

Measurability of functionals and of ideal point forecasts

Tobias Fissler

*Vienna University of Economics and Business (WU),
Department of Finance, Accounting and Statistics,
Welthandelsplatz 1, 1020 Vienna, Austria.
e-mail: tobias.fissler@wu.ac.at*

and

Hajo Holzmann

*Philipps-Universität Marburg, Fachbereich Mathematik und Informatik,
Hans-Meerwein-Straße, 35043 Marburg, Germany.
e-mail: holzmann@mathematik.uni-marburg.de*

Abstract: The ideal probabilistic forecast for a random variable Y based on an information set \mathcal{F} is the conditional distribution of Y given \mathcal{F} . In the context of point forecasts aiming to specify a functional T such as the mean, a quantile or a risk measure, the ideal point forecast is the respective functional applied to the conditional distribution. This paper provides a theoretical justification why this ideal forecast is actually a forecast, that is, an \mathcal{F} -measurable random variable. To that end, the appropriate notion of measurability of T is clarified and this measurability is established for a large class of practically relevant functionals, including elicitable ones. More generally, the measurability of T implies the measurability of any point forecast which arises by applying T to a probabilistic forecast. Similar measurability results are established for proper scoring rules, the main tool to evaluate the predictive accuracy of probabilistic forecasts.

MSC2020 subject classifications: Primary 62C99; secondary 91B06.

Keywords and phrases: Bayes act, elicibility, forecast, information set, scoring function, scoring rule.

Received April 2022.

Contents

1	Introduction	5020
2	Statistical functionals and forecasting	5021
3	Measurability of functionals	5022
4	Elicitable functionals	5027
5	Measurability of scoring rules	5029
	Acknowledgments	5031
	References	5031

1. Introduction

A typical aspect of decision making, be it in business, politics, or private life, is that decisions should account for unknown or future events. Hence, these decisions commonly base on predictions or forecasts for these events. Fortunately, one regularly has at least partial information about the events of interest, such as regressors or feature variables in a cross-sectional setting or past observations of the event in a time series framework, which help to improve the predictions. Expressed in terms of mathematical statistics, the event of interest is a random element Y on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ which often attains values in the real numbers, and the given information is modelled as a sub- σ -algebra $\mathcal{F} \subseteq \mathcal{A}$.

It has been argued in the literature that probabilistic forecasts, taking the form of probability measures, distributions, or densities, are most informative and should be preferred, since they directly express the uncertainty about the future event (Dawid, 1984, Gneiting, Balabdaoui and Raftery, 2007). The *ideal* or truthful probabilistic forecast of Y based on \mathcal{F} is the conditional distribution of Y given \mathcal{F} , denoted by $\mathbb{P}_{Y|\mathcal{F}}$. “However, practical situations may require single-valued point forecasts, for reasons of decision making, reporting requirements, or communications, among others” (Gneiting and Katzfuss, 2014, pp. 134–135). In such situations, one summarises the uncertainty in a typically real-valued functional T of the distribution of Y such as the mean, a quantile, an expectile, or a law-determined risk measure such as Expected Shortfall. Then, the ideal or truthful point forecast for Y based on \mathcal{F} is the functional T applied to the conditional distribution of Y given \mathcal{F} , that is, $T(Y|\mathcal{F}) := T(\mathbb{P}_{Y|\mathcal{F}})$ (Nolde and Ziegel, 2017). This concatenation also turns out to be the main idea in building dynamic or conditional risk measures (Weber, 2006).

Ideal forecasts based on \mathcal{F} need to be \mathcal{F} -measurable random elements, or random variables, for the case of point forecasts. On the one hand, this ensures that they actually only exploit the information in \mathcal{F} and no additional sources. From a more technical perspective, on the other hand, measurability ensures that one can assign probabilities to events induced by forecasts and, more generally, that forecasts are amenable to statistical analysis.

This paper studies when an ideal point forecast based on \mathcal{F} , or more generally, a forecast which arises by applying the functional to a probabilistic forecast based on \mathcal{F} , is indeed \mathcal{F} -measurable. To this end, the appropriate notion of measurability of the target functional T , viewed as a map from a class of probability measures to the reals, is clarified. Moreover, we establish corresponding measurability results for many practically relevant examples, including the class of elicitable ones, which can be written as the minimiser of an expected loss function (Theorem 9). Our results provide a formal justification for the \mathcal{F} -measurability of ideal \mathcal{F} -based point forecasts which seems to have been taken for granted in large parts of the forecast evaluation literature, sometimes tacitly, sometimes with somewhat incomplete arguments, or which has been established under restrictive assumptions (Weber, 2006, Holzmann and Eulert, 2014, Nolde and Ziegel, 2017, Pohle, 2020, Fissler and Hoga, 2021, Gneiting and Resin, 2021, Fissler et al., 2021, Hoga and Dimitriadis, 2022). From a regression

or machine learning perspective where Y is observed together with explanatory variables X , our results justify the measurability of the oracle regression function $x \mapsto T(Y|X = x)$ beyond the classical situation of mean regression to quantile regression (Koenker, 2005), expectile regression (Newey and Powell, 1987) or Expected Shortfall regression (Dimitriadis and Bayer, 2019). In the field of sensitivity analysis, measuring the information content with score improvements also calls for the measurability results established in our paper (Borgonovo et al., 2021, Fissler and Pesenti, 2022).

In Section 2, we provide the formal definitions and concepts of the article. The appropriate concept of measurability of functionals is discussed in Section 3, which leads to the \mathcal{F} -measurability of the ideal \mathcal{F} -based forecast. We show measurability of important functionals such as moments or weighted quantiles and discuss the connection to continuity and robustness results discussed in statistics and finance. Measurability of elicitable functionals is discussed and established in Section 4. Finally, we provide sufficient conditions for the measurability of scoring rules in Section 5. They are used to measure the predictive accuracy of probabilistic forecasts, and map a pair of a predictive distribution P and an observation y to the real number $S(P, y)$. Hence, technically speaking, for fixed y they constitute a real-valued statistical functional, which aligns them with the main topic of this paper. Regarding the application, the measurability of scoring rules formally justifies common statistical practices such as computing Diebold–Mariano tests which test predictive dominance (Diebold and Mariano, 1995).

2. Statistical functionals and forecasting

Real-valued statistical functionals map probability measures on the Borel- σ -algebra of events of the observation domain to real numbers. More formally, the *observation domain* $O \subseteq \mathbb{R}^d$ is a Borel-measurable subset of a Euclidean space \mathbb{R}^d , where we will mainly consider $d = 1$, with *Borel- σ -algebra* \mathcal{O} of O . A statistical functional on some family \mathcal{P} of (Borel-)probability measures on (O, \mathcal{O}) is a map $T : \mathcal{P} \rightarrow \mathbb{R}$ from \mathcal{P} to the real numbers.

