1 = 4/7$, $a_2 = 2/7$ and $a_3 = 1/7$ the information value of X_i expressed as percentage improvement over the expected score is equal to 51% for the first asset, to 10% for second asset and to 2.4% for the third asset. Note that for the linear combination of normal random variables the same expected percentage improvements would be obtained if we were to consider information value for reporting any quantile, because the quantile score is given by (20) and because the CRPS is the weighted average of quantile scores.

6 Choosing the Right Sensitivity Measure

Here we investigate the practical (or managerial) implications of the findings in the previous sections. Consider an analyst choosing the proper sensitivity measure for the application at hand. Table 3 lists eight of the probabilistic sensitivity measures discussed earlier, classified according to the type of report and its corresponding scoring rule (if any).

What sensitivity measure to use? The analyst should first consider the best type of report (mean, quantile(s), or distribution) to produce. This may depend on the analyst's anticipated audience, or there may be a requirement for a specific report type. Should the desired report include some measure of central tendency and/or a prediction interval, then one of the first three sensitivity measures η_X, ϵ_X^Q in Table 3 would be appropriate. Although these are not transformation invariant and do not obey nullity-implies-independence, the analyst may nevertheless support their use as sensitivity measures by noting they are equal to information value should the report quality be evaluated respectively by the quadratic, log, or piecewise-linear scoring rules. Of course, if the analyst reports, say quantiles, there would be no rationale to use η_X as the sensitivity measure, since this is an information value under scoring rules whose optimal report is the mean. It makes no sense to attempt to justify a sensitivity measure as information value under a scoring rule if the optimal report under the rule differs from the analyst's desired report.

Should a *distribution* report be requested or allowed, the analyst can choose from one of the last seven sensitivity measures in the table. All satisfy nullity-implies-independence. Should transformation invariance be desired, there are three possibilities, δ_X, ϵ_X^{KL} and β_X^{Ku} but only the Kullback-Leibler score ϵ_X^{KL} is known to be information value under a scoring rule, a consideration that might justify preference for its use. If transformation invariance is not crucial, then ϵ_X^{KL} as well as three others of these seven measures $\epsilon_X^{Power}, \epsilon_X^{Spherical}$ and ϵ_X^{CRPS} could be justified as information value under a corresponding scoring rule. If instead it is desired to use one of the others such as δ_X and β_X^{Ku}, perhaps for computational convenience, then our Theorem 1 provides partial reassurance that these are also information value, although not under any model-independent scoring rule. And much as we have already noted, there would be no rationale to use any of the first three sensitivity measures from the table, since they are information value under scoring rules whose optimal report is not a distribution.

Although these rationales do not uniquely identify the appropriate sensitivity measure under all reporting circumstances, they do narrow the field and provide reasonable justification for the measure or measures eventually chosen.

7 Application: Radioactive Waste Management

Many developed countries are dealing with the problem of long-term disposal of nuclear waste in deep geological formations. This is a complex decision making problem with consequences that reach far into the future in an intergenerational perspective. The Nuclear Energy Agency (NEA) of the the Organization for Economic Co-Operation and Development (OECD) established the Radioactive Waste Management Committee in 1975. The mandate of the committee ranges from facilitating *the elaboration of waste management strategies that respect societal requirements, from helping in providing common bases to the national regulatory frameworks*[11]. As part of this exercise, mathematical models are developed for predicting flow and transport of radionuclides in actual geologic formations for assessing the safety of deep repository systems of long-lived radioactive waste (NEA, 2001). The validation of these computer codes and the quantification of the relevant uncertainties have been identified as key-elements for ensuring the quality of the assessment process. In 1985 the NEA started the development of a mathematical code that could serve as a benchmark for Member Countries (OECD, 1989). The code, later named LevelE, has since then become a benchmark for Monte Carlo simulation and sensitivity analysis studies in general (Saltelli et al., 2008).

[11]https://www.oecd-nea.org/rwm/

Table 4: Parameters and Respective Distributions for the LevelE Model (Saltelli and Tarantola, 2002).

X	Definition	Distribution	Range	Units
X_1	Containment Time to Failure	Uniform	$[100, 1000]$	yr
X_2	Iodine Leach Rate	Log-Uniform	$[10^{-3}, 10^{-2}]$	$mols/yr$
X_3	Np Chain Leach Rate	Log-Uniform	$[10^{-6}, 10^{-5}]$	$mols/yr$
X_4	Water Velocity Geosphere Layer I	Log-Uniform	$[10^{-3}, 10^{-1}]$	m/yr
X_5	Length Geosphere Layer I	Uniform	$[100, 500]$	m
X_6	Retention Factor Iodine Layer I	Uniform	$[1, 5]$	-
X_7	Retention Factor for NP Layer I	Uniform	$[3, 30]$	-
X_8	Water Velocity Geosphere Layer II	Log-Uniform	$[10^{-2}, 10^{-1}]$	m/yr
X_9	Length of Geosphere Layer II	Uniform	$[10^{-2}, 10^{-1}]$	m
X_{10}	Retention Factor Iodine Layer II	Uniform	$[10^{-2}, 10^{-1}]$	-
X_{11}	Retention Factor for NP Layer II	Uniform	$[10^{-2}, 10^{-1}]$	-
X_{12}	Stream Flow Rate	Log-Uniform	$[10^{-2}, 10^{-1}]$	m^3/yr

From a technical viewpoint, LevelE consists of a series of nested differential equations. The modelled processes comprise radioactive decay, dispersion and advection through the soil of the radionuclide ^{129}Iodine and the chain ^{237}Neptunium$\mapsto$^{233}Uranium$\mapsto$^{229}Thorium — see Saltelli and Tarantola (2002, p. 703) for the detailed equations. The output of the model is the total dose ingested by an individual at time t, calculated as sum of the doses of the four radionuclides coming from the repository. The time spans geological eras from 2×10^4 to 2×10^9 years into the future.

Parametric uncertainty has been addressed in dedicated exercises over the years and has lead to the identification of the uncertain exogenous variables, which have been assigned official distributions that, since then, have become the standard in subsequent numerical experiments on the model. The distributions are listed in Table 4.

We consider the dose D (in Sievert/year) at $t = 300000$ years as output of interest. A Monte Carlo simulation conducted with a quasi-random sample of size $N = 2^{18}$ produces a skewed distribution of D with a 5^{th} quantile estimate of $\hat{Q}^D_{05} = 5.57 \cdot 10^{-36}$, a 95^{th} quantile estimate of $\hat{Q}^D_{95} = 5.03 \cdot 10^{-08}$, a median estimate of $\hat{Q}^D_{50} = 8.96 \cdot 10^{-13}$ and an expected value estimate of $\hat{\mathbb{E}}[D] = 1.76 \cdot 10^{-08}$.

The barplots in Figure 2 display the importance of the uncertain exogenous variables as the forecast report varies. Each panel corresponds to an alternative report or score function. The first three panels refer to the 5^{th}, 50^{th} and 95^{th} quantiles, the fourth panel to the quadratic score, the fifth to the logarithmic score on densities, the eight to CRPS. The sixth and seventh panels report results for the δ and β^{Ku} sensitivity measures.

A visual inspection of these panels suggests that the selection of the most informative variables depends notably on which sensitivity measure and associated forecast report the analyst selects. We give what could be the subjective impression an analyst receives from these panels in Table 5. The fourth column in Table 5 reports the most important variables for each forecast report. The criterion is quantitative that is, for each report, we list the exogenous variables with sensitivity at least 33% of the most important variable.

