

Forecast evaluation of quantiles, prediction intervals, and other set-valued functionals

Fissler, Tobias; Frongillo, Rafael; Hlavinova, Jana; Rudloff, Birgit

Published in:
Electronic Journal of Statistics

DOI:
[10.1214/21-EJS1808](https://doi.org/10.1214/21-EJS1808)

Published: 01/01/2021

Document Version
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Fissler, T., Frongillo, R., Hlavinova, J., & Rudloff, B. (2021). Forecast evaluation of quantiles, prediction intervals, and other set-valued functionals. *Electronic Journal of Statistics*, 15(1), 1034 - 1084. <https://doi.org/10.1214/21-EJS1808>

Forecast evaluation of quantiles, prediction intervals, and other set-valued functionals*

Tobias Fissler¹, Rafael Frongillo², Jana Hlavinová¹
and Birgit Rudloff¹

¹Vienna University of Economics and Business (WU), Institute for Statistics and
Mathematics, Welthandelsplatz 1, 1020 Vienna, Austria,
e-mail: tobias.fissler@wu.ac.at; jana.hlavinova@wu.ac.at; birgit.rudloff@wu.ac.at

²University of Colorado Boulder, Department of Computer Science, 430 UCB, 1111
Engineering Dr, Boulder, CO 80309, e-mail: raf@colorado.edu

Abstract: We introduce a theoretical framework of elicibility and identifiability of set-valued functionals, such as quantiles, prediction intervals, and systemic risk measures. A functional is elicitable if it is the unique minimiser of an expected scoring function, and identifiable if it is the unique zero of an expected identification function; both notions are essential for forecast ranking and validation, and M - and Z -estimation. Our framework distinguishes between *exhaustive* forecasts, being set-valued and aiming at correctly specifying the entire functional, and *selective* forecasts, content with solely specifying a single point in the correct functional. We establish a mutual exclusivity result: A set-valued functional can be either selectively elicitable or exhaustively elicitable or not elicitable at all. Notably, since quantiles are well known to be selectively elicitable, they fail to be exhaustively elicitable. We further show that the classes of prediction intervals and Vorob'ev quantiles turn out to be exhaustively elicitable, hence not selectively elicitable, but still selectively identifiable. In particular, we provide a mixture representation of elementary exhaustive scores, leading the way to Murphy diagrams. We establish that the shortest prediction interval and those specified by an endpoint or midpoint in general fail to be elicitable with respect to either notion, unless an endpoint is given via a quantile. We end with a comprehensive literature review on common practice in forecast evaluation of set-valued functionals.

MSC2020 subject classifications: 62C05, 62F07, 91B06; secondary 62H11.

Keywords and phrases: Consistency, convex level sets, elicibility, identifiability, M -estimation, prediction intervals, random sets, Vorob'ev quantiles.

Received August 2020.

*T. Fissler received financial support via his Chapman Fellowship from Imperial College London for parts of this work. R. Frongillo was supported in part by the U.S. National Science Foundation under Grant No. 1657598. B. Rudloff acknowledges support from the OeNB anniversary fund, project number 17793.

Contents

1	Introduction	1035
2	Two types of elicibility and identifiability	1038
2.1	Scoring and identification functions for single-valued functionals	1038
2.2	Selective and exhaustive scoring and identification functions	1040
2.3	Practical implications on M -estimation and forecast comparison	1042
3	Structural results	1043
3.1	CxLS properties and their implications	1043
3.2	Mutual exclusivity	1046
3.3	Specifications of set-valued functionals	1049
4	Prediction intervals	1050
4.1	Notation	1052
4.2	Elicibility and identifiability of the class of α -prediction intervals	1054
4.3	Prediction interval with an endpoint or the midpoint given by an identifiable functional	1058
4.4	Shortest prediction intervals	1061
5	Vorob'ev quantiles	1065
6	Connections to forecast evaluation in the literature	1070
6.1	Statistical forecast evaluation	1070
6.2	Statistical theory and risk measurement	1071
6.3	Spatial statistics	1071
6.4	Regression and machine learning	1072
6.5	Philosophy	1072
A	Prediction intervals with a fixed endpoint or midpoint	1073
B	Injectivity results for prediction interval variants	1074
C	Omitted technical proofs	1077
	Acknowledgements	1080
	References	1080

1. Introduction

Decision making in the presence of uncertainty is a regular task faced by many stakeholders in society. This uncertainty may have many sources, one of which stems from unknown or random future events. For example, in agriculture the right time for harvesting depends on the weather in the forthcoming days; in business, an investment decision in a production facility depends on future demand; in politics, decisions for travel restrictions or lockdowns heavily depend on the anticipated future development of a pandemic. Predicting uncertain future events is therefore urgent and ubiquitous.

The presence of various sorts of forecasts and society's reliance on them calls for a careful assessment and evaluation, basically focusing on two complementary aspects: First, are given forecasts good or reliable in absolute terms? And second, how well have certain forecasts performed in comparison to some alternative predictions, assessing their relative quality? Clearly, these questions can only be answered *ex post*, given observations of the future events in question. Then, the reliability or calibration can be assessed in terms of *moment* or *identification functions*. Forecast comparison and ranking, in turn, commonly employ loss or scoring functions.

To perform forecast evaluation properly, one must specify a certain quality criterion or directive the forecasts should follow. This directive might be given indirectly in terms of a cost, loss, or score, such that good forecasts aim to minimise this criterion (in expectation). This directive can also be formulated directly, e.g. as the whole probability distribution of the uncertain event, capturing the entire inherent uncertainty, or as a summary measure thereof, called a *functional*, such as the mean, variance, or a certain risk measure. When the directive is formulated directly, it is crucial that the tools of forecast evaluation, chiefly scoring and identification functions, are in line with this directive. This alignment leads to the notions of strictly consistent scores, which are minimised in expectation by the correctly specified forecasts, and strict identification functions, whose roots in expectation are the correctly specified forecasts.