A prediction space setting (Gneiting and Ranjan, 2013) consists of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ together with a random variable $Y : \Omega \rightarrow O$, the response variable modelling the quantity of interest, and a sub- σ -algebra \mathcal{F} of \mathcal{A} , the information set on which the forecast is based. \mathcal{F} can be generated, for example, by observable explanatory variables, often called, regressors or features, which can also contain past observations of Y in a time series setting. Then, the conditional distribution $\mathbb{P}_{Y|\mathcal{F}}$ of Y given \mathcal{F} is called the *ideal probabilistic forecast* (Gneiting and Ranjan, 2013, Definition 2.2). If a real-valued quantity shall be issued as a forecast, it is essential to specify which statistical functional T of the conditional distribution $\mathbb{P}_{Y|\mathcal{F}}$ is the target.

In analogy, if one is interested in point forecasts with target functional T , we call $T(Y|\mathcal{F}) := T(\mathbb{P}_{Y|\mathcal{F}})$ the *ideal point forecast*. For $T(Y|\mathcal{F})$ to be defined, we first require that conditional distributions in $\mathbb{P}_{Y|\mathcal{F}}$ are contained in \mathcal{P} , the

domain of T . Second, we require the \mathcal{F} -measurability of the forecast $T(Y|\mathcal{F})$. Our aim is to provide conditions on T which guarantee that, when well-defined, $T(Y|\mathcal{F})$ is automatically a forecast, that is, \mathcal{F} -measurable, and to show that these conditions are satisfied by essentially all statistical functionals which arise in applications.

To proceed, let us recall the formal definition of the conditional distribution $\mathbb{P}_{Y|\mathcal{F}}$. A *Markov kernel* from (Ω, \mathcal{F}) to (O, \mathcal{O}) is a map $\kappa : \Omega \times \mathcal{O} \rightarrow [0, 1]$ such that

1. for each $\omega \in \Omega$, the map $B \mapsto \kappa(\omega, B)$, $B \in \mathcal{O}$, is a probability measure on (O, \mathcal{O}) ,
2. for each $B \in \mathcal{O}$, the map $\omega \mapsto \kappa(\omega, B)$, $\omega \in \Omega$, is $\mathcal{F} - \mathcal{B}([0, 1])$ -measurable.

A *regular version* of the conditional probability of Y given \mathcal{F} , denoted by $\mathbb{P}_{Y|\mathcal{F}}$, is a Markov kernel from (Ω, \mathcal{F}) to (O, \mathcal{O}) such that for each $F \in \mathcal{F}$ and $B \in \mathcal{O}$,

$$\mathbb{P}(F \cap \{Y \in B\}) = \int_F \mathbb{P}_{Y|\mathcal{F}}(\omega, B) \, d\mathbb{P}(\omega).$$

Recall that $\mathbb{P}_{Y|\mathcal{F}}$ is unique only up to almost sure equality.

When forecasting the functional T for Y based on \mathcal{F} , we aim at the map

$$\omega \mapsto T(\mathbb{P}_{Y|\mathcal{F}}(\omega, \cdot)) =: T(Y|\mathcal{F})(\omega), \quad \omega \in \Omega, \quad (1)$$

and we shall assume that the version $\mathbb{P}_{Y|\mathcal{F}}$ can be and is chosen such that $\mathbb{P}_{Y|\mathcal{F}}(\omega, \cdot) \in \mathcal{P}$ for all $\omega \in \Omega$. If κ is a (not necessarily ideal) probabilistic forecast based on \mathcal{F} , that is, a Markov kernel from (Ω, \mathcal{F}) to (O, \mathcal{O}) for which $\kappa(\omega, \cdot) \in \mathcal{P}$ for $\omega \in \Omega$, we may form the resulting point forecast by

$$\omega \mapsto T(\kappa(\omega, \cdot)), \quad \omega \in \Omega. \quad (2)$$

To show that $T(Y|\mathcal{F})$ or, more generally, that $T(\kappa(\cdot, \cdot))$ is a forecast, that is, \mathcal{F} -measurable, it suffices to find a σ -algebra $\mathcal{A}(\mathcal{P})$ such that (a) $\mathbb{P}_{Y|\mathcal{F}}$ or κ , considered as a map from Ω to \mathcal{P} , is $\mathcal{F} - \mathcal{A}(\mathcal{P})$ -measurable, and (b) to show that $T : \mathcal{P} \rightarrow \mathbb{R}$ is $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\mathbb{R})$ -measurable, where $\mathcal{B}(\mathbb{R})$ is the Borel- σ -algebra of \mathbb{R} .

3. Measurability of functionals

We shall choose $\mathcal{A}(\mathcal{P})$ as the projection σ -algebra. More formally, let \mathcal{Q} be the family of all probability measures P on (O, \mathcal{O}) , and for each $B \in \mathcal{O}$, consider the evaluation map

$$\pi_B(P) = P(B), \quad P \in \mathcal{Q}.$$

Then, on \mathcal{Q} we can consider the smallest σ -algebra $\mathcal{A}(\mathcal{Q})$ which makes all evaluation maps $\mathcal{A}(\mathcal{Q}) - \mathcal{B}([0, 1])$ -measurable. For a subset $\mathcal{P} \subseteq \mathcal{Q}$ we denote by $\mathcal{A}(\mathcal{P}) = \{\mathcal{P} \cap A \mid A \in \mathcal{A}(\mathcal{Q})\}$ the trace σ -algebra. Note that $\mathcal{A}(\mathcal{P})$ is also the σ -algebra generated by the restrictions of the evaluation maps π_B to \mathcal{P} , which

follows from the simple fact that for any $C \in \mathcal{B}([0, 1])$ and any $B \in \mathcal{O}$ we have $\pi_{B|\mathcal{P}}^{-1}(C) = \mathcal{P} \cap \pi_B^{-1}(C)$. Moreover, if O is a separable metric space and \mathcal{O} its Borel σ -algebra, $\mathcal{A}(\mathcal{P})$ coincides with the Borel σ -algebra generated by the weak convergence (Gaudard and Hadwin, 1989).

Lemma 1. *Let $\kappa : \Omega \times \mathcal{O} \rightarrow [0, 1]$ be a map such that for each $\omega \in \Omega$, $\kappa(\omega, \cdot) \in \mathcal{P}$. Then κ is a Markov kernel from (Ω, \mathcal{F}) to (O, \mathcal{O}) if and only if the map $\omega \mapsto \kappa(\omega, \cdot)$ from Ω to \mathcal{P} is $\mathcal{F} - \mathcal{A}(\mathcal{P})$ -measurable.*

Lemma 1 follows from Kallenberg (2002, Lemma 1.40). We provide the argument for the sake of completeness.