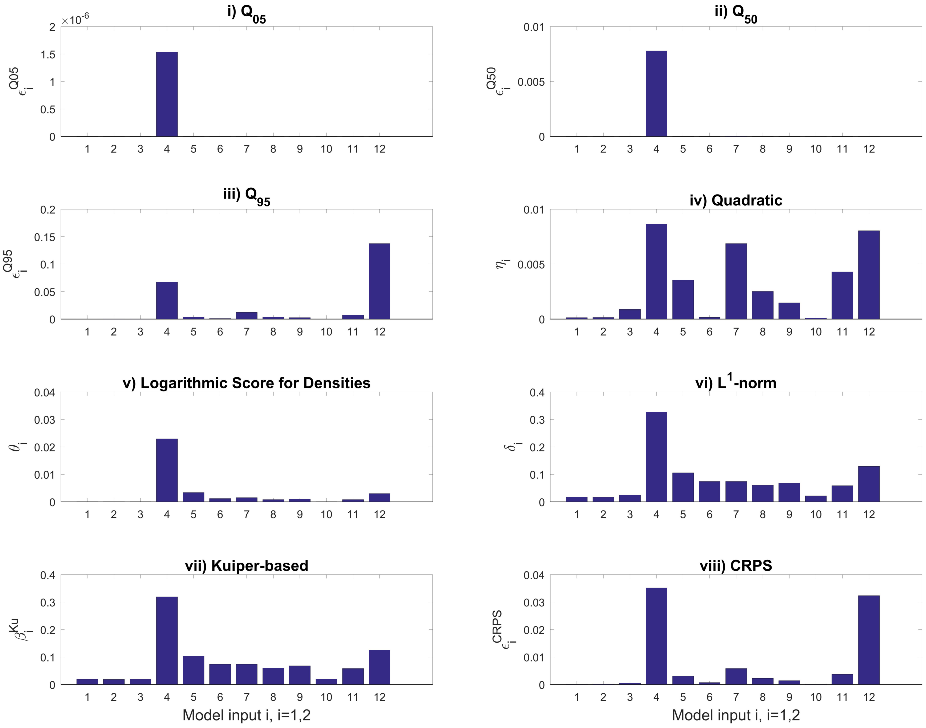

Figure 2: Sensitivity Measures for the Level E dose at $t = 3 \cdot 10^5$.

Table 5: Summary of the most important uncertain exogenous variables for the dose D at $t = 300000$ estimated using the LevelE code, across eight alternative probabilistic sensitivity measures. The symbol $\sim$ denotes difference in sensitivity within 10%, and the symbol $>$ denotes a greater difference.

Report	Scoring Rule	Measure	Important Variables
Fifth Quantile, Q_{05}	Linear	ϵ_X^{05}	X_4
Median, Q_{50}	Linear	ϵ_X^{50}	X_4
95^{th} Quantile, Q_{95}	Linear	ϵ_X^{95}	$X_{12} > X_4$
Mean	Quadratic	$\epsilon_X^{Quad} = \eta_X$	$X_4 \sim X_{12} > X_7 > X_{11} \sim X_5$
PDF	Logarithmic	$\epsilon_X^{KL} = \theta_X$	X_4
PDF	N/A	δ_X	$X_4 > X_{12}$
CDF	N/A	β_X	$X_4 > X_{12}$
CDF	CRPS	ϵ_X^{CRPS}	$X_4 \sim X_{12}$

As the table shows, there is some variability across scoring rules in the variables to which dose D is most sensitive, although X_4 and X_{12} are consistently highly ranked. To return to the point of the prior section, however, note that a danger could arise should an analyst default to a single sensitivity measure without considering the desired report type, and without appreciating the interpretation of sensitivity as information value. For instance, consider an analyst reporting a mean — row 4 in Table 5. If she defaults to logarithmic sensitivity then far too few variables would be of concern compared to what would have been reported under the quadratic score that is consistent with reporting a mean. In the same way, suppose an analyst reporting a density defaults to the quadratic sensitivity measure, as it is commonly used and widely recommended. Then far too many variables would appear to be of concern (comparing rows 4 and 5 in the Table) than if she had simply used logarithmic scoring that is consistent with a density report, or used the δ-importance (row 6) that can be defended as information value under a suitable proper utility function (Section 3.2). The same problem occurs if an analyst is reporting a median or the 5th quantile — rows 1 and 2 in Table 5. These results do confirm the suggestions of Section 6: Randomly picking a sensitivity measure exposes the analyst to the risk of miscommunication about the most important exogenous variables. These results also suggest there is an advantage in using a variety of sensitivity measures. In case the choice of the forecast report is not clear or the analyst (in accordance with the decision maker) does not feel confident enough to rely on a single report, she can use the ensemble of results for communicating insights. To illustrate, for the Dose at $t = 300,000$ the analyst can confidently say that X_4 and X_{12} are the two most important variables. The analyst can also explain this assertion: Water Velocity in Layer I (X_4) plays an important role on the quantiles, on the mean as well as on the entire distribution. Stream Flow Rate (X_{12}) plays a more relevant role than X_4 when the report is the 95th quantile and always ranks second when the report of interest is the entire distribution (either cdf or density). These two uncertain variables are followed by a group comprising X_5, X_7 and X_{11}, with X_{11} ranking fourth only under a quadratic score and thus being overall less relevant than X_5 and X_7.

In either case (if attention is focused on a given report or if a holistic view on the sensitivity measures is adopted), the analyst has a way to provide solid recommendations about which variables are more relevant for further information collection. This has the potential of reducing uncertainty in predictions and consequently of making the decision process better informed. These observations, while illustrated through a case study in a particular sector, are not restricted to the context and are applicable to generic decision problems in which a quantitative model is used to support a decision.

8 Conclusions

This work has established a bridge between three relevant decision analysis topics that have not been simultaneously studied so far. This synthesis yields a variety of results and insights. First, it allows one to better characterize the conditions under which a probabilistic sensitivity measure can be interpreted as information value. Second, it permits us to identify the value-of-information sensitivity measure consistent with a specified scoring rule. Third, it provides a better understanding of various properties of probabilistic sensitivity measures such as nullity-implies-independence. The work introduces several new probabilistic sensitivity measures that retain a value-of-information interpretation, and in particular, sensitivity measures related to the reporting of densities and of the entire cdf. Among others, we prove that Szekely's popular energy statistic is information value consistent with the CRPS score. Some of our results give rise to broad characterizations, such as the fact that well-known sensitivity measures based on metrics such as the L^1-norm and Kuiper metric, while unlikely in our view to be information value under any model-independent scoring rule, are nevertheless information value under some model-specific proper utility function. Our results provide discipline in the selection of sensitivity measures for prediction problems, by giving analysts a structured rationale.

Acknowledgements

We wish to thank Casey Lichtendahl for useful comments and discussions on an earlier version of the manuscript.