The literature on forecast evaluation has mainly focused on single-valued functionals, such as real-valued and vector-valued point forecasts or probabilistic forecasts, where a single correct value is specified for each possible distribution (for technical definitions and an account of the literature, we refer to Section 2.1). Yet set-valued functionals are abundant; prominent examples include quantiles (see also option (ii) of the interesting discussion provided in Mizera (2010, p. 170)), the mode, and prediction intervals, which may all be non-unique and therefore set-valued. Applications bring many other examples, such as the set-valued systemic risk measures introduced in Feinstein, Rudloff and Weber (2017) which specify the entire set of capital allocations adequate to render a financial system's risk acceptable. Set-valued functionals also naturally arise via expectations or quantiles of random sets (Molchanov, 2017), such as from climatology and meteorology (the area affected by a flood), reliability engineering (parts of a machine being affected by extreme heat) or medicine (tumorous tissue in the human body); cf. Bolin and Lindgren (2015). We discuss these applications and several others in Section 6 with a special focus on quantiles of random sets in Section 5.

Techniques developed to evaluate single-valued forecasts do not in general suffice for set-valued functionals. For functionals like the mode or the α -quantile, it is common to restrict to the set of distributions with a unique mode or α -quantile (Heinrich, 2014; Fissler and Ziegel, 2019), yet one may be interested in distributions with multiple modes or quantiles. Moreover, many functionals, such as quantiles of random sets, are inherently set-valued and such a restriction is not available, or too much of a simplification. E.g. Buansing, Golan and Ullah (2020) consider interval-valued forecasts for interval-valued observations,

proposing several different loss functions to evaluate their forecasts without providing a discussion of which one is best suited for the situation at hand. We therefore see the need for a comprehensive theoretical framework for the evaluation of forecasts for set-valued functionals.

It turns out that already for the definition of elicitation (or identification) of a set-valued functional, there are several possibilities, depending on the form the forecasts take. One may ask for an arbitrary element of the functional, an arbitrary subset, or perhaps even the entire set itself. For the case of the α -prediction interval, and the uniform distribution on $[0, 1]$, these definitions correspond to any subinterval of $[0, 1]$ of length at least α , any set of such intervals, or the entire set of all such intervals. Moreover, if a set-valued functional is elicitable in one of these corresponding senses, does it continue to be elicitable if one specifies a particular element of the functional to be elicited, such as the shortest α -prediction interval; or if it is not elicitable in one of the senses above, could such a specification render it elicitable?

In this paper, we present a general theoretical framework to evaluate the forecasts of set-valued functionals, which clarifies and expands upon these questions. We begin with a thorough definition of elicibility and identifiability of set-valued functionals (Section 2). In particular, as alluded to in the above questions, we define two types of set-valued elicitation (identification). For the *selective* type, we follow Lambert and Shoham (2009) and Gneiting (2011a) where a single-valued forecast must be among the set of correct values specified by the functional, as is also typical for quantiles (Koenker, 2005). In contrast, the *exhaustive* type more ambitiously asks one to forecast the entire set of correct values, and requires this set to be the unique minimiser (zero) of the expected score (identification function). We comment on the implications of the two different modes of forecasts and of elicibility on M -estimation and on forecast comparison in Subsection 2.3.

The main result of this article, Theorem 3.9, states that the two types of elicibility are mutually exclusive: a set-valued functional is either selectively elicitable or exhaustively elicitable, or not elicitable at all, subject to mild regularity conditions. The proof follows from a refinement of the classical result (Proposition 3.4) that elicitable functionals have Convex Level Sets (CxLS), meaning that if two distributions have the same functional value, any mixture of the two has the same functional value. This mutual exclusivity result is powerful in its ability to rule out elicibility of one type or the other. For example, it is natural to ask for the full set of α -quantiles of a random variable, yet as quantiles are known to be selectively elicitable, they must fail to be exhaustively elicitable. Interestingly, any specification of the quantile, such as the lower quantile, or Value at Risk in the risk management literature, is also not elicitable in general; see Corollary 3.15 and the discussion thereafter.

We illustrate our framework with new results for prediction intervals (Section 4) and Vorob'ev quantiles (Section 5). Prediction intervals are a relevant statistical tool that aims at issuing forecasts that account for the uncertainty about the future outcome, reporting an interval into which a given random variable will fall with an *ex ante* specified certainty or coverage $\alpha \in$

$(0, 1]$ rather than merely reporting a single point. They are used in a wide range of applications, most recently e.g. in the field of epidemiology where many forecasts of case or death counts are issued in the form of intervals, cf. <https://covid19forecasthub.org> or Bracher et al. (2020) and references therein. We show that the whole class of α -prediction intervals is exhaustively elicitable (Theorem 4.2). While this immediately rules out their selective elicibility (Corollary 4.7), we consider certain interesting specifications of prediction intervals. If the midpoint or an endpoint is given by a constant, this specification is elicitable (Appendix A). However, if the midpoint or an endpoint is given via a general identifiable functional, it is in general not elicitable, unless the endpoint is specified via a quantile (Propositions 4.10 and 4.12). We also show that the shortest prediction interval is not elicitable in either sense (Theorem 4.16). The results of Section 4 complement and generalise recent results established independently in Brehmer and Gneiting (2020). In Section 5 we establish the exhaustive elicibility and selective identifiability of Vorob'ev quantiles of random closed sets. For an application to systemic risk measures, we refer the reader to Fissler, Hlavinová and Rudloff (2021), where the theoretical framework of this paper has been applied to establish selective identifiability, exhaustive elicibility, and mixture representations of exhaustive scoring functions, leading to Murphy diagrams. We close in Section 6 with a comprehensive literature review on forecast evaluation for set-valued quantities, covering spatial statistics, machine learning, engineering, climatology and meteorology, and philosophy, leaving many interesting avenues for future research. Technical proofs and further interesting results are deferred to Appendix C and B, respectively.