Proof of Lemma 1. κ is a Markov kernel if and only if for each $B \in \mathcal{O}$, the map $\omega \mapsto \kappa(\omega, B) = \pi_B(\kappa(\omega, \cdot))$ is $\mathcal{F} - \mathcal{B}([0, 1])$ -measurable, which is the case if and only if $\omega \mapsto \kappa(\omega, \cdot)$ is $\mathcal{F} - \mathcal{A}(\mathcal{P})$ -measurable. \square

Concerning the measurability of the forecasts $T(Y|\mathcal{F})$ of functionals T , we observe the following.

Lemma 2. *If the functional $T : \mathcal{P} \rightarrow \mathbb{R}$ is $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\mathbb{R})$ -measurable, then $T(Y|\mathcal{F})$ defined in (1) is $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable.*

Proof. This is clear since by Lemma 1, the conditional distribution $\mathbb{P}_{Y|\mathcal{F}}$ is $\mathcal{F} - \mathcal{A}(\mathcal{P})$ -measurable as a map from Ω to \mathcal{P} , and T is $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\mathbb{R})$ -measurable by assumption. \square

Similarly to Lemma 2, the general forecast $\omega \mapsto T(\kappa(\omega, \cdot))$, $\omega \in \Omega$, where κ is a Markov kernel from (Ω, \mathcal{F}) to (O, \mathcal{O}) , is $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable. Thus, to clarify the measurability of $T(\kappa(\cdot, \cdot))$ and of $T(Y|\mathcal{F})$ in (1) it suffices to investigate T . Before we turn to a general result for elicitable functionals, let us investigate several important examples.

Example 3 (Moments). Let $h : O \rightarrow \mathbb{R}$ be a measurable function. Let \mathcal{P}_h be a family of probability measures on (O, \mathcal{O}) satisfying $\int_O |h(y)| dP(y) < \infty$, and consider the mean functional

$$T_h(P) = \int_O h(y) dP(y), \quad P \in \mathcal{P}_h.$$

Then T_h is $\mathcal{A}(\mathcal{P}_h) - \mathcal{B}(\mathbb{R})$ -measurable.

Indeed, if $h(y) = \mathbb{1}_B(y)$ for some $B \in \mathcal{O}$, then $T_h = \pi_B$ and the claim follows from the measurability of the evaluation map. If h is a simple function, the resulting functional will be a finite linear combination of evaluation maps, and hence also measurable. For $h \geq 0$, there is a sequence of non-negative simple functions (h_n) with $h_n \uparrow h$ pointwise, and then $T_{h_n}(P) \uparrow T_h(P)$ for every P . Hence, as a pointwise limit of the measurable functions T_{h_n} , T_h is also measurable. For a general h , the measurability follows by the decomposition $T_h = T_{h^+} - T_{h^-}$.

Example 3 directly yields the measurability of Borel-measurable functions of multiple moments. This yields, for example, the measurability of the variance,

the skewness, the kurtosis, or of the Sharpe ratio. In the following we identify any probability distribution $P \in \mathcal{Q}$ with its distribution function $F = F_P$ defined by $F(x) = P(O \cap (-\infty, x])$, $x \in O$.

Example 4 (Quantiles). Let $O = \mathbb{R}$, fix $\alpha \in [0, 1]$ and let

$$T(F) = q_\alpha^-(F) = \inf\{x \in \mathbb{R} \mid F(x) \geq \alpha\} \in \bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}, \quad F \in \mathcal{P},$$

be the lower α -quantile of F , which is its essential supremum if $\alpha = 1$. Recall that $\mathcal{B}(\bar{\mathbb{R}}) = \{A \cup E \mid A \in \mathcal{B}(\mathbb{R}), E \subseteq \{-\infty, \infty\}\}$.

To show that T is $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\bar{\mathbb{R}})$ -measurable, consider the family of evaluation maps $\pi_{(-\infty, x]}(F) = F(x)$, for $F \in \mathcal{P}$, $x \in \mathbb{R}$. Each $\pi_{(-\infty, x]}$ is $\mathcal{A}(\mathcal{P}) - \mathcal{B}([0, 1])$ -measurable.

Now, given $p \in \mathbb{R}$, the lower α -quantile $q_\alpha^-(F)$ of F is strictly larger than p , $q_\alpha^-(F) > p$, if and only if $\alpha > F(p)$. Therefore

$$\{F \in \mathcal{P} \mid q_\alpha^-(F) > p\} = \{F \in \mathcal{P} \mid \pi_{(-\infty, p]}(F) < \alpha\},$$

showing measurability of the lower α -quantile, where we exploit the fact that it suffices to show measurability on a generator of $\mathcal{B}(\bar{\mathbb{R}})$.

Similar considerations yield the $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\bar{\mathbb{R}})$ -measurability of the upper α -quantile, $\alpha \in [0, 1]$,

$$T(F) = q_\alpha^+(F) = \sup\{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leq \alpha\} \in \bar{\mathbb{R}}, \quad F \in \mathcal{P},$$

where $q_0^+(F)$ corresponds to the essential infimum of F . For $\alpha \in (0, 1)$, we can also consider the α -quantile, q_α , as an interval-valued functional $q_\alpha(F) = [q_\alpha^-(F), q_\alpha^+(F)] \subset \mathbb{R}$. Then, the previously established measurability results yield that q_α is an *Effros measurable* closed-valued multifunction in the sense that for all open sets $G \subseteq \mathbb{R}$, $\{F \in \mathcal{P} \mid q_\alpha(F) \cap G \neq \emptyset\} \in \mathcal{A}(\mathcal{P})$, see Definition 1.3.1 in Molchanov (2017). Therefore, the ideal forecast $q_\alpha(Y|\mathcal{F})$ is an \mathcal{F} -measurable random closed set in the sense of Definition 1.1.1 in Molchanov (2017), which coincides with the recent findings of de Castro et al. (2021), which are established with different methods.

In the risk management literature, the lower α -quantile is known as Value at Risk at level α , VaR_α . The following example is concerned with another important quantitative risk measure, the Expected Shortfall.

Example 5 (L-functionals). Let $w: [0, 1] \rightarrow [0, \infty)$ be a measurable weight function such that the left-sided and the right-sided limits exist for all $x \in [0, 1]$. Consider the L -functional

$$T_w(F) = \int_0^1 q_\gamma^-(F) w(\gamma) d\gamma \in \bar{\mathbb{R}}, \quad F \in \mathcal{P}_w,$$

where \mathcal{P}_w is the domain such that T_w is well-defined. The measurability of q_γ^- from Example 4 and a Riemann approximation argument yield that T_w is $\mathcal{A}(\mathcal{P}_w) - \mathcal{B}(\bar{\mathbb{R}})$ -measurable.