Appendix A: Details on the Introductory Example

First, let us write the equations of the trapezoidal densities of X_1 and X_2:

$$f(x,b,\mu) = \begin{cases} (1-\dfrac{x}{b})K_0(b,\mu) + \dfrac{x}{b}K_1(b,\mu) & if \quad 0 \le x \le b \\ 0 & otherwise \end{cases}$$

with $K_0(b,\mu) = \frac{2}{b^2}(2b-3\mu)$, and$K_1(b,\mu) = \frac{2}{b^2}(-b+3\mu)$. The conditional distribution of Y given $X_1 = \lambda_1$ and $X_2 = \lambda_2$ is

$$F_Y(y|\lambda_1,\lambda_2) = e^{-(\lambda_0+\lambda_1+\lambda_2)y}.$$

The unconditional distribution is therefore

$$F_Y(y) = \mathbb{E}_{X_1,X_2}[e^{-(\lambda_0+X_1+X_2)y}].$$

As we assume independence, we obtain

$$F_Y(y) = e^{-\lambda_0 y}L_1(y)L_2(y),$$

where $L_i(y) = \mathbb{E}_{X_i}[e^{-X_i y}]$, $i = 1,2$. The conditional distributions of Y given X_1 or X_2 are

$$F_Y(y|X_1=\lambda_1) = e^{-(\lambda_0+\lambda_1)y}L_2(y) \qquad and \qquad F_Y(y|X_2=\lambda_2) = e^{-(\lambda_0+\lambda_2)y}L_1(y) \ .$$

Let us consider that the analyst chooses a quadratic scoring rule, i.e., a variance based sensitivity measure, as we are to see. Then as we show in Section 4, value of information is given by

$$\varepsilon^{\text{Quad}}_{X_i} = \mathbb{V}_{X_i}[\mathbb{E}[Y|X_i]]. \tag{33}$$

We then need to compute $\mathbb{E}[Y|X_i]$, $i = 1, 2$. This conditional expectation is given for X_1 by

$$\mathbb{E}[Y|X_1 = \lambda_1] = \mathbb{E}_{X_2}[(\lambda_0 + \lambda_1 + X_2)^{-1}]. \tag{34}$$

Let now $K_2(v) = \mathbb{E}_{X_2}[(v + X_2)^{-1}]$. By substituting (34) into (33) one obtains

$$\varepsilon_{X_1}^{\text{Quad}} = \mathbb{E}_{X_1}[K_2(\lambda_0 + X_1)^2] - \mathbb{E}_{X_1}[K_2(\lambda_0 + X_1)]^2.$$

For the second contributing rate X_2 one proceeds analogously.
For $\varepsilon_{X_1}^{\text{CRPS}}$ we obtain:

$$\varepsilon_1^{\text{CRPS}} = \mathbb{E}[\int_0^\infty (F_Y(y) - F_Y(y|X_1))^2 dy$$
$$= \int_0^\infty \mathbb{E}[(F_Y(y) - F_Y(y|X_1))^2]dy = \int_0^\infty \mathbb{E}[\mathbb{V}\{F_Y(y|X_1)\}]dy.$$

The last term of the previous equality is given by

$$\mathbb{V}\{F_Y(y|X_1)\} = \mathbb{E}[F_Y(y|X_1)^2] - F_Y(y)^2 = e^{-2\lambda_0 y}\mathbb{E}[e^{-2X_1 y}]L_2(y)^2 - e^{-2\lambda_0 y}L_2(y)^2 L_1(y)^2$$
$$= e^{-2\lambda_0 y}L_2(y)^2[L_1(2y) - L_1(y)^2],$$

so that

$$\varepsilon_1^{\text{CRPS}} = \int_0^\infty e^{-2\lambda_0 y}L_2(y)^2[L_1(2y) - L_1(y)^2]dy. \tag{35}$$

The expressions in (34) and (35) are easily implemented in a software such as Mathcad, Matlab, Mathematica (the first two are used by the authors). For the parameterization described in Section 1, one obtains the values of the sensitivity measures in Table 1.

Appendix B: Proofs

Proof of Proposition 2

We can write $\epsilon_X^{\mathcal{S}} = \mathbb{E}_X[\zeta^{\mathcal{S}}(\mathbb{P}_Y, \mathbb{P}_{Y|X})]$, where

$$\zeta^{\mathcal{S}}(\mathbb{P}_Y, \mathbb{P}_{Y|X}) = \mathbb{E}_Y[S(Y, a^*(X)) - S(Y, a^*)|X].$$

We also know that $\zeta^{\mathcal{S}}(\mathbb{P}_Y, \mathbb{P}_{Y|X})$ is greater than or equal to zero for all values of X. Then, if the scoring rule is strictly proper, $a^*(X) = F_{Y|X}$ and $a^* = F_Y$ must maximize the expected score. Therefore,

$$\zeta^{\mathcal{S}}(\mathbb{P}_Y, \mathbb{P}_{Y|X}) = \mathbb{E}_Y[S(Y, F_{Y|X}) - S(Y, F_Y)|X].$$

We already know $\epsilon_X^{\mathcal{S}} = 0$ if Y, X are independent, because $\epsilon_X^{\mathcal{S}}$ is a sensitivity measure. Conversely, suppose that $\epsilon_X^{\mathcal{S}} = 0$. Then, the nonnegativity of $\zeta^{\mathcal{S}}(\mathbb{P}_Y, \mathbb{P}_{Y|X})$ forces $\zeta^{\mathcal{S}}(\mathbb{P}_Y, \mathbb{P}_{Y|X})$ to be zero for almost all X. That is, for almost all X, $\mathbb{E}_Y[S(Y, F_{Y|X})|X] = \mathbb{E}_Y[S(Y, F_Y)|X]$. Now, because S is strictly proper, the distribution $F_{Y|X}$ is, for each value X, the unique maximizer of $\mathbb{E}_Y[S(Y, F_{Y|X})|X]$. Therefore $F_{Y|X} = F_Y$ for almost all X, which shows that Y and X are independent.

Proof of Theorem 1

For brevity, let u_B be the random variable $U(\mathbb{P}_{Y|B})$ for $B \in \mathcal{B}$, so that $\mathbb{E}[u_B|A] = \mathbb{E}[U(\mathbb{P}_{Y|B})|A]$. Because the σ-algebra $\mathcal{B}$ contains only finitely many sets, it has a finite basis $B_1, ..., B_m$ of mutually exclusive and collectively exhaustive event sets. Because u_B is $\mathcal{B}$-measurable, it follows that u_B has only finitely many values, one for each B_i, and can be regarded as a vector. Let π_A be the vector of conditional probabilities given A, having in the same way one value $p_{i|A}$ for each B_i. Then $\mathbb{E}[u_B|A] = \pi_A \cdot u_B$. To reduce notation clutter, we take all statements $A \in \mathcal{B}$ below to mean $A \in \mathcal{B} \setminus \{\emptyset\}$. Then the assumptions that U is proper and consistent with ζ can be written

$$\begin{aligned} \pi_A \cdot u_B &\leq \zeta_A + u_0 \qquad A \in \mathcal{B}, A \neq B \\ \pi_B \cdot u_B &= \zeta_B + u_0. \end{aligned} \tag{36}$$

Our goal is to show that this system has a solution, implying the existence of a proper utility function consistent with ζ.

Following Ch.1 in Stoer and Witzgall (1970), we use the Kuhn-Fourier Theorem to write down necessary and sufficient conditions for this system to have a solution. The possible legal linear combinations of the system (36) are

$$\sum_{\substack{A \in \mathcal{B} \\ A \neq B}} V_A \pi_A + W \pi_B \leq \sum_{\substack{A \in \mathcal{B} \\ A \neq B}} V_A(\zeta_A + u_0) + W(\zeta_B + u_0) \tag{37}$$

where $V_A \geq 0$ for all A, $V_A > 0$ for some A, and

$$W \pi_B = W(\zeta_B + u_0) \tag{38}$$

with W$\in \mathbb{R}$.