2. Two types of elicibility and identifiability

2.1. Scoring and identification functions for single-valued functionals

We use the decision-theoretic framework described for example in Gneiting (2011a); cf. Savage (1971); Osband (1985); Lambert, Pennock and Shoham (2008); Fissler and Ziegel (2016, 2019). Let $(\Omega, \mathfrak{F}, \mathbf{P})$ be some complete, atomless probability space rich enough to accommodate all random elements mentioned in the sequel. With Y we denote an observation of interest, taking values in some measurable space $(\mathcal{O}, \mathcal{O})$, called *observation domain*. Forecasts for Y are denoted by X , taking values in a measurable space $(\mathcal{A}, \mathcal{A})$ called *action domain*. We assume that the directive for an ideal forecast is given in terms of a statistical functional of the (conditional) distribution F of Y (given some σ -algebra which makes X measurable). Mathematically, this is a map $T: \mathcal{M} \rightarrow \mathcal{A}$, where \mathcal{M} is some class of probability measures or probability distribution functions on $(\mathcal{O}, \mathcal{O})$. All functions are tacitly assumed to be measurable. A scoring function is a map $S: \mathcal{A} \times \mathcal{O} \rightarrow \mathbb{R}^* := (-\infty, \infty]$. It is negatively oriented, meaning that a forecast $x \in \mathcal{A}$ receives the penalty $S(x, y)$ if $y \in \mathcal{O}$ materialises. Statistically, the relative quality of a prediction observation sequence (X_t, Y_t) , $t = 1, \dots, N$,

is evaluated by S in terms of the *realised score*

$$\frac{1}{N} \sum_{t=1}^N S(X_t, Y_t). \quad (2.1)$$

Invoking an expected utility maximisation argument or a suitable law of large numbers, it has been widely argued (Murphy and Daan, 1985; Engelberg, Manski and Williams, 2009) that a scoring function should incentivise truthful forecasts in that the Bayes-act coincides with the given directive. We say that S is incentive compatible or \mathcal{M} -consistent for T if

$$\bar{S}(T(F), F) \leq \bar{S}(x, F) \quad (2.2)$$

for all $x \in \mathbf{A}$ and $F \in \mathcal{M}$, implicitly assuming that $\bar{S}(x, F) := \int S(x, y) dF(y)$ exists. If additionally equality in (2.2) implies that $x = T(F)$, S is *strictly* \mathcal{M} -consistent for T . A functional that admits a strictly consistent scoring function is called *elicitable* (Osband, 1985; Lambert, Pennock and Shoham, 2008). As such, the elicibility of a functional opens the way to meaningful forecast comparison (Gneiting, 2011a) which is closely related to comparative backtests in finance (Fissler, Ziegel and Gneiting, 2016; Nolde and Ziegel, 2017). Similarly, it is crucial for M -estimation (Huber, 1967; Huber and Ronchetti, 2009) and regression, such as quantile regression (Koenker and Basset, 1978; Koenker, 2005) or expectile regression (Newey and Powell, 1987).

While scoring functions serve the purpose of forecast comparison and ranking, we employ *identification functions* when it comes to forecast *validation*. Similarly to a scoring function, an identification function is a map $V: \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}^k$ where we again make the tacit assumption that $\bar{V}(x, F) := \int V(x, y) dF(y)$ exists for all $x \in \mathbf{A}$, $F \in \mathcal{M}$ with the additional assumption that the expectation be finite. V is an \mathcal{M} -*identification function* for T if $\bar{V}(T(F), F) = 0$. It is a *strict* \mathcal{M} -identification function for T if additionally $\bar{V}(x, F) = 0$ implies that $x = T(F)$. In the literature on point-valued functionals, it turned out to be appropriate that k coincides with the dimension of the forecasts (Osband, 1985; Frongillo and Kash, 2015; Fissler and Ziegel, 2016). Since statistical practice demands to evaluate the *realised identification function*, which is the counterpart of (2.1) upon replacing S by V , V simply needs to map to a real vector space. One can even be more flexible and use an infinite-dimensional space. E.g. in Proposition 4.20 the identification function maps to $\mathbb{R}^{[-1,1]}$, the space of functions from $[-1, 1]$ to \mathbb{R} . One can also relax the requirement that the expected identification function attains a 0 at the correctly specified forecast. It can rather attain some predefined particular value(s)—the important requirement is that this value be *identifiable* in the common sense.

In statistics and econometrics, identification functions are often called *moment functions* and give rise to the (*generalised*) *method of moments* (Newey and McFadden, 1994) or Z -estimation. For a discussion of identifiability and calibration in the context of backtesting risk measures, we refer the reader to the insightful papers Davis (2016) and Nolde and Ziegel (2017). For a recent general perspective on identification, see Basse and Bojinov (2020).

2.2. Selective and exhaustive scoring and identification functions

When T is set-valued, we may write $T: \mathcal{M} \rightarrow 2^W$, where W is some generic space. As mentioned in the Introduction, we distinguish two types of forecasts with the corresponding notions of scoring and identification functions. In decision-theoretic terms, this translates into two sensible choices for the action domain A :

- (i) $A = A_{\text{sel}} \subseteq W$: The elements of the action domain A_{sel} representing possible forecasts are *points* in the space W . Truthful reporting means that there are generally *multiple* best actions, namely all *selections* $t \in T(F) \subseteq A_{\text{sel}}$ for $F \in \mathcal{M}$. Mnemonically, we shall refer to A_{sel} as a *selective* action domain.
- (ii) $A = A_{\text{exh}} \subseteq 2^W$: The elements of the action domain A_{exh} representing possible forecasts are *subsets* of the space W . Truthful reporting means that there is a *unique* best action, namely the *exhaustive* functional $T(F) \in A_{\text{exh}}$ for $F \in \mathcal{M}$. Similarly, we shall refer to A_{exh} as an *exhaustive* action domain.

The two different choices of action domains lay claim to different levels of precision and ambition of the forecasts. For a certain functional $T: \mathcal{M} \rightarrow 2^W$, the connection between the choice of the selective action domain $A_{\text{sel}} \subseteq W$ and the exhaustive action domain $A_{\text{exh}} \subseteq 2^W$ will be specified if needed for a certain result, otherwise remaining unspecified. However, a sensible connection between the two choices we have in mind is $A_{\text{sel}} = \bigcup_{B \in A_{\text{exh}}} B$.

We continue to use the dichotomy introduced above also for scoring functions evaluating forecasts for some set-valued functional $T: \mathcal{M} \rightarrow 2^W$. Let $\mathcal{M}' \subseteq \mathcal{M}$ be some generic subset.