This directly yields the measurability of the lower and upper Expected Shortfall (by choosing $w = \mathbb{1}_{[0,\alpha]}/\alpha$ or $w = \mathbb{1}_{[\alpha,1]}/(1 - \alpha)$) and of the Range Value at Risk (Cont, Deguest and Scandolo, 2010, Fissler and Ziegel, 2021) (by choosing $w = \mathbb{1}_{[\alpha,\beta]}/(\beta - \alpha)$ for $0 < \alpha < \beta < 1$).

To treat the next Example 7, we first need to establish the following technical lemma.

Lemma 6. *Let $T: \mathcal{P} \rightarrow \mathbb{R}$ be an $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\mathbb{R})$ -measurable functional. Then the composed evaluation maps $\mathcal{P} \rightarrow [0, 1]$,*

$$F \mapsto \pi_{(T(F),\infty)}(F), \quad F \mapsto \pi_{(-\infty,T(F))}(F), \quad F \mapsto \pi_{\{T(F)\}}(F) \quad (3)$$

are $\mathcal{A}(\mathcal{P}) - \mathcal{B}([0, 1])$ -measurable.

Proof. For the first map in (3), let $p \in [0, 1]$ and consider

$$\begin{aligned} \{F \in \mathcal{P} \mid \pi_{(T(F),\infty)}(F) \leq p\} &= \{F \in \mathcal{P} \mid F(T(F)) \geq 1 - p\} \\ &= \{F \in \mathcal{P} \mid T(F) \geq q_{1-p}^-(F)\} \\ &= \{F \in \mathcal{P} \mid (T - q_{1-p}^-)(F) \geq 0\} \in \mathcal{A}(\mathcal{P}). \end{aligned}$$

The last assertion comes from the fact that q_{1-p}^- is $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\mathbb{R})$ -measurable and that differences of measurable functionals are also measurable.

For the second map in (3), the arguments are similar. Using the shorthand $F(x-) := \lim_{t \uparrow x} F(t)$ for the left-sided limit at $x \in \mathbb{R}$, we obtain for $p \in [0, 1]$

$$\begin{aligned} \{F \in \mathcal{P} \mid \pi_{(-\infty,T(F))}(F) \leq p\} &= \{F \in \mathcal{P} \mid F(T(F)-) \leq p\} \\ &= \{F \in \mathcal{P} \mid T(F) \leq q_p^+(F)\} \\ &= \{F \in \mathcal{P} \mid (T - q_p^+)(F) \leq 0\} \in \mathcal{A}(\mathcal{P}). \end{aligned}$$

Finally, note that $\pi_{\{T(F)\}}(F) = 1 - \pi_{(T(F),\infty)}(F) - \pi_{(-\infty,T(F))}(F)$. □

Example 7. Conditional Value at Risk (CoVaR) and conditional Expected Shortfall (CoES) are influential systemic risk measures due to Adrian and Brunnermeier (2016). In a nutshell, they assess the riskiness of a position Y given that a reference position X is at risk. Following Girardi and Tolga Ergün (2013), the latter conditioning event is interpreted as X exceeding its Value at Risk at level β , VaR_β . Hence, these risk measures can be viewed as functionals, first mapping a bivariate distribution $P_{X,Y}$ to the univariate conditional distribution $P_{Y|X \geq \text{VaR}_\beta(X)}$ and then applying a risk measure such as VaR_α (in the case of CoVaR) or ES_α (in the case of CoES) to this univariate distribution. Since VaR_α and ES_α are measurable functionals from univariate distributions to \mathbb{R} , we obtain the measurability of these conditional risk measures if the map $P_{X,Y} \mapsto P_{Y|X \geq \text{VaR}_\beta(X)}$ is measurable. To account for possible discontinuities in the marginal distribution of the first component X , we follow Remark C.1 in Fissler and Hoga (2021) and incorporate a correction term to this map. Hence, we formally consider $\eta: \mathcal{P}^2 \rightarrow \mathcal{P}$, where \mathcal{P}^2 is a set of Borel-distributions on \mathbb{R}^2

and \mathcal{P} is a set of such distributions on \mathbb{R} , such that for a Borel set $B \in \mathcal{B}(\mathbb{R})$ we get

$$\begin{aligned} \eta(\mathbb{P}_{X,Y})(B) &= \mathbb{P}_{Y|X \succcurlyeq \text{VaR}_\beta(X)}(B) := \frac{1}{1-\beta} \left[\mathbb{P}_{X,Y}((\text{VaR}_\beta(\mathbb{P}_X), \infty) \times B) \right. \\ &\quad \left. + \left(1 - \beta - \mathbb{P}_X((\text{VaR}_\beta(\mathbb{P}_X), \infty)) \right) \mathbb{P}_{X,Y}(\{\text{VaR}_\beta(\mathbb{P}_X)\} \times B) / \mathbb{P}_X(\{\text{VaR}_\beta(\mathbb{P}_X)\}) \right] \end{aligned}$$

For the last term, note that $\mathbb{P}_X(\{\text{VaR}_\beta(\mathbb{P}_X)\}) = 0$ implies that $1 - \beta - \mathbb{P}_X((\text{VaR}_\beta(\mathbb{P}_X), \infty)) = 0$ for which case we set $0/0 = 0$. To show the $\mathcal{A}(\mathcal{P}^2) - \mathcal{A}(\mathcal{P})$ -measurability of η it is sufficient to show that for any $B \in \mathcal{B}(\mathbb{R})$, the concatenation with the evaluation map $\pi_B \circ \eta$ is $\mathcal{A}(\mathcal{P}^2) - \mathcal{B}([0, 1])$ -measurable. This measurability follows from Lemma 6 and the fact that for any $C \in \mathcal{B}(\mathbb{R})$ the map $\mathbb{P}_{X,Y} \mapsto \mathbb{P}(\cdot \times C)$ is $\mathcal{A}(\mathcal{P}^2) - \mathcal{A}(\mathcal{P})$ -measurable.

Remark 8 (Continuity of statistical functionals). Apart from mere measurability, continuity or even differentiability properties of statistical functionals are of interest in statistics, finance and econometrics. A statistical functional is weakly continuous at $\mathbb{P} \in \mathcal{P}$ if the weak convergence $\mathbb{P}_n \Rightarrow \mathbb{P}$ of a sequence (\mathbb{P}_n) in \mathcal{P} to \mathbb{P} implies that $T(\mathbb{P}_n) \rightarrow T(\mathbb{P})$. Let $\mathcal{B}(\mathcal{P})$ denote the Borel- σ -algebra generated by the weak topology on \mathcal{P} . If a statistical functional T is weakly continuous at every $\mathbb{P} \in \mathcal{P}$ it is also $\mathcal{B}(\mathcal{P})$ -measurable. Gaudard and Hadwin (1989) show that $\mathcal{A}(\mathcal{P}) \subseteq \mathcal{B}(\mathcal{P})$ with equality if \mathcal{O} is a separable metric space and \mathcal{O} is its Borel σ -algebra. For these practically relevant cases, in particular for $\mathcal{O} = \mathbb{R}$, weak continuity of T readily implies $\mathcal{A}(\mathcal{P})$ -measurability. However, apart from the Range Value at Risk the functionals discussed above are known not to be weakly continuous at every \mathbb{P} under the general assumptions that we impose: For the mean, the discontinuity follows from Huber and Ronchetti (2009, Lemma 2.1), while the lower α -quantile, $\alpha \in (0, 1)$, is weakly continuous at F if and only if the quantile function $\gamma \mapsto q_\gamma^-(F)$ is continuous at α , that is, if and only if $q_\alpha^-(F) = q_\alpha^+(F)$; see van der Vaart (1998, Lemma 21.2).