A legal linear combination of a system of equations and inequalities is a *legal linear dependence* if its left side is zero but not all V_A and W are zero. The Kuhn-Fourier Theorem states that the system (36) has a solution if and only if every legal linear dependence is always true. For (37), this means that if $V_A \geq 0$ all A, $V_A > 0$ for some A, and $W \in \mathbb{R}$ then

$$\sum_{\substack{A \in \mathcal{B} \\ A \neq B}} V_A \pi_A + W \pi_B = 0 \Rightarrow \sum_{\substack{A \in \mathcal{B} \\ A \neq B}} V_A(\zeta_A + u_0) + W(\zeta_B + u_0) \geq 0. \tag{39}$$

For (38), this means that for $W \neq 0$

$$W \pi_B = 0 \Rightarrow W(\zeta_B + u_0) = 0. \tag{40}$$

The last implication is vacuously true, since π_B is never zero for $B \neq \emptyset$. Consider then (39). Note that $\sum_{\substack{A \in \mathcal{B} \\ A \neq B}} V_A \pi_A > 0$, because $V_A > 0$ for some A. Therefore, in the premise of (39), W must be strictly negative. Replace W by its negative and solve on both sides of (39) to get the following equivalent version of (39):

$$\pi_B = \sum_{\substack{A \in \mathcal{B} \\ A \neq B}} V_A \pi_A \Rightarrow \zeta_B + u_0 \leq \sum_{\substack{A \in \mathcal{B} \\ A \neq B}} V_A(\zeta_A + u_0), \tag{41}$$

where now we are using new V_A equal to the old $V_A/(-W)$. Multiply each side of the premise to (41) by a vector of ones to obtain $1 = \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A$. Therefore, the scalar u_0 on the right side of (41) cancels, and (41) is equivalent to

$$\pi_B = \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A \pi_A \Rightarrow \zeta_B \leq \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A \zeta_A, \tag{42}$$

for every collection $\{V_A | A \in \mathcal{B}, A \neq B\}$ with $V_A \geq 0$ and $\sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A. = 1$. If we can demonstrate this, then it follows that the system (36) has a solution.

So suppose $V_A \geq 0$ and $\sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A. = 1$, and the premise of (42) holds. In terms of the variables $p_{i|B}$ mentioned at the beginning of this proof, this premise is equivalent to

$$p_{i|B} = \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A p_{i|A} \qquad i = 1, ..., m$$

Therefore

$$\begin{aligned}\mathbb{P}_{Y|B}(dy) &= \sum_i \mathbb{P}_{Y|B_i}(dy) p_{i|B} = \sum_i \mathbb{P}_{Y|B_i}(dy) \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A p_{i|A} \\ &= \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A \sum_i \mathbb{P}_{Y|B_i}(dy) p_{i|A} = \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A \mathbb{P}_{Y|A}(dy)\end{aligned}$$

Then invoking the convexity of ζ in its second argument, we have

$$\xi_B = \zeta(\mathbb{P}_Y, \mathbb{P}_{Y|B}) = \zeta(\mathbb{P}_Y, \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A \mathbb{P}_{Y|A}) \leq \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A \zeta(\mathbb{P}_Y, \mathbb{P}_{Y|A}) = \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A \xi_A.$$

Therefore the conclusion of (42) holds, and we have demonstrated (42). Therefore, the system (36) has a solution.

Proofs of Convexity Claims for Corollary 1

For the sensitivity measure η_X, we have

$$\zeta^\eta(\mathbb{P}, \mathbb{Q}) = (\mu_\mathbb{P} - \mu_\mathbb{Q})^2 = \left(\mu_\mathbb{P} - \int_\mathbb{R} y\mathbb{Q}(dy)\right)^2.$$

This is a convex quadratic function of a linear function $\mathbb{Q} \mapsto \int_\mathbb{R} y\mathbb{Q}(dy)$, hence is convex.
For the sensitivity measure δ_X, we have

$$\zeta^{L1}(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} \int_\mathbb{R} |f_\mathbb{P}(y) - f_\mathbb{Q}(y)|\, dy.$$

This is a composition of mappings $f_{\mathbb{Q}} \mapsto |f_{\mathbb{P}} - f_{\mathbb{Q}}| \mapsto \frac{1}{2}\int_{\mathbb{R}} |f_{\mathbb{P}}(y) - f_{\mathbb{Q}}(y)|\, dy$, which is a linear functional (on a space of functions) following a convex function (from a space of densities to a function space) of $f_{\mathbb{Q}}$. Hence the composition is convex in $f_{\mathbb{Q}}$.
For the sensitivity measure θ_X, we have

$$\zeta^{KL}(\mathbb{P},\mathbb{Q}) = \int_{\mathbb{R}^+} f_{\mathbb{Q}}(y)\left(\ln f_{\mathbb{Q}}(y) - \ln f_{\mathbb{P}}(y)\right) dy = \int_{\mathbb{R}^+} f_{\mathbb{Q}}(y) \ln f_{\mathbb{Q}}(y) dy - \int_{\mathbb{R}^+} f_{\mathbb{Q}}(y) \ln f_{\mathbb{P}}(y) dy.$$

The second term in this difference is linear in $f_{\mathbb{Q}}$, so the overall function will be convex in $f_{\mathbb{Q}}$ if the first term is. Note that the first term is a composition $f_{\mathbb{Q}} \mapsto f_{\mathbb{Q}} \cdot \ln g \mapsto \int_{\mathbb{R}+} f_{\mathbb{Q}}(y) \ln f_{\mathbb{Q}}(y) dy$, which is a linear function following the transformation $f_{\mathbb{Q}} \mapsto f_{\mathbb{Q}} \cdot \ln f_{\mathbb{Q}}$, and the latter is convex because its pointwise analog $y \mapsto y \cdot \ln y$ is convex, as may be verified by checking the second derivative. Therefore the overall transformation is convex in $f_{\mathbb{Q}}$.
For the sensitivity measure β^{KS}, we have

$$\zeta^{KS}(\mathbb{P},\mathbb{Q}) = \sup_{y\in\mathbb{R}} |F_{\mathbb{P}}(y) - F_{\mathbb{Q}}(y)|\,,$$

which is a composition $F_{\mathbb{Q}} \mapsto |F_{\mathbb{P}} - F_{\mathbb{Q}}| \mapsto \sup_{y\in\mathbb{R}} |F_{\mathbb{P}}(y) - F_{\mathbb{Q}}(y)|$, that is, a linear function following a convex function of $F_{\mathbb{Q}}$. The composition is therefore convex. For sensitivity measure β^{Ku}, $\zeta^{Ku}(\mathbb{P},\mathbb{Q})$ is convex in $F_{\mathbb{Q}}$ by the same logic as for $\zeta^{KS}(\mathbb{P},\mathbb{Q})$.