Definition 2.1 (Consistency, elicibility). (i) A *selective* scoring function $S_{\text{sel}}: A_{\text{sel}} \times \mathcal{O} \rightarrow \mathbb{R}^*$ is \mathcal{M}' -consistent for $T: \mathcal{M} \rightarrow 2^{A_{\text{sel}}}$ if

$$\bar{S}_{\text{sel}}(t, F) \leq \bar{S}_{\text{sel}}(x, F) \quad \forall x \in A_{\text{sel}}, \forall t \in T(F), \forall F \in \mathcal{M}'. \quad (2.3)$$

The selective score S_{sel} is *strictly \mathcal{M}' -consistent* for T if it is \mathcal{M}' -consistent for T and if equality in (2.3) implies that $x \in T(F)$. T is *selectively elicitable* on \mathcal{M}' if there is a strictly \mathcal{M}' -consistent selective scoring function for T .

- (ii) An *exhaustive* scoring function $S_{\text{exh}}: A_{\text{exh}} \times \mathcal{O} \rightarrow \mathbb{R}^*$ is \mathcal{M}' -consistent for $T: \mathcal{M} \rightarrow A_{\text{exh}}$ if

$$\bar{S}_{\text{exh}}(T(F), F) \leq \bar{S}_{\text{exh}}(B, F) \quad \forall B \in A_{\text{exh}}, \forall F \in \mathcal{M}'. \quad (2.4)$$

The exhaustive score S_{exh} is *strictly \mathcal{M}' -consistent* for T if it is \mathcal{M}' -consistent for T and if equality in (2.4) implies that $B = T(F)$. T is *exhaustively elicitable* on \mathcal{M}' if there is a strictly \mathcal{M}' -consistent exhaustive scoring function for T .

Unless mentioned explicitly otherwise, we tacitly assume that all scoring functions are \mathcal{M}' -finite in the sense that $\bar{S}_{\text{sel}}(x, F), \bar{S}_{\text{exh}}(B, F) < \infty$ for all $x \in A_{\text{sel}}, B \in A_{\text{exh}}, F \in \mathcal{M}'$.

If we merely say that $T: \mathcal{M} \rightarrow 2^W$ is (selectively or exhaustively) elicitable, we mean it is (selectively or exhaustively) elicitable on \mathcal{M} . Assuming \mathcal{M} -finiteness is convenient in many proofs and is a standard assumption in the literature (Ziegel, 2016a; Fissler and Ziegel, 2016; Brehmer and Strokov, 2019; Wang and Wei, 2020). Note that without stipulating \mathcal{M} -finiteness, the *strict* \mathcal{M} -consistency of a selective (exhaustive) scoring function S_{sel} (S_{exh}) already implies that $\bar{S}_{\text{sel}}(t, F) \in \mathbb{R}$ for all $F \in \mathcal{M}$, $t \in T(F)$ ($\bar{S}_{\text{exh}}(T(F), F) \in \mathbb{R}$ for all $F \in \mathcal{M}$). According to Gneiting and Raftery (2007) we call any two (selective or exhaustive) scoring functions $S, S': \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}$ equivalent if there is some $\lambda > 0$ and some function $a: \mathbf{O} \rightarrow \mathbb{R}$ such that $S'(x, y) = \lambda S(x, y) + a(y)$. It is immediate to see that this equivalence relation preserves \mathcal{M} -consistency, and also strict \mathcal{M} -consistency, subject to a being \mathcal{M} -integrable. If there is no risk of confusion, we shall drop the indices “sel” and “exh” to indicate the difference between selective and exhaustive interpretations, respectively.

For identification functions, we again make the distinction between selective and exhaustive identification functions to allow for a rigorous treatment of set-valued functionals.

Definition 2.2 (Identification function, identifiability). (i) A map $V_{\text{sel}}: \mathbf{A}_{\text{sel}} \times \mathbf{O} \rightarrow \mathbb{R}^k$ is a *selective \mathcal{M}' -identification function* for $T: \mathcal{M} \rightarrow 2^{\mathbf{A}_{\text{sel}}}$ if $\bar{V}_{\text{sel}}(t, F) = 0$ for all $t \in T(F)$ and for all $F \in \mathcal{M}'$. Moreover, V_{sel} is a strict selective \mathcal{M}' -identification function for T if

$$\bar{V}_{\text{sel}}(x, F) = 0 \iff x \in T(F), \quad \forall x \in \mathbf{A}_{\text{sel}}, \quad \forall F \in \mathcal{M}'. \quad (2.5)$$

T is *selectively identifiable* on \mathcal{M}' if it possesses a strict selective \mathcal{M}' -identification function.

(ii) A map $V_{\text{exh}}: \mathbf{A}_{\text{exh}} \times \mathbf{O} \rightarrow \mathbb{R}^k$ is an *exhaustive \mathcal{M}' -identification function* for $T: \mathcal{M} \rightarrow \mathbf{A}_{\text{exh}}$ if $\bar{V}_{\text{exh}}(T(F), F) = 0$ for all $F \in \mathcal{M}'$. Moreover, V_{exh} is a strict exhaustive \mathcal{M}' -identification function for T if

$$\bar{V}_{\text{exh}}(B, F) = 0 \iff B = T(F), \quad \forall B \in \mathbf{A}_{\text{exh}}, \quad \forall F \in \mathcal{M}'. \quad (2.6)$$

T is *exhaustively identifiable* on \mathcal{M}' if it possesses a strict exhaustive \mathcal{M}' -identification function.

Again, we say that $T: \mathcal{M} \rightarrow 2^W$ is (selectively or exhaustively) identifiable, if we mean that it is (selectively or exhaustively) identifiable on \mathcal{M} .

For single-valued functionals such as the mean, the distinction between selective and exhaustive elicibility is obsolete, since any choice of an action domain leads to a unique best action. Hence, one is actually always in the exhaustive setting, and there is no point in mentioning this fact explicitly. Of course, we could formally identify a point-valued functional $T': \mathcal{M} \rightarrow \mathbf{A}'$ with the set-valued functional $T: \mathcal{M} \rightarrow \mathbf{A} = \{\{a\} \mid a \in \mathbf{A}'\}$ where $T(F) = \{T'(F)\}$. Then the following lemma holds.

Lemma 2.3. *Let $T': \mathcal{M} \rightarrow \mathbf{A}'$ be some point-valued functional. Define the set-valued functional $T(F) := \{T'(F)\}$, $F \in \mathcal{M}$. Then T , considered as a map to the*

power set $2^{A'}$, is selectively elicitable (identifiable) if and only if T , considered as a map to the exhaustive action domain $A = \{\{a\} \mid a \in A'\}$, is exhaustively elicitable (identifiable). Moreover, the selective elicibility (identifiability) of $T: \mathcal{M} \rightarrow 2^{A'}$ is equivalent to the elicibility (identifiability) of T' .