In the finance literature, statistical functionals arise from law-determined (often called law-invariant) risk measures and are often called risk functionals (Krättschmer, Schied and Zähle, 2014). Assuming monotonicity with respect to first order stochastic dominance and translation equivariance, Weber (2006, Lemma 2.1 and Corollary 2.1) establishes the weak continuity of risk functionals on the class of probability distributions with compact support on \mathbb{R} , implying the $\mathcal{A}(\mathcal{P})$ -measurability. While this applies to lower quantiles and weighted averages of lower quantiles with weight function w satisfying $\int w(\gamma) d\gamma = 1$, we do not require compactly supported distributions for our results.

Generalising the class of distributions, Krättschmer, Schied and Zähle (2014) show that monotone and translation equivariant risk functionals on \mathbb{R} arising from convex risk measures on $\mathcal{P} \subseteq \mathcal{Q}^p = \{\mathbb{P} \in \mathcal{Q} \mid \int |x|^p d\mathbb{P}(x) < \infty\}$ are continuous in the Wasserstein metric of order $p \in [1, \infty)$. Similarly, Kiesel et al. (2016) show the continuity of L -functionals (see Example 5) in the Wasserstein metric of order $p \in [1, \infty)$. Since the topology generated by the Wasserstein metric of order p is larger than the weak topology, this continuity result does not imply the $\mathcal{A}(\mathcal{P})$ -measurability of L -functionals established in Example 5.

4. Elicitable functionals

Let us now turn to general elicitable functionals. These are functionals such that the ideal forecast corresponds to a Bayes act in the sense that it minimises an expected loss function. Again, let $O \subseteq \mathbb{R}$ be a measurable subset, let A be an interval, and let \mathcal{P} be a family of Borel probability measures on (O, \mathcal{O}) . Let

$$L : A \times O \rightarrow \mathbb{R}$$

be a scoring or loss function, which assigns a loss $L(a, y) \in \mathbb{R}$ to a point forecast $a \in A$ if the observation $y \in O$ materialises. We impose the following properties on L :

1. For each $a \in A$ and $P \in \mathcal{P}$, $L(a, \cdot)$ is P -integrable. Set

$$\bar{L}(a, P) = \int_O L(a, y) dP(y).$$

2. For each $P \in \mathcal{P}$, the function $a \mapsto \bar{L}(a, P)$, $a \in A$,
 - (a) is continuous,
 - (b) has a compact, non-empty interval of minimisers, denoted by $I_L(P)$,
 - (c) is monotonically decreasing to the left of $I_L(P)$, and monotonically increasing to the right.

Then L is \mathcal{P} -consistent for the functionals $T_{\min}(P) = \min I_L(P)$ and $T_{\max}(P) = \max I_L(P)$ meaning that $T_{\min}(P)$ and $T_{\max}(P)$ minimise the expected loss function $a \mapsto \bar{L}(a, P)$ for each $P \in \mathcal{P}$. If $I_L(P)$ is a singleton for each $P \in \mathcal{P}$, then $T_{\min} = T_{\max} =: T$, and then L is strictly \mathcal{P} -consistent for T , meaning that T uniquely minimises the expected loss. A functional is *elicitable* on \mathcal{P} if there exists a strictly \mathcal{P} -consistent scoring function for it.

If the class \mathcal{P} is convex and if the continuity assumption (a) on the expected loss holds, then the strict \mathcal{P} -consistency of L for $T_{\min} = T_{\max}$ already implies properties (b) and (c) as shown in Fissler and Ziegel (2019, Proposition 2.2), Nau (1985, Proposition 3), Bellini and Bignozzi (2015, Proposition 3.4), Lambert (2019, Proposition 11); see also Steinwart et al. (2014, Theorem 5) for a related result. Properties (b) and (c) are known as *order-sensitivity* or *accuracy-rewarding* in the literature; see Lambert, Pennock and Shoham (2008) and the references above. As noted in the discussion below Proposition 2.2 in Fissler and Ziegel (2019), a sufficient condition for the continuity of the expected loss $a \mapsto \bar{L}(a, P)$ for each $P \in \mathcal{P}$ are conditions (a), (b) and (c) on the level of the loss $a \mapsto L(a, y)$ itself for each $y \in O$. The continuity of $L(a, y)$ in its first argument is a common regularity condition; see e.g. Gneiting (2011).

Define the *Bayes risk* (sometimes also called *entropy* or *uncertainty*) of the \mathcal{P} -consistent loss L for T as $T_{\text{risk}}(P) = \bar{L}(T(P), P)$, $P \in \mathcal{P}$.

Theorem 9. *Under the above assumptions, the functionals T_{\min} , T_{\max} and T_{risk} are $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\mathbb{R})$ -measurable.*

Proof. First consider T_{\min} . By Example 3, for each $a \in A$ the functional $P \mapsto \bar{L}(a, P)$ is $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\mathbb{R})$ -measurable. We then argue similarly as in Example 4: Let $a \in A$ and $P \in \mathcal{P}$. We claim that under our assumptions, $T_{\min}(P) > a$ holds if and only if there exists $b \in A \cap \mathbb{Q}$ with $b > a$ such that $\bar{L}(a, P) > \bar{L}(b, P)$. Indeed, by definition of T_{\min} , $T_{\min}(P) > a$ holds if and only if a is to the left of $I_L(P)$. By monotonicity and continuity of $\bar{L}(\cdot, P)$, this holds if and only if there exists $b \in (a, \infty) \cap \mathbb{Q} \cap A$ such that $\bar{L}(a, P) > \bar{L}(b, P)$. Therefore,

$$\{P \in \mathcal{P} \mid T_{\min}(P) > a\} = \bigcup_{b \in (a, \infty) \cap \mathbb{Q} \cap A} \{P \in \mathcal{P} \mid \bar{L}(a, P) - \bar{L}(b, P) > 0\},$$

showing measurability of $T_{\min}(P)$. The arguments for T_{\max} work similarly.