Calculations for Equations (8) and (9) in Section 3.3

For the inner operator in (8), we have:

$$\begin{aligned}
\eta'_X &= \mathbb{E}_X\left[\zeta'^{\eta}(\mathbb{P}_Y,\mathbb{P}_{Y|X})\right] = \mathbb{E}_X\left[(\mu_Y - \mu_{Y|X} + k)^2 - k^2\right] \\
&= \mathbb{E}_X\left[(\mu_Y - \mu_{Y|X})^2 + 2k(\mu_Y - \mu_{Y|X})\right] \mathbb{E}_X\left[(\mu_Y - \mu_{Y|X})^2\right] + 2k\mathbb{E}_X\left[\mu - \mu_{Y|X}\right] \\
&= \mathbb{E}_X\left[(\mu_Y - \mu_{Y|X})^2\right] + 0 = \eta_X.
\end{aligned}$$

for arbitrary Y, X.
For the inner operator in (9), because $z^+ = \frac{1}{2}(z + |z|)$, we have for any Y, X

$$\begin{aligned}
\mathbb{E}_X[\zeta^+(\mathbb{P}_Y,\mathbb{P}_{Y|X})] &= \mathbb{E}_X\left[\int_{\mathbb{R}} \left(f_Y(y) - f_{Y|X}(y)\right)^+ dy\right] \\
&= \mathbb{E}_X\left[\int_{\mathbb{R}} \left(\frac{1}{2}\left|f_Y(y) - f_{Y|X}(y)\right| + \frac{1}{2}\left(f_Y(y) - f_{Y|X}(y)\right)\right) dy\right] \\
&= \mathbb{E}_X\left[\frac{1}{2}\int_{\mathbb{R}} \left|f_Y(y) - f_{Y|X}(y)\right|\right] dy + 0 = \delta_X.
\end{aligned}$$

Proof of Proposition 3

To prove Proposition 3, we first prove the following result.

Proposition 7. *Consider a scoring rule $S(y,a)$ and suppose $a^S_{\mathbb{P}}$ is the set of optimal reports whenever Y has distribution $\mathbb{P}$. A necessary condition for the sensitivity measure ξ_X with inner operator ζ to be, for all X, the value ϵ^S_X of information X under scoring rule S is that for all distributions $\mathbb{Q}_1, \mathbb{Q}_0$, and all α with $0 < \alpha < 1$,*

$$a^S_{\mathbb{Q}_0} = a^S_{\mathbb{Q}_1} \Rightarrow \zeta\left((1-\alpha)\mathbb{Q}_0 + \alpha\mathbb{Q}_1, \mathbb{Q}_0\right) = \zeta\left((1-\alpha)\mathbb{Q}_0 + \alpha\mathbb{Q}_1, \mathbb{Q}_1\right) = 0. \tag{43}$$

The proof of the above proposition is divided into two steps. First, we state and prove the following lemma:

Lemma 1. *If a^S_X does not depend on X, then $\epsilon^S_X = 0$.*

Proof of Lemma 1 Suppose $a^S_X = a_0$ for all X, that is, $a = a_0$ optimizes $\mathbb{E}[S(Y,a)|X]$ regardless of X. Then $a = a_0$ must also optimize $\mathbb{E}[\mathbb{E}[S(Y,a)|X]]$. But the latter is equal to $\mathbb{E}[S(Y,a)]$. Therefore $a = a_0$ and $a = a^S$ are both optimizers of $\mathbb{E}[S(Y,a)]$, so that $\mathbb{E}[S(Y,a_0)] = \mathbb{E}[S(Y,a^S)]$. Consequently $\epsilon^S_X = \mathbb{E}[S(Y,a_0)] - \mathbb{E}[S(Y,a^S)] = 0$. □
We now prove Proposition 7. As in the statement of the proposition, let $\mathbb{Q}_0$ and $\mathbb{Q}_1$ be arbitrary distributions over the possible values of Y such that $a^S_{\mathbb{Q}_0} = a^S_{\mathbb{Q}_1}$, and let

$$\mathbb{P} = (1-\alpha)\mathbb{Q}_0 + \alpha\mathbb{Q}_1.$$

for $0 < \alpha < 1$. Let X be a binary variable with $\alpha = P(X=1) = 1 - P(X=0)$, and suppose Y has conditional distributions $\mathbb{P}_{Y|X=0} = \mathbb{Q}_0$, and $\mathbb{P}_{Y|X=1} = \mathbb{Q}_1$. Then $\mathbb{P}_Y = \mathbb{P}$. Under scoring rule S, the optimal report set given $X = 1$ is $a^S_{Q_1}$ and the optimal report set given $X = 0$ is $a^S_{Q_0}$. Because these two sets are by hypothesis the same, the information value of X under score S must be zero according to the lemma. Suppose the information value of X under S is equal to the sensitivity measure ξ_X. We therefore have

$$0 = \xi_X = \mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X})] = (1-\alpha)\zeta(\mathbb{P}, \mathbb{Q}_0) + \alpha\zeta(\mathbb{P}, \mathbb{Q}_1)$$

for $0 < \alpha < 1$, as desired.□
We can then prove Proposition 3. The necessary condition of the proposition is equivalent to a simplified version of (43), namely to

$$a^S_{\mathbb{Q}_0} = a^S_{\mathbb{Q}_1} \implies \zeta(\mathbb{Q}_0, \mathbb{Q}_1) = \zeta(\mathbb{Q}_1, \mathbb{Q}_0) = 0 \tag{44}$$

which we now demonstrate. The continuity hypothesis of the proposition implies

$$\lim_{\alpha\uparrow 1}\zeta\left((1-\alpha)\mathbb{Q}_0 + \alpha\mathbb{Q}_1, \mathbb{Q}_0\right) = \zeta\left(\mathbb{Q}_1, \mathbb{Q}_0\right).$$

Then from (43), because $\zeta\left((1-\alpha)\mathbb{Q}_0 + \alpha\mathbb{Q}_1, \mathbb{Q}_0\right) = 0$ for all $\alpha \in (0,1)$, we obtain $\zeta\left(\mathbb{Q}_1, \mathbb{Q}_0\right) = 0$. Similarly, the continuity hypothesis of the proposition implies

$$\lim_{\alpha\downarrow 0}\zeta\left((1-\alpha)\mathbb{Q}_0 + \alpha\mathbb{Q}_1, \mathbb{Q}_1\right) = \zeta\left(\mathbb{Q}_0, \mathbb{Q}_1\right),$$

whence we obtain for a similar reason $\zeta\left(\mathbb{Q}_0, \mathbb{Q}_1\right) = 0$.

Proof of Proposition 4

By definition and because the optimal action under Bregman scoring is the mean,

$$\begin{aligned}\epsilon_X^B &= \mathbb{E}_X\left[\mathbb{E}[\psi(\mu_{Y|X}) + \psi'(\mu_{Y|X})(Y-\mu_{Y|X}) - \psi(Y)|X]\right.\\ &\quad \left. - \mathbb{E}[\psi(\mu_Y) + \psi'(\mu_Y)(Y-\mu_Y) - \psi(Y)]\right]\\ &= \mathbb{E}_X\left[\mathbb{E}[\psi(\mu_{Y|X})] + \psi'(\mu_{Y|X})\mathbb{E}[(Y-\mu_{Y|X})] - \mathbb{E}[\psi(Y)|X]]\right.\\ &\quad \left. - \mathbb{E}[\psi(\mu_Y)] - \psi'(\mu_Y)\mathbb{E}[(Y-\mu_Y)] + \mathbb{E}[\psi(Y)]\right]\\ &= \mathbb{E}_X\left[\mathbb{E}[\psi(\mu_{Y|X})] + 0 - \mathbb{E}[\psi(Y)|X]] - \mathbb{E}[\psi(\mu_Y)] - 0 + \mathbb{E}[\psi(Y)]\right]\\ &= \mathbb{E}_X[\psi(\mu_{Y|X})] - \mathbb{E}[\psi(Y)] - \mathbb{E}[\psi(\mu_Y)] + \mathbb{E}[\psi(Y)] = \mathbb{E}_X[\psi(\mu_{Y|X})] - \psi(\mu_Y).\end{aligned}$$