While we are aware of contributions to the literature which consider either the selective or the exhaustive interpretation only (see Section 6), one novelty in the present paper is that we thoroughly study and compare these two alternative notions, which is the content of the Section 3.

2.3. Practical implications on M -estimation and forecast comparison

Concerning M -estimation, the two modes of elicibility appear equally powerful: Indeed, given an i.i.d. sample $Y_1, \dots, Y_n \sim F \in \mathcal{M}$, for a strictly consistent selective score S_{sel} or strictly consistent exhaustive score S_{exh} for $T: \mathcal{M} \rightarrow 2^W$, the corresponding M -estimators are given via

$$\hat{T}_{\text{sel}}(F) = \arg \min_{x \in W} \frac{1}{n} \sum_{i=1}^n S_{\text{sel}}(x, Y_i), \quad \hat{T}_{\text{exh}}(F) \in \arg \min_{B \subseteq W} \frac{1}{n} \sum_{i=1}^n S_{\text{exh}}(B, Y_i), \quad (2.7)$$

where the arg min on the right hand side is a singleton in the limit, due to the strict consistency of S_{exh} . From a computational angle, the minimisation over points $x \in W$ in $\hat{T}_{\text{sel}}(F)$ is practically often easier than the minimisation over all subsets $B \subseteq W$ in $\hat{T}_{\text{exh}}(F)$. In a regression context with i.i.d. explanatory variables Z_1, \dots, Z_n , the situation is slightly different. If T is selectively elicitable, one can model $T(F_{Y_i|Z_i})$ via a selective model $m(Z_i, \theta) \in W$, where any correctly specified parameter $\theta_0 \in \Theta$ fulfils $m(Z_i, \theta_0) \in T(F_{Y_i|Z_i})$, where $F_{Y_i|Z_i}$ is the conditional distribution of Y_i given Z_i . Exhaustive models, on the other hand, are set-valued, $M(Z_i, \theta) \subseteq W$. Any correct parameter $\theta_0 \in \Theta$ fulfils $M(Z_i, \theta_0) = T(F_{Y_i|Z_i})$. Once the different models have been formulated, the actual approach for estimating θ_0 is basically similar:

$$\hat{\theta}_{\text{sel}} \in \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n S_{\text{sel}}(m(Z_i, \theta), Y_i), \quad \hat{\theta}_{\text{exh}} \in \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n S_{\text{exh}}(M(Z_i, \theta), Y_i). \quad (2.8)$$

So the main consequence of the classification as selectively or exhaustively elicitable manifests in the form of the regression model, being either point-valued or set-valued.

Forecast comparison and ranking, in turn, is straightforward if the form of the forecasts corresponds to the actual mode of elicibility, that is, if we have point-valued forecasts in case of selective elicibility, or set-valued forecasts in case of exhaustive elicibility of T . If the form of the forecasts does not match the mode of elicibility, however, the situation appears virtually hopeless: If the forecasts are set-valued, B_i^j for $i = 1, \dots, n$, $j = 1, 2$, but T is

selectively elicitable, one could come up with an ad-hoc-procedure. For fixed $x_i^j \in B_i^j$ for all $i = 1, \dots, n$, $j = 1, 2$, consider the average score differences $\frac{1}{n} \sum_{i=1}^n S_{\text{sel}}(x_i^1, Y_i) - S_{\text{sel}}(x_i^2, Y_i)$. If the sign is the same for all possible choices of $x_i^j \in B_i^j$ and for all $i = 1, \dots, n$, we have a conclusive ranking. However, ignoring the computational complexity of this ad-hoc-procedure for a moment, and the statistical challenges of performing a uniform test, this procedure will generally not produce a conclusive ranking. For the corresponding situation of selective forecasts x_i^j for $i = 1, \dots, n$, $j = 1, 2$ for an exhaustively elicitable functional, an ad-hoc-procedure could consist in considering the average score differences $\frac{1}{n} \sum_{i=1}^n S_{\text{exh}}(\{x_i^1\}, Y_i) - S_{\text{exh}}(\{x_i^2\}, Y_i)$. If one can expect the true functional to be singleton-valued in relevant cases, this might be reasonable. But for situations when the true functional is a proper set, e.g. for Vorob'ev quantiles of random sets (see Section 5), this procedure will not produce reasonable results.

The implications of the two modes of identifiability on Z -estimation and forecast validation are similar. Notably, only the ad-hoc procedure of validating exhaustive forecasts B_i , $i = 1, \dots, n$, with a selective strict identification function V_{sel} appears to have slightly less complications: The forecasts can be considered valid if $\frac{1}{n} \sum_{i=1}^n V_{\text{sel}}(x_i, y_i)$ is 0 (or sufficiently close to it) for all choices $x_i \in B_i$ and for all $i = 1, \dots, n$. Despite the additional computational complexity, this procedure produces conclusive results.

3. Structural results

Our structural results consist of generalisations of the classical Convex Level Sets (CxLS) property due to Osband (1985) and their immediate implications (Section 3.1), the main result on the mutual exclusivity of selective and exhaustive elicibility (Section 3.2), and implications of certain specifications of set-valued functionals in Section 3.3. The most important structural results of Sections 3.1 and 3.2 are illustrated and summarised in Figure 1.

3.1. CxLS properties and their implications

Definition 3.1. Let $T: \mathcal{M} \rightarrow 2^W$ be a set-valued functional and $\mathcal{M}' \subseteq \mathcal{M}$.