For T_{risk} , we obtain the measurability upon noting that for any $x \in \mathbb{R}$

$$\{P \in \mathcal{P} \mid T_{\text{risk}}(P) < x\} = \bigcup_{a \in \mathbb{Q} \cap A} \{P \in \mathcal{P} \mid \bar{L}(a, P) < x\}. \quad \square$$

Example 10 (Expectiles and generalised quantiles). On the class of distributions with finite mean, Newey and Powell (1987) introduce the τ -expectile of a distribution P , $e_\tau(P)$, $\tau \in (0, 1)$, as the unique solution to the equation

$$\tau \int_{(x, \infty)} (y - x) dP(y) = (1 - \tau) \int_{(-\infty, x]} (x - y) dP(y)$$

in x . On the class of square-integrable distributions, e_τ is elicitable with the asymmetric piecewise quadratic loss $L_\tau(a, y) = |\mathbb{1}\{y \leq a\} - \tau|(a - y)^2$ as a strictly consistent loss. Hence, Theorem 9 establishes the measurability of e_τ on this class. For the measurability on the class of integrable distributions, one can consider the slightly modified loss function $\tilde{L}_\tau(a, y) = L_\tau(a, y) - L_\tau(0, y)$. This result complements the results of Bellini, Bigozzi and Puccetti (2018) establishing the existence and uniqueness of conditional expectiles.

Bellini et al. (2014) study *generalised* τ -quantiles which arise as minimisers of the expected loss $L_{\tau, \phi_1, \phi_2}(a, y) = \mathbb{1}\{y \leq a\}(1 - \tau)\phi_1(|a - y|) + \mathbb{1}\{y > a\}\tau\phi_2(|a - y|)$ for two convex and strictly increasing functions $\phi_1, \phi_2 : [0, \infty) \rightarrow [0, \infty)$. Clearly, if ϕ_1 and ϕ_2 are both the absolute function (the squared function) the τ -quantile (τ -expectile) arises. Elliott, Timmermann and Komunjer (2005) consider the situation of general power functions. Theorem 9 then establishes the $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\mathbb{R})$ -measurability of these generalised τ -quantiles.

Example 11 (Variance and Expected Shortfall). On the class of square-integrable distributions \mathcal{P} , the squared loss $L(a, y) = (a - y)^2$ is a strictly \mathcal{P} -consistent loss for the mean functional. The corresponding entropy functional T_{risk} is the variance functional. Therefore, Theorem 9 establishes the $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\mathbb{R})$ -measurability of the variance as an alternative argument to the direct application of Example 3.

For the lower or the upper Expected Shortfall at level α it is well known (Embrechts and Wang, 2015, Lemmas 2.3 and 3.3) that both versions can be written as the entropy of a consistent loss function for q_α^- . Therefore, Theorem 9

establishes an alternative argument for the measurability to that provided in Example 5.

An alternative approach to establish the measurability of a functional is to exploit identifiability instead of elicibility. A functional is called identifiable if there exists a moment or identification function $V: A \times O \rightarrow \mathbb{R}$ such that the functional is the unique zero of the expected identification function. Indeed, under some technical assumptions, elicibility and identifiability are equivalent for one-dimensional functionals (Steinwart et al., 2014, Theorem 5), and concerning measurability, one ends up with similar results and examples as considered in this section.

Remark 12. Proposition 3.7 in Bellini and Bignozzi (2015) establishes ψ -weak continuity of elicitable functionals on classes of probability measures with compact support on \mathbb{R} under additional regularity conditions on the corresponding strictly consistent loss function. ψ -weak continuity has been discussed in Krätschmer, Schied and Zähle (2014) and generalises continuity in the Wasserstein metric. Since the topology generated by ψ -weak convergence is larger than the one generated by weak convergence, this continuity result does not imply the $\mathcal{A}(\mathcal{P})$ -measurability of elicitable functionals, which we establish directly.

Remark 13. For an elicitable target functional $T : \mathcal{P} \rightarrow \mathbb{R}$ with strictly \mathcal{P} -consistent loss function L , an alternative way to define the ideal forecast for Y based on \mathcal{F} is to set

$$T(Y|\mathcal{F}) := \arg \min_{Z \in \mathcal{F}} \mathbb{E}[L(Z, Y)]. \tag{4}$$

This has been done in Bellini, Bignozzi and Puccetti (2018) for T being a τ -expectile. In contrast to our concatenation approach (1), the approach (4) ensures \mathcal{F} -measurability of $T(Y|\mathcal{F})$ by definition, while the existence needs to be established. The advantage of the concatenation approach is that, first, (1) also works if T is not elicitable. Second, it can be easily used to define general, not necessarily ideal forecasts for T based on \mathcal{F} as outlined in (2).

5. Measurability of scoring rules

As outlined in the introduction, the \mathcal{F} -measurability of the ideal probabilistic forecast for Y based on \mathcal{F} is guaranteed since this is the conditional distribution $\mathbb{P}_{Y|\mathcal{F}}$. In contrast to Section 4, we shall focus here on the measurability of *scoring rules*, which are the analogon to loss functions and which are used to compare probabilistic forecasts. Formally, a scoring rule is a map

$$S : \mathcal{P} \times O \rightarrow (-\infty, \infty], \tag{5}$$

assigning the penalty $S(P, y) \in (-\infty, \infty]$ to a probabilistic forecast $P \in \mathcal{P}$ if $y \in O$ is observed. We assume that for each $P \in \mathcal{P}$, $S(P, \cdot)$ is Borel-measurable and that for each $Q \in \mathcal{P}$,

$$\mathbb{E}_Q[S(P, Y)] = \int_O S(P, y) dQ(y)$$

exists. Note that we allow ∞ in the range of S in (5) to treat the common logarithmic score, see Example 17. The score S is called a strictly \mathcal{P} -proper scoring rule if

$$\mathbb{E}_Q[S(Q, Y)] < \mathbb{E}_Q[S(P, Y)] \quad \text{for all } P, Q \in \mathcal{P}, P \neq Q,$$

meaning that forecasting the true distribution Q uniquely minimises the expected score.

Recall that the ideal probabilistic forecast based on an information set \mathcal{F} is the conditional distribution $\mathbb{P}_{Y|\mathcal{F}}$. That means in a forecasting situation, with available information \mathcal{F} (e.g. given by regressors), it is natural that the forecasts have the form of Markov kernels from (Ω, \mathcal{F}) to (O, \mathcal{O}) . Inserting a Markov kernel κ for P leads to the map

$$\omega \mapsto S(\kappa(\omega, \cdot), Y(\omega)), \quad (6)$$

and averages of such expressions, which estimate expected values, are used when comparing different forecasts. But in order to form the expected value in (6) we in particular require measurability, i.e. we need that (6) is an $\mathcal{A} - \mathcal{B}(\mathbb{R})$ -measurable random variable. Since $\omega \mapsto (\kappa(\omega, \cdot), Y(\omega))$ is $\mathcal{A} - \mathcal{A}(\mathcal{P}) \otimes \mathcal{O}$ -measurable, this will be guaranteed if S in (5) is $\mathcal{A}(\mathcal{P}) \otimes \mathcal{O} - \mathcal{B}(\mathbb{R})$ -measurable. Note that the latter measurability is assumed in Nielsen (2022), whereas the following lemma provides primitive conditions for it.