Also from Proposition 1, we have

$$\begin{aligned}\zeta^B(\mathbb{P}_Y, \mathbb{P}_{Y|X}) &= \mathbb{E}\left[S^B(Y,\mu_{Y|X}) - S^B(Y,\mu_Y)\middle|X\right]\\ &= \mathbb{E}\left[\psi(\mu_{Y|X}) + \psi'(\mu_{Y|X})(Y-\mu_{Y|X}) - \psi(Y)\right.\\ &\qquad \left. - (\psi(\mu_Y) + \psi'(\mu_Y)(Y-\mu_Y) - \psi(Y))\middle|X\right]\\ &= \mathbb{E}\left[\psi(\mu_{Y|X}) - (\psi(\mu_Y) + \psi'(\mu_Y)(Y-\mu_Y))\middle|X\right]\\ &= \psi(\mu_{Y|X}) - (\psi(\mu_Y) + \psi'(\mu_Y)(\mu_{Y|X}-\mu_Y))\end{aligned}$$

as desired.

Proof of Proposition 5

By definition, we have:

$$\begin{aligned}\mathbb{E}[S_p^Q(Y,a)] &= h\int_{-\infty}^{\infty}\left[p(t(y)-t(a))^+ + (1-p)(t(a)-t(y))^+\right]dF_Y(y)\\ &= h\left[p\int_a^{\infty}(t(y)-t(a))\,dF_Y(y) + (1-p)\int_{-\infty}^{a}(t(a)-t(y))\,dF_Y(y)\right]\\ &= h\big[p\left(\mathbb{E}^{a,\infty}[t(Y)] - t(a)\left(1-F_Y(a)\right)\right) + (1-p)\left(t(a)F_Y(a) - \mathbb{E}^{-\infty,a}[t(Y)]\right)\big]\\ &= h\left[p\mathbb{E}^{a,\infty}[t(Y)] - (1-p)\mathbb{E}^{-\infty,a}[t(Y)] + t(a)\left(F_Y(a)-p\right)\right]\end{aligned}$$

Substituting $a^* = Q_Y(p)$ and using the identity $\mathbb{E}^{b,\infty}(g(Z)) = \mathbb{E}(g(Z)) - \mathbb{E}^{-\infty,b}(g(Z))$ yields

$$\begin{aligned}\mathbb{E}[S_p^Q(Y,a^*)] &= h\left[(1-p)\mathbb{E}_Y^{-\infty,Q_Y(p)}[t(Y)] - p\mathbb{E}^{Q_Y(p),\infty}[t(Y)] + t(Q_Y(p))\left(p - F(Q_Y(p))\right)\right]\\ &= h\left[\mathbb{E}^{-\infty,Q_Y(p)}[t(Y)] - p\mathbb{E}[t(Y)] + t(Q_Y(p))\left(p - F_Y(Q_Y(p))\right)\right].\end{aligned}$$

An analogous expression holds for the conditional r.v. $Y|X$. Inserting these into (3), we get

$$\begin{aligned}\epsilon_X^Q &= h\mathbb{E}_X\left\{\mathbb{E}^{-\infty,Q_{Y|X}(p)}\left[t(Y)|X\right] - p\mathbb{E}\left[t(Y)|X\right] + t(Q_{Y|X}(p))\left(p - F_{Y|X}(Q_{Y|X}(p))\right)\right\} \\ &\quad - h\left\{\mathbb{E}^{-\infty,Q_Y(p)}\left[t(Y)\right] - p\mathbb{E}\left[t(Y)\right] + t(Q_Y(p))\left(p - F_Y(Q_Y(p))\right)\right\} \\ &= h\mathbb{E}_X\left\{\mathbb{E}^{-\infty,Q_{Y|X}(p)}\left[t(Y)|X\right] + t(Q_{Y|X}(p))\left(p - F_{Y|X}(Q_{Y|X}(p))\right)\right\} \\ &\quad - h\left\{\mathbb{E}^{-\infty,Q_Y(p)}\left[t(Y)\right] + t(Q_Y(p))\left(p - F_Y(Q_Y(p))\right)\right\},\end{aligned}$$

which completes the proof.

Derivation for Equation (20)

The conditional and unconditional densities of Y in (19) are

$$f_Y(y) = \nu(y, m_Y, \sigma_Y) \quad \text{and} \quad f_{Y|X}(y, x_i) = \nu(y - a_i x_i, m_{Y|X}(x_i), \sigma_{Y|X}), \tag{45}$$

where $\nu(y, m_Y, \sigma_Y)$ is the normal density with mean $m_Y = \sum_{\ell=1}^n a_\ell \mu_\ell$ and variance $\sigma_Y^2 = \mathbf{a}\Sigma\mathbf{a}^T$, and $\nu(y - a_i x_i, m_{Y|X}(x_i), \sigma_{Y|X})$ is the normal density with mean $m_{Y|X}(x_i) = \sum_{\ell=1}^{n} a_\ell \left[\mu_\ell + (x_i - \mu_i)\frac{\sigma_{i,\ell}}{\sigma_{i,i}}\right]$, variance $\sigma_{Y|X}^2 = \mathbf{a}\Sigma_{Y|X}\mathbf{a}^T$, with $\Sigma_{Y|X} = \left[\sigma_{\ell,s} - \frac{\sigma_{\ell,i} \cdot \sigma_{i,s}}{\sigma_{i,i}}\right]_{\ell,s=1,2,\ldots,n}$. Then, because the normal distribution is a member of the location-scale family, we can write $Q_Y(\alpha) = \mu_Y + \Phi^{-1}(\alpha)\sigma_Y$ and by Winkler et al. (1972)

$$\mathbb{E}^{-\infty,z}(Y) = \mu_Y \Phi\left(\tfrac{z-\mu_Y}{\sigma_Y}\right) - \sigma_Y \varphi\left(\tfrac{z-\mu_Y}{\sigma_Y}\right). \tag{46}$$

Substituting this into (18), we obtain (20).

Derivation for Equations (26) and (28)

For (26),

$$\begin{aligned}\epsilon_X^{Power} &= \mathbb{E}_X\mathbb{E}[S^{Power}(Y,q)|X] - \mathbb{E}[S^{Power}(Y,q)] \\ &= \mathbb{E}_X\left[\int_{\mathbb{R}} \omega \left[s q_{Y|X}^{s-1}(y) - (s-1)\int_{\mathbb{R}} q_{Y|X}^s(t)dt\right] s q_{Y|X} dy\right] \\ &\quad - \int_{\mathbb{R}} \omega \left[s q_Y^{s-1}(y) - (s-1)\int_{\mathbb{R}} q_Y^s(t)dt\right] s q_Y dy \\ &= \omega\mathbb{E}_X\left[\int_{\mathbb{R}} \left(q_{Y|X}^s(y) - s q_{Y|X}(y) q_Y^{s-1}(y) + (s-1) q_Y^s(y)\right) dy\right],\end{aligned}$$

which is the desired result. Similarly, for (28), we have:

$$\begin{aligned}\epsilon_X^{Spherical} &= \mathbb{E}_X\mathbb{E}[S^{Spherical}(Y,q)|X] - \mathbb{E}[S^{Spherical}(Y,q)] \\ &= \mathbb{E}_X\left[\int_{\mathbb{R}} \left(\frac{q_{Y|X}(y)}{\|\mathbf{q}_{Y|X}\|_s}\right)^{s-1} q_{Y|X}(y) dy\right] - \int_{\mathbb{R}} \left(\frac{q_Y(y)}{\|\mathbf{q}_Y\|_s}\right)^{s-1} dy,\end{aligned}$$

where $\|\mathbf{q}\|_s = (\int q(y)^s dy)^{1/s}$ and which gives us (28).