- (i) T has the *selective* CxLS property on \mathcal{M}' if for all $F_0, F_1 \in \mathcal{M}'$ and for all $\lambda \in (0, 1)$ such that $(1 - \lambda)F_0 + \lambda F_1 \in \mathcal{M}'$:

$$T(F_0) \cap T(F_1) \subseteq T((1 - \lambda)F_0 + \lambda F_1).$$

- (ii) T has the *selective* CxLS* property on \mathcal{M}' if for all $F_0, F_1 \in \mathcal{M}'$ and for all $\lambda \in (0, 1)$ such that $(1 - \lambda)F_0 + \lambda F_1 \in \mathcal{M}'$:

$$T(F_0) \cap T(F_1) \neq \emptyset \implies T(F_0) \cap T(F_1) = T((1 - \lambda)F_0 + \lambda F_1).$$

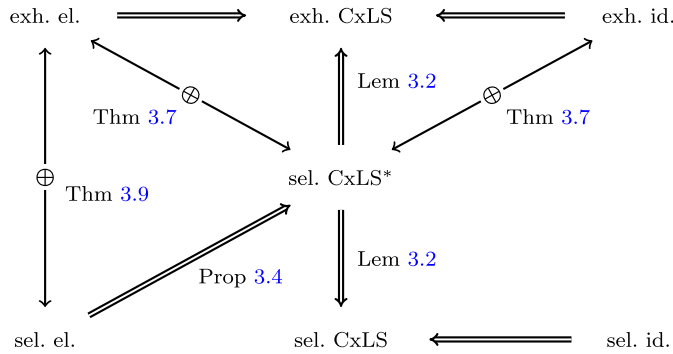


FIG 1. An illustration of the most important structural results of a set-valued functional T satisfying the proper-subset property. We abbreviate as follows: (sel)ective, (exh)haustive, (el)icitable, (id)entifiable. For two statements A and B , we write $A \implies B$ to mean A implies B , and $A \leftarrow \oplus \rightarrow B$ to mean A and B are mutually exclusive. The three implications without a reference are standard and follow along the lines of the proof of Proposition 3.4.

- (iii) T has the *exhaustive CxLS* property on \mathcal{M}' if for all $F_0, F_1 \in \mathcal{M}'$ and for all $\lambda \in (0, 1)$ such that $(1 - \lambda)F_0 + \lambda F_1 \in \mathcal{M}'$:

$$T(F_0) = T(F_1) \implies T(F_0) = T((1 - \lambda)F_0 + \lambda F_1).$$

If we omit to mention the class \mathcal{M}' explicitly, we mean that T has the corresponding CxLS property on \mathcal{M} . The exhaustive CxLS property is the most common one in the literature, and the one used for point-valued functionals (Steinwart et al., 2014; Bellini and Bigozzi, 2015; Delbaen et al., 2016; Wang and Wei, 2020). The selective CxLS property follows the one proposed in Gneiting (2011a), while the selective CxLS* property is novel. However, it is noteworthy that the recent paper Brehmer and Storkorb (2019) introduced the notion of *max-functionals*. Using our notation, a real-valued functional $T: \mathcal{M} \rightarrow \mathbb{R}$ is called a *max-functional* if for any $F_0, F_1 \in \mathcal{M}$ and $\lambda \in (0, 1)$

$$T((1 - \lambda)F_0 + \lambda F_1) = \max(T(F_0), T(F_1)).$$

It is immediate that a real-valued functional $T: \mathcal{M} \rightarrow \mathbb{R}$ is a max-functional if and only if the set-valued functional $T^+(F) := [T(F), \infty)$ satisfies the selective CxLS* property. The following implications are immediate and are illustrated in the middle column of Figure 1.

Lemma 3.2. *Let $T: \mathcal{M} \rightarrow 2^W$ be a set-valued functional and $\mathcal{M}' \subseteq \mathcal{M}$.*

- (i) *If T has the selective CxLS* property on \mathcal{M}' , then it also has the selective and the exhaustive CxLS property on \mathcal{M}' .*
- (ii) *If T is singleton-valued on \mathcal{M}' , then the selective CxLS property on \mathcal{M}' , the exhaustive CxLS property on \mathcal{M}' and the selective CxLS* property on \mathcal{M}' are equivalent.*

The second point of Lemma 3.2 underpins why the distinction of the CxLS properties is obsolete for the point-valued case.

Example 3.3. Prototypical set-valued functionals of distributions over \mathbb{R} are α -quantiles, q_α . Straightforward (graphical) considerations show that q_α satisfies the selective CxLS* property. Further examples satisfying or violating the respective CxLS properties from Definition 3.1 are listed in Example 3.10.

It follows from the seminal work of Osband (1985) that the exhaustive CxLS property is necessary for exhaustive elicibility and exhaustive identifiability, and that the selective CxLS property is necessary for selective elicibility and selective identifiability. Under additional regularity assumptions and for real-valued functionals, Steinwart et al. (2014) established that the CxLS property is also sufficient for both elicibility and identifiability (which is generally not the case for set-valued functionals; see Figure 1). A novelty is the following necessity-result.

Proposition 3.4. *If $T: \mathcal{M} \rightarrow 2^{\mathbf{A}_{\text{sel}}}$ is selectively elicitable on $\mathcal{M}' \subseteq \mathcal{M}$, then it satisfies the selective CxLS* property on \mathcal{M}' .*

Proof. Let $F_0, F_1 \in \mathcal{M}'$, $\lambda \in (0, 1)$ such that $F_\lambda = (1 - \lambda)F_0 + \lambda F_1 \in \mathcal{M}'$, and suppose there exists $t \in T(F_0) \cap T(F_1)$. Let $S: \mathbf{A}_{\text{sel}} \times \mathcal{O} \rightarrow \mathbb{R}^*$ be a strictly \mathcal{M}' -consistent selective scoring function for T . Moreover, let $x \in \mathbf{A}_{\text{sel}}$. Note that for $i \in \{0, 1\}$

$$\bar{S}(x, F_i) - \bar{S}(t, F_i) \begin{cases} = 0, & \text{if } x \in T(F_i) \\ > 0, & \text{if } x \notin T(F_i) \end{cases}$$

due to the strict \mathcal{M}' -consistency of S . This implies that

$$\begin{aligned} & \bar{S}(x, F_\lambda) - \bar{S}(t, F_\lambda) \\ &= (1 - \lambda)(\bar{S}(x, F_0) - \bar{S}(t, F_0)) + \lambda(\bar{S}(x, F_1) - \bar{S}(t, F_1)) \quad (3.1) \\ & \begin{cases} = 0, & \text{if } x \in T(F_0) \cap T(F_1) \\ > 0, & \text{if } x \notin T(F_0) \cap T(F_1). \end{cases} \end{aligned}$$

The identity in (3.1) stems from the fact that the expected score $\bar{S}(\cdot, \cdot)$ behaves “linearly” in its second argument, which is the integration measure. Again, invoking the strict \mathcal{M}' -consistency of S , the assertion follows. \square

For our next result, we need to introduce a property which essentially excludes any degenerate cases of set-valued functionals, e.g. being singleton-valued. Recall that any point-valued functional, such as the mean functional, can be identified with a singleton-valued functional.