Lemma 14. *Suppose that O is a one-dimensional interval, and that*

1. *for each $y \in O$, the map $S(\cdot, y) : \mathcal{P} \rightarrow (-\infty, \infty]$ is $\mathcal{A}(\mathcal{P}) - \mathcal{B}(\mathbb{R})$ -measurable,*
2. *for each $P \in \mathcal{P}$, the map $S(P, \cdot)$ is right-continuous (or left-continuous).*

Then S in (5) is $\mathcal{A}(\mathcal{P}) \otimes \mathcal{O} - \mathcal{B}(\mathbb{R})$ -measurable.

Proof. For a finite interval, say $O = (0, 1]$, for $n \in \mathbb{N}$, by (i) the map

$$S_n(P, y) = \sum_{k=1}^{2^n} S(P, k/2^n) \mathbb{1}_{((k-1)/2^n, k/2^n]}(y)$$

is $\mathcal{A}(\mathcal{P}) \otimes \mathcal{O} - \mathcal{B}(\mathbb{R})$ -measurable and by (ii), $S_n(P, y) \rightarrow S(P, y)$ for $n \rightarrow \infty$. An infinite interval can be treated analogously by extending the sum to an infinite one. \square

Remark 15. The argument in Lemma 14 can be used to obtain a similar result for multivariate observations, where right-continuity is defined, as for multivariate distribution functions, by taking limits from above in each coordinate. However, the form of the observation domain O is also an issue, and to proceed as above we require that O is a possibly infinite hyperrectangle.

Example 16 (Continuous ranked probability score). The continuous ranked probability score (CRPS) of a distribution F with finite first moment is

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(z) - \mathbb{1}\{y \leq z\})^2 dz, \quad F \in \mathcal{P}, \quad (7)$$

see Gneiting and Raftery (2007). It has an alternative representation in terms of an integrated quantile loss, and also as

$$\text{CRPS}(F, y) = \mathbb{E}_F[|y - X|] - \frac{1}{2} \mathbb{E}_F[|X' - X|], \quad (8)$$

where X and X' are independent copies with distribution F . From both (8) and (7) one can readily see that $\text{CRPS}(F, y)$ is continuous as a function of y , invoking the dominated convergence theorem. Further, for fixed y , (8) implies measurability of $\text{CRPS}(F, y)$ as a map in F , while from (7) the measurability can be obtained with a Riemann sum approximation similarly as in Example 5. Similar arguments can be made and hence the measurability is obtained for weighted versions of the CRPS as introduced in Gneiting and Ranjan (2011) and Holzmann and Klar (2017).

Example 17 (Logarithmic score and local scoring rules). Assume that \mathcal{P} is a family of right-continuous Lebesgue densities on the real line. We write $p \in \mathcal{P}$ and denote the distribution function associated with p by F_p . The logarithmic score is

$$\text{LOGS}(p, y) = -\log p(y), \quad p \in \mathcal{P},$$

see Good (1952) and Gneiting and Raftery (2007).

To apply Lemma 14, observe that $\text{LOGS}(p, y)$ is right-continuous as a function of y by our assumption on p and continuity of the logarithm. For the measurability in p for fixed y , note that by right-continuity,

$$\lim_{n \rightarrow \infty} \log \left(n (F_p(y + 1/n) - F_p(y)) \right) = \log p(y),$$

showing measurability, since the evaluation of the distribution function is measurable. Measurability extends to weighted versions of the logarithmic score as in Diks, Panchenko and van Dijk (2011) and in Holzmann and Klar (2017), and under appropriate assumptions on higher derivatives to higher-order local scoring rules (Ehm and Gneiting, 2012, Parry, Dawid and Lauritzen, 2012) such as the Hyvärinen score (Hyvärinen, 2005).

Acknowledgments

We would like to thank Johannes Resin for valuable feedback on an earlier version of the paper. We are grateful to two anonymous referees for constructive comments and for pointing out important references.

References

- [1] ADRIAN, T. and BRUNNERMEIER, M. K. (2016). CoVaR. *Am. Econ. Rev.* **106** 1705–1741.
- [2] BELLINI, F. and BIGNOZZI, V. (2015). On elicitable risk measures. *Quant. Finance* **15** 725–733. [MR3334566](#)

- [3] BELLINI, F., BIGNOZZI, V. and PUCCETTI, G. (2018). Conditional expectiles, time consistency and mixture convexity properties. *Insurance Math. Econom.* **82** 117–123. [MR3850612](#)
- [4] BELLINI, F., KLAR, B., MÜLLER, A. and ROSAZZA GIANIN, E. (2014). Generalized quantiles as risk measures. *Insurance Math. Econom.* **54** 41–48. [MR3145849](#)
- [5] BORGONOVO, E., HAZEN, G. B., JOSE, V. R. R. and PLISCHKE, E. (2021). Probabilistic sensitivity measures as information value. *Eur. J. Oper. Res.* **289** 595–610. [MR4170762](#)
- [6] CONT, R., DEGUEST, R. and SCANDOLO, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quant. Finance* **10** 593–606. [MR2676786](#)
- [7] DAWID, A. P. (1984). Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach. *J. Roy. Statist. Soc. Ser. A* **147** 278–292. [MR0763811](#)
- [8] DE CASTRO, L. I., COSTA, B. N., GALVAO, A. F. and ZUBELLI, J. (2021). Conditional Quantiles: An Operator-Theoretical Approach. *Preprint*. <https://dx.doi.org/10.2139/ssrn.3924597>.
- [9] DIEBOLD, F. X. and MARIANO, R. S. (1995). Comparing predictive accuracy. *J. Bus. Econom. Statist.* **13** 253–263. [MR3303732](#)
- [10] DIKS, C., PANCHENKO, V. and VAN DIJK, D. (2011). Likelihood-based scoring rules for comparing density forecasts in tails. *J. Econom.* **163** 215–230. [MR2812867](#)
- [11] DIMITRIADIS, T. and BAYER, S. (2019). A joint quantile and expected shortfall regression framework. *Electron. J. Statist.* **13** 1823–1871. [MR3959874](#)
- [12] EHM, W. and GNEITING, T. (2012). Local proper scoring rules of order two. *Ann. Statist.* **40** 609–637. [MR3014319](#)
- [13] ELLIOTT, G., TIMMERMANN, A. and KOMUNJER, I. (2005). Estimation and Testing of Forecast Rationality under Flexible Loss. *Rev. Econ. Stud.* **72** 1107–1125. [MR2176557](#)
- [14] EMBRECHTS, P. and WANG, R. (2015). Seven proofs for the subadditivity of expected shortfall. *Depend. Model.* **3** 126–140. [MR3418661](#)
- [15] FISSLER, T. and HOGA, Y. (2021). Backtesting Systemic Risk Forecasts using Multi-Objective Elicitability. *Preprint* version 4. <https://arxiv.org/abs/2104.10673v4>.
- [16] FISSLER, T. and PESENTI, S. M. (2022). Sensitivity measures based on scoring functions. *Preprint*. <https://doi.org/10.48550/arXiv.2203.00460>.
- [17] FISSLER, T. and ZIEGEL, J. F. (2019). Order-Sensitivity and Equivariance of Scoring Functions. *Electron. J. Stat.* **13** 1166–1211. [MR3935847](#)
- [18] FISSLER, T. and ZIEGEL, J. F. (2021). On the elicibility of range value at risk. *Stat. Risk Model.* **38** 25–46. [MR4343188](#)
- [19] FISSLER, T., FRONGILLO, R., HLAVINOVÁ, J. and RUDLOFF, B. (2021). Forecast evaluation of quantiles, prediction intervals, and other set-valued functionals. *Electron. J. Statist.* **15** 1034–1084. [MR4255295](#)