Proof of Proposition 6

By definition of information value, we have

$$\epsilon_X^{CRPS} = \mathbb{E}_X \left\{ \max_{F \in A} \mathbb{E}\left[S^{CRPS}(Y,F) \middle| X \right] \right\} - \max_{F \in A} \mathbb{E}\left[S^{CRPS}(Y,F) \right].$$

and by strict properness, the two maximizers are $F = F_{Y|X}$ and $F = F_Y$, respectively. Then

$$\begin{aligned}
\epsilon_X^{CRPS} &= \mathbb{E}_X \left\{ \mathbb{E}_Y \left[S^{CRPS}(Y, F_{Y|X}) \middle| X \right] \right\} - \mathbb{E}_Y[S^{CRPS}(Y, F_Y)] \\
&= \mathbb{E}\left\{ \mathbb{E}_Y \left[S^{CRPS}(Y, F_{Y|X}) - S^{CRPS}(Y, F_Y) \middle| X \right] \right\} \\
&= \mathbb{E}_X \left\{ \mathbb{E}_Y \left[\int_{\mathbb{R}} (F_Y(z) - \mathbf{1}\{z \geq Y\})^2 \, dz - \int_{\mathbb{R}} \left(F_{Y|X}(z) - \mathbf{1}\{z \geq Y\} \right)^2 dz \middle| X \right] \right\} \\
&= \mathbb{E}_X \left\{ \mathbb{E}_Y \left[\int_{\mathbb{R}} \left(F_Y^2(z) - F_{Y|X}^2(z) - 2\left(F_Y(z) - F_{Y|X}(z) \right) \cdot \mathbf{1}\{z \geq Y\} \right) dz \middle| X \right] \right\} \\
&= \mathbb{E}_X \left\{ \int_{\mathbb{R}} \left(F_Y^2(z) - F_{Y|X}^2(z) - 2\left(F_Y(z) - F_{Y|X}(z) \right) \cdot F_{Y|X}(z) \right) dz \right\} \\
&= \mathbb{E}\left\{ \int_{\mathbb{R}} \left(F_Y(z) - F_{Y|X}(z) \right)^2 dz \right\}.
\end{aligned}$$

References

Anderson, T.W. 1962. On the Distribution of the Two-Sample Cramer von Mises Criterion. *Annals of Mathematical Statistics* **33**(3) 1148–1159.

Artzner, P., F. Delbaen, J.M. Eber, D. Heath. 1999. Coherent Measures of Risk. *Mathematical Finance* **9** 203–228.

Baucells, M., E. Borgonovo. 2013. Invariant Probabilistic Sensitivity Analysis. *Management Science* **59**(11) 2536–2549.

Bernardo, J.M. 1979a. Expected Information as Expected Utility. *The Annals of Statistics* **7**(3) 686–690.

Bernardo, J.M. 1979b. Reference Posterior Distributions for Bayesian Inference. *Journal of the Royal Statistical Society. Series B (Methodological)* **41**(2) 113–147.

Bernardo, J.M., A.F.M. Smith. 1994. *Bayesian Theory*. First Edition, Wiley&Sons, Chichester.

Borgonovo, E. 2007. A New Uncertainty Importance Measure. *Reliability Engineering and System Safety* **92**(6) 771–784.

Borgonovo, E., G. Hazen, E. Plischke. 2016. A Common Rationale for Global Sensitivity Measures and Their Estimation. *Risk Analysis* **36**(10) 1871–1895.

Borgonovo, E., E. Plischke. 2016. Sensitivity Analysis: A Review of Recent Advances. *European Journal of Operational Research* **3**(1) 869–887.

Borgonovo, E., S. Tarantola, E. Plischke, M.D. Morris. 2014. Transformation and Invariance in the Sensitivity Analysis of Computer Experiments. *Journal of the Royal Statistical Society Series B* **76**(5) 925–947.

Bratvold, R.B., J.E. Bickel, H.P. Lohne. 2009. Value of Information in the Oil and Gas Industry: Past, Present and Future. *SPE Reservoir Evaluation and Engineering* **12**(4) 630–638.

Browne, T., J.-C. Fort, B. Iooss, L. Le Gratiet. 2017. Estimate of Quantile-Oriented Sensitivity Indices. *HAL repository Working Paper* 1–24.

Cervera, J.L., J. Muñoz. 1996. Proper Scoring Rules for Fractiles. In: J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith, eds., *Bayesian Statistics 5*. Oxford University Press, Oxford, UK, 513–519.

Clemen, R.T. 1997. Making Hard Decisions: An Introduction to Decision Analysis. Duxbury Press, Pacific Grove, Calif, USA.

Critchfield, G.G., K.E. Willard. 1986. Probabilistic Analysis of Decision Trees using Monte Carlo Simulation. *Medical Decision Making* **6**(2) 85–92.

Dawid, A.P. 2007. The Geometry of Proper Scoring Rules. *Annals of the Institute of Statistical Mathematics* **59**(1) 77–93.

Eschenbach, T.G. 1992. Spiderplots versus Tornado Diagrams for Sensitivity Analysis. *Interfaces* **22**(6) 40–46.

Felli, J.C., G.B. Hazen. 1998. Sensitivity Analysis and the Expected Value of Perfect Information. *Medical Decision Making* **18** 95–109.

Felli, J.C., G.B. Hazen. 1999. A Bayesian Approach to Sensitivity Analysis. *Health Economics* **8** 263–268.

Fermanian, J.-D., O. Scaillet. 2005. Sensitivity Analysis of VaR and Expected Shortfall for Portfolios under Netting Agreements. *Journal of Banking & Finance* **29**(4) 927–958.

Ferretti, F., Saltelli, A., and S. Tarantola. 2016. Trends in Sensitivity Analysis Practice in the Last Decade. *The Science of the Total Environment* **568** 666–670.

Fort, J., T. Klein, N. Rachdi. 2015. New Sensitivity Analysis Subordinated to a Contrast. *Communications in Statistics: Theory and Methods* **45**(15) 4349–4364.

French, S. 2017. Societal Risk and Decision Analyses and the Political Process. *Fifth Games Decision and Reliability Symposium,* Madrid, June 7–9 2017.

Gamboa, F., T. Klein, A. Lagnoux. 2015. Sensitivity Analysis Based on Cramer von Mises Distance. *arXiv:1506.04133v1*, *Working Paper*.

Ghanem, R., D. Higdon, D. Owhadi, eds. 2016. *Handbook of Uncertainty Quantification*. Springer.