Definition 3.5. A set-valued functional $T: \mathcal{M} \rightarrow 2^{\mathcal{W}}$ has the *proper-subset property* on $\mathcal{M}' \subseteq \mathcal{M}$ if there are $F, G \in \mathcal{M}'$ such that

$$\emptyset \neq T(G) \subsetneq T(F)$$

and for all $\varepsilon \in (0, 1)$ there exists a $\lambda_0 \in (0, \varepsilon)$ such that $(1 - \lambda_0)F + \lambda_0 G \in \mathcal{M}'$.

Again, if we omit to mention the class \mathcal{M}' explicitly, we mean that T has the proper-subset property on \mathcal{M} .

Example 3.6. The proper-subset property of Definition 3.5 generally depends very much on the choice of \mathcal{M}' . This can be seen easily when considering q_α , the α -quantile. If \mathcal{M}' is the (convex) class of all distributions on \mathbb{R} with finite support, q_α satisfies the proper-subset property. E.g., $q_{1/2}(\delta_0/2 + \delta_1/2) = [0, 1]$ and $q_{1/2}(\delta_0/2 + \delta_2/2) = [0, 2]$. If, on the other hand, \mathcal{M}' is the set of all strictly increasing distribution functions, q_α is singleton-valued, and hence clearly violating the proper-subset property.

Theorem 3.7. *If $T: \mathcal{M} \rightarrow \mathbf{A}_{\text{exh}}$ satisfies the proper-subset property and the selective CxLS* property on $\mathcal{M}' \subseteq \mathcal{M}$, it is neither exhaustively elicitable nor exhaustively identifiable on \mathcal{M}' .*

Proof. Assume S is a strictly \mathcal{M}' -consistent exhaustive scoring function and V is a strict \mathcal{M}' -identification function for T . Let $F, G \in \mathcal{M}'$ be such that $\emptyset \neq T(G) \subsetneq T(F)$. Then

$$\bar{S}(T(F), F) - \bar{S}(T(G), F) < 0 < \bar{S}(T(F), G) - \bar{S}(T(G), G)$$

and $\bar{V}(T(G), F) \neq 0$. The proper-subset property implies that there is a sufficiently small $\lambda_0 \in (0, 1)$, such that $H := (1 - \lambda_0)F + \lambda_0 G \in \mathcal{M}'$ and, exploiting the selective CxLS* property yielding $T(H) = T(G)$, we end up with

$$\begin{aligned} & \bar{S}(T(F), H) - \bar{S}(T(H), H) \\ &= (1 - \lambda_0)(\bar{S}(T(F), F) - \bar{S}(T(G), F)) + \lambda_0(\bar{S}(T(F), G) - \bar{S}(T(G), G)) < 0, \end{aligned}$$

which violates the strict \mathcal{M}' -consistency. Note that the last inequality only holds under the tacit assumption that S is \mathcal{M}' -finite. Moreover

$$\bar{V}(T(H), H) = (1 - \lambda_0) \underbrace{\bar{V}(T(G), F)}_{\neq 0} + \lambda_0 \underbrace{\bar{V}(T(G), G)}_{=0} \neq 0.$$

Hence, V cannot be a strict \mathcal{M}' -identification function. \square

Remark 3.8. Remarkably, the combination of the selective CxLS* property and the proper-subset property implies that there are $F, G \in \mathcal{M}$ with $T(F) \neq T(G)$ such that for all $\lambda \in (0, 1)$ it holds that $T((1 - \lambda)F + \lambda G) \in \{T(F), T(G)\}$. That means, in our Theorem 3.7 we directly recover the condition of Theorem 3.3 in Brehmer and Strokov (2019). Even though their result is stated for real-valued functionals only, it immediately generalises to the set-valued case. Hence, the conclusions coincide in both instances implying that T fails to be (exhaustively) elicitable.

3.2. Mutual exclusivity

We now present our main result, which states that, for functionals satisfying the proper-subset property, selective and exhaustive elicibility are mutually

exclusive. The proof follows immediately from Proposition 3.4 and Theorem 3.7; see Figure 1.

Theorem 3.9 (Mutual exclusivity). *Let $T: \mathcal{M} \rightarrow \mathbf{A}_{\text{exh}} \subseteq 2^{\mathbf{A}_{\text{sel}}}$ be a set-valued functional with the proper-subset property. Then T cannot be both selectively elicitable and exhaustively elicitable.*

This result gives a broad insight into the structure of set-valued elicibility. It basically establishes the following partition of set-valued functionals:

- (1) The class of selectively elicitable functionals.
- (2) The class of exhaustively elicitable functionals.
- (3) The class of functionals which are not elicitable at all.

The result also gives a powerful tool to rule out elicibility without the need for a direct argument, which in some cases may appear quite challenging a priori. We give several example applications of Theorem 3.9 in Example 3.10. They should be considered within the context of the practical implications on M -estimation and forecast comparison outlined in Section 2.3.

- Example 3.10.**
- (i) Any α -quantile, $\alpha \in (0, 1)$ is selectively elicitable. If the class \mathcal{M} is reasonably large, then the α -quantile clearly satisfies the proper-subset property (see Example 3.6). Hence, it fails to be exhaustively elicitable on such a class.
 - (ii) If \mathcal{M} is the class of distributions on \mathbb{R} with discrete support, then the mode functional is selectively elicitable on \mathcal{M} with the strictly \mathcal{M} -consistent selective scoring function $S(x, y) = \mathbf{1}\{x \neq y\}$ (Heinrich, 2014; Gneiting, 2017). Since the mode functional satisfies the proper-subset property on \mathcal{M} , it also fails to be exhaustively elicitable on \mathcal{M} .
 - (iii) Any elicitable real-valued functional $T: \mathcal{M} \rightarrow \mathbb{R}$ induces trivial set-valued functionals $T^-(F) := (-\infty, T(F)]$ and $T^+(F) = [T(F), \infty)$. Clearly, the elicibility of T is equivalent to the exhaustive elicibility of T^- and T^+ considered as maps to $\mathbf{A}^- = \{(-\infty, x] \mid x \in \mathbb{R}\}$ and $\mathbf{A}^+ = \{[x, \infty) \mid x \in \mathbb{R}\}$, e.g. by invoking the revelation principle (Osband, 1985; Gneiting, 2011a; Fissler, 2017). If T is not constant on \mathcal{M} , then T^- and T^+ also satisfy the proper-subset property, which means they violate the selective CxLS* property such that they are not selectively elicitable.
Vice versa, if T^+ or T^- satisfies the selective CxLS* property, then T or $-T$ is a max-functional in the sense of Brehmer and Strokov (2019) such that T (and $-T$) is not elicitable unless it is constant, which recovers their Corollary 3.4.
 - (iv) In Fissler, Hlavínová and Rudloff (2021), the exhaustive elicibility of the set-valued systemic risk measures defined by Feinstein, Rudloff and Weber (2017) has been established. For a random vector Y representing a financial system, a measure of systemic risk is defined as a collection of capital allocations $k \in \mathbb{R}^d$ such that $\rho(\Lambda(Y + k)) \leq 0$ where ρ is a scalar risk measure and $\Lambda: \mathbb{R}^d \rightarrow \mathbb{R}$ a non-decreasing aggregation function. The cash-invariance property of these risk measures implies that they satisfy

the proper-subset property. This means that they cannot be selectively elicitable.

- (v) In Section 4, we consider the class of α -prediction intervals, i.e., of intervals a random variable will fall into with a probability of at least α . We show that, on a suitable class of probability distributions \mathcal{M} , this class of α -prediction intervals is exhaustively elicitable on \mathcal{M} , and in Lemma 4.6 we construct an example showing that the proper-subset property is satisfied on \mathcal{M} . The combination of these results then rules out the selective elicibility of the class of α -prediction intervals on \mathcal{M} .
- (vi) Theorem 4.16 shows that there are classes of distributions where the collection of all shortest α -prediction intervals fails to be elicitable in either sense—selectively and exhaustively; see Remark 4.18 for details. Up to our knowledge, this is the first non-degenerate example of a set-valued functional which is not elicitable in either sense. (Clearly, any non-elicitable single-valued functional, such as the variance, would trivially satisfy such a statement by virtue of Lemma 2.3).
- (vii) In Section 5, we establish that Vorob'ev quantiles of random sets are selectively identifiable and exhaustively elicitable. Under the additional mild proper-subset property, which is satisfied in a lot of settings, this means that Vorob'ev quantiles cannot be selectively elicitable.

Remark 3.11. Elicibility and identifiability have structural differences in this context. Most importantly, as illustrated in Figure 1, while selective elicibility implies the selective CxLS* property, it is only possible to establish the necessity of the selective CxLS property for selective identifiability, due to possible cancellation effects. Hence, it does not seem possible to establish a mutual exclusivity result for selective and exhaustive identifiability in the spirit of Theorem 3.9. However, we are unaware of a specific set-valued example (satisfying the proper-subset property) which is identifiable in both senses.

Figure 1 also informs which of the four combinations of selective / exhaustive elicibility / identifiability are possible, given the proper-subset property.

- (i) Exhaustive elicibility and exhaustive identifiability are jointly possible. Invoking the revelation principle for identifiability (Fissler, 2017, Proposition 2.3.2), T^+ and T^- from Example 3.10(iii) are a specific example for this, given that the real-valued functional T is both elicitable and identifiable.
- (ii) Set-valued systemic risk measures, the class of α -prediction intervals and Vorob'ev quantiles (Example 3.10(iv), (v), (vii)) all constitute functionals which are exhaustively elicitable and selectively identifiable.
- (iii) The usual α -quantile (Example 3.10(i)) is a prototypical example for a selectively elicitable and selectively identifiable functional on the class of all continuous distributions.
- (iv) Notably, a combination of Proposition 3.4 and Theorem 3.7 rules out the existence of a selectively elicitable and exhaustively identifiable functional.

3.3. Specifications of set-valued functionals

Let us now take a look at *specifications* of set-valued functionals. For a set $W \neq \emptyset$ and a set-valued functional $T: \mathcal{M} \rightarrow 2^W$, a point-valued functional $T': \mathcal{M} \rightarrow W$ is a specification of T if $T'(F) \in T(F)$ for all $F \in \mathcal{M}$. We start with a lemma, the proof of which is straightforward.

Lemma 3.12. *Let $T: \mathcal{M} \rightarrow 2^A$ be a set-valued functional with $T(F) \neq \emptyset$ for all $F \in \mathcal{M}$, and let $T': \mathcal{M} \rightarrow A$ be a specification of T .*

- (i) *If $S: A \times \mathcal{O} \rightarrow \mathbb{R}^*$ is a (strictly) \mathcal{M} -consistent selective scoring function for T , then it is an \mathcal{M} -consistent scoring function for T' .*
- (ii) *If $V: A \times \mathcal{O} \rightarrow \mathbb{R}$ is a (strict) selective \mathcal{M} -identification function for T , then it is an \mathcal{M} -identification function for T' .*

Clearly, the scoring function S (identification function V) that appears in Lemma 3.12 is only *strictly* consistent (*strict*) for T' if T is a singleton on \mathcal{M} , that is, $T(F) = \{T'(F)\}$ for all $F \in \mathcal{M}$. This suggests the question as to whether the specification can be elicitable (identifiable) at all, which the following proposition is concerned with.

Proposition 3.13. *Let $T: \mathcal{M} \rightarrow 2^A$ be selectively elicitable and $T': \mathcal{M} \rightarrow A$ a specification of T . Let $\mathcal{M}_1 := \{F \in \mathcal{M} \mid T(F) = \{T'(F)\}\}$ and suppose that $\mathcal{M} \setminus \mathcal{M}_1 \neq \emptyset$.*

Let $\mathcal{S}_{\mathcal{M}}$ ($\mathcal{S}_{\mathcal{M}_1}$) be the class of strictly \mathcal{M} -consistent (\mathcal