- [20] GAUDARD, M. and HADWIN, D. (1989). Sigma-Algebras on Spaces of Probability Measures. *Scand. J. Stat.* **16** 169–175. [MR1028976](#)
- [21] GIRARDI, G. and TOLGA ERGÜN, A. (2013). Systemic Risk Measurement: Multivariate GARCH Estimation of CoVaR. *J. Bank. Financ.* **37** 3169–3180.
- [22] GNEITING, T. (2011). Making and Evaluating Point Forecasts. *J. Amer. Statist. Assoc.* **106** 746–762. [MR2847988](#)
- [23] GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 243–268. [MR2325275](#)
- [24] GNEITING, T. and KATZFUSS, M. (2014). Probabilistic Forecasting. *Annu. Rev. Stat. Appl.* **1** 125–151.
- [25] GNEITING, T. and RAFTERY, A. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Amer. Statist. Assoc.* **102** 359–378. [MR2345548](#)
- [26] GNEITING, T. and RANJAN, R. (2011). Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules. *J. Bus. Econom. Statist.* **29** 411–422. [MR2848512](#)
- [27] GNEITING, T. and RANJAN, R. (2013). Combining predictive distributions. *Electron. J. Statist.* 1747–1782. [MR3080409](#)
- [28] GNEITING, T. and RESIN, J. (2021). Regression Diagnostics meets Forecast Evaluation: Conditional Calibration, Reliability Diagrams, and Coefficient of Determination. *Preprint*. <https://doi.org/10.48550/arXiv.2108.03210>.
- [29] GOOD, I. J. (1952). Rational Decisions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **14** 107–114. [MR0077033](#)
- [30] HOGA, Y. and DIMITRIADIS, T. (2022). On Testing Equal Conditional Predictive Ability Under Measurement Error. *J. Bus. Econom. Statist.* **0** 1–13.
- [31] HOLZMANN, H. and EULERT, M. (2014). The role of the information set for forecasting – with applications to risk management. *Ann. Appl. Stat.* **8** 79–83. [MR3192004](#)
- [32] HOLZMANN, H. and KLAR, B. (2017). Focusing on regions of interest in forecast evaluation. *Ann. Appl. Stat.* **11** 2404–2431. [MR3743302](#)
- [33] HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, Second ed. John Wiley & Sons, Inc., Hoboken, New Jersey. [MR2488795](#)
- [34] HYVÄRINEN, A. (2005). Estimation of Non-Normalized Statistical Models by Score Matching. *J. Mach. Learn. Res.* **6** 695–709. [MR2249836](#)
- [35] KALLENBERG, O. (2002). *Foundations of modern probability*, second ed. *Probability and its applications*. Springer, New York, NY. [MR1876169](#)
- [36] KIESEL, R., RÜHLICKE, R., STAHL, G. and ZHENG, J. (2016). The Wasserstein Metric and Robustness in Risk Management. *Risks* **4** 1–14.
- [37] KOENKER, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge. [MR2268657](#)
- [38] KRÄTSCHMER, V., SCHIED, A. and ZÄHLE, H. (2014). Comparative and qualitative robustness for law-invariant risk measures. *Finance Stoch.* **18**

- 271–295. [MR3177407](#)
- [39] LAMBERT, N. (2019). Elicitation and Evaluation of Statistical Functionals. *Preprint*. <http://ai.stanford.edu/~nlambert/papers/elicitability.pdf>.
- [40] LAMBERT, N., PENNOCK, D. M. and SHOHAM, Y. (2008). Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce* 129–138. ACM, Chicago, IL, USA.
- [41] MOLCHANOV, I. (2017). *Theory of Random Sets*, 2 ed. *Probability Theory and Stochastic Modelling*. Springer-Verlag London, London. [MR3751326](#)
- [42] NAU, R. F. (1985). Should Scoring Rules Be ‘Effective’? *Manag. Sci.* **31** 527–535.
- [43] NEWEY, W. K. and POWELL, J. L. (1987). Asymmetric Least Squares Estimation and Testing. *Econometrica* **55** 819–847. [MR0906565](#)
- [44] NIELSEN, M. (2022). Regular conditional probabilities and strictly proper loss functions. *Statist. Probab. Lett.* **185** 1–5. [MR4383205](#)
- [45] NOLDE, N. and ZIEGEL, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *Ann. Appl. Stat.* **11** 1833–1874. [MR3743276](#)
- [46] PARRY, M., DAWID, A. P. and LAURITZEN, S. (2012). Proper local scoring rules. *Ann. Statist.* **40** 561–592. [MR3014317](#)
- [47] POHLE, M. O. (2020). The Murphy Decomposition and the Calibration-Resolution Principle: A New Perspective on Forecast Evaluation. *Preprint*. <https://doi.org/10.48550/arXiv.2005.01835>.
- [48] STEINWART, I., PASIN, C., WILLIAMSON, R. and ZHANG, S. (2014). Elicitation and Identification of Properties. *JMLR Workshop Conf. Proc.* **35** 1–45.
- [49] VAN DER VAART, A. W. (1998). *Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press. [MR1652247](#)
- [50] WEBER, S. (2006). Distribution-Invariant Risk Measures, Information, and Dynamic Consistency. *Math. Finance* **16** 419–441. [MR2212272](#)