Gilboa, I., E. Lehrer. 1991. The Value of Information: An Axiomatic Approach. *Journal of Mathematical Economics* **20** 443–459.

Gneiting, T.G. 2011. Making and Evaluating Point Forecasts. *Journal of the American Statistical Association* **106**(74) 746–762.

Gneiting, T.G., M. Katzfuss. 2014. Probabilistic Forecasting. *Annual Reviews of Statistics and Its Applications* **1** 25–51.

Gneiting, T.G., A.E. Raftery. 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102**(477) 359–377.

Gourieroux, C., J.P. Laurent, O. Scaillet. 2000. Sensitivity Analysis of Values at Risk. *Journal of Empirical Finance* **7**(3-4) 225–245.

Hazen, G.B., J. Sounderpandian. 1999. Lottery Acquisition versus Information Acquisition: Prices and Preference Reversals. *Journal of Risk and Uncertainty* **18**(2) 125–136.

Hilton, R.W. 1981. The Determinants of Information Value: Synthesizing Some General Results. *Management Science* **27**(1) 57–64.

Hoeffding, W. 1948. A Non Parametric Test of Independence. *Annals of Mathematical Statistics* **90**(4) 546–557.

Howard, R. A. 1966. Decision Analysis: Applied Decision Theory. *Proc. Fourth Internat. Conf. Oper. Res., Wiley-Interscience, New York* .

Howard, R.A. 1983. The Evolution of Decision Analysis. *Readings on the Principles and Applications of Decision Analysis. Vol. I*. Strategic Decisions Group, Menlo Park, CA.

Hu, Z., J. Cao, L.J. Hong. 2012. Robust Simulation of Global Warming Policies Using the DICE Model. *Management Science* **58**(12) 2190–2206.

Iman, R.L., S.C. Hora. 1990. A Robust Measure of Uncertainty Importance for Use in Fault Tree System Analysis. *Risk Analysis* **10** 401–406.

Ishigami, T., T. Homma. 1990. An Importance Quantification Technique in Uncertainty Analysis for Computer Models. *ISUMA'90, First International Symposium on Uncertainty Modelling and Analysis*. University of Maryland.

Jose, V.R.R., R.F. Nau, R.L. Winkler. 2009. Sensitivity to Distance and Baseline Distributions in Forecast Evaluation. *Management Science* **55**(4) 582–590.

Keisler, J.M., Z.A. Collier, E. Chu, N. Sinatra, I. Linkov. 2014. Value of Information Analysis: The State of Application. *Environment Systems and Decisions* **34**(1) 3–23.

Kullback, S., R.A. Leibler. 1951. On Information and Sufficiency. *Annals of Mathematical Statistics* **22**(1) 79–86.

LaValle, I. 1968. On Cash Equivalents and Information Evaluation in Decisions Under Uncertainty: Part I: Basic Theory. *Journal of the American Statistical Association* **63**(321) 252–276.

Marangoni, G., M. Tavoni, V. Bosetti, E. Borgonovo, P. Capros, O. Fricko, D.E.H.J. Gernaat, C. Guivarch, P. Havlik, D. Huppmann, N. Johnson, P. Karkatsoulis, I. Keppo, V. Krey, E. Ó Broin, J. Price, D. P. Van Vuuren. 2017. Sensitivity of Projected Long-Term CO_2 Emissions Across the Shared Socioeconomic Pathways. *Nature Climate Change* **7**(2) 113–117.

NEA. 2001. Confidence in Models of Radionuclide Transport for Site-specific Assessments. Technical report, OECD.

Oakley, J., A. O'Hagan. 2004. Probabilistic Sensitivity Analysis of Complex Models: a Bayesian Approach. *Journal of the Royal Statistical Society, Series B* **66**(3) 751–769.

Oakley, J.E. 2009. Decision-Theoretic Sensitivity Analysis for Complex Computer Models. *Technometrics* **51**(2) 121–129.

OECD. 1989. PSACOIN Level E Intercomparison. An International Code Intercomparison Exercise on a Hypothetical Safety Assessment Case Study for Radioactive Waste Disposal Systems. Technical report, OECD - NEA publication.

Pearson, K. 1905. *On the General Theory of Skew Correlation and Non-linear Regression*, *Mathematical Contributions to the Theory of Evolution, Drapers' Company Research Memoirs*, vol. XIV. Dulau & Co., London.

Pflug, G.C. 2006. A Value-of-Information Approach to Measuring Risk in Multi-Period Economic Activity. *Journal of Banking & Finance* **30**(2) 695–715.

Plischke, E., E. Borgonovo, C.L. Smith. 2013. Global Sensitivity Measures from Given Data. *European Journal of Operational Research* **226**(3) 536–550.

Rabitz, H., O.F. Alis. 1999. General Foundations of High-Dimensional Model Representations. *J. Math. Chem.* **25**(2-3) 197–233.

Rahman, S. 2016. The f-Sensitivity Index. *SIAM/ASA Journal on Uncertainty Quantification* **4**(1) 130–162.

Rényi, A. 1959. On Measures of Dependence. *Acta Mathematica Hungarica* **10**(3-4) 441–451.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola. 2008. *Global Sensitivity Analysis – The Primer*. Wiley&Sons, Chichester.

Saltelli, A., S. Tarantola. 2002. On the Relative Importance of Input Factors in Mathematical Models: Safety Assessment for Nuclear Waste Disposal. *Journal of the American Statistical Association* **97**(459) 702–709.

Saltelli, A, S Tarantola, F Campolongo. 2000. Sensitivity Analysis as An Ingredient of Modelling. *Statistical Science* **19**(4) 377–395.

Savage, L.J. 1971. Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association* **66**(336) 783–801.

Schervish, M.J. 1989. A General Method for Comparing Probability Assessors. *The Annals of Statistics***17**(4) 1856–1879.

Shuford, E.H., A. Albert, H.E. Massengill. 1966. Admissible Probability Measurement Procedures. *Psychometrika* **31**(2) 125–145.

Stoer, J., C.W. Witzgall. 1970. *Convexity and Optimization in Finite Dimensions I, 1970*. Springer, Berlin.

Strong, M., J.E. Oakley. 2013. An Efficient Method for Computing Partial Expected Value of Perfect Information for Correlated Inputs. *Medical Decision Making* **33** 755–766.

Strong, M., J.E. Oakley, A. Brennan. 2014. Estimating Multiparameter Partial Expected Value of Perfect Information from a Probabilistic Sensitivity Analysis Sample: A Nonparametric Regression Approach. *Medical Decision-Making* **34(3)** 311–326.

Szekely, G.J. 1989. *Potential and Kinetic Energy in Statistics*. Budapest Technical University, Budapest.

Szekely, G.J., M.L. Rizzo. 2013. Energy Statistics: Statistics Based on Distances. *Journal of Statistical Planning and Inference* **143**(8) 1249–1272.

Szekely, G.J., M.L. Rizzo. 2017. The Energy of Data. *Annual Review of Statistics and Its Applications* **4** 447–479.

Thomson, W. 1979. Eliciting Production Possibilities from a Well-Informed Manager. *Journal of Economic Theory* **20**(3) 360–380.

Winkler, R.L., V.R.R. Jose. 2011. Scoring Rules. *Wiley Encyclopedia of Operations Research and Management Science*.

Winkler, R.L., G.M. Roodman, R.R. Britney. 1972. The Determination of Partial Moments. *Management Science* **19**(3) 290–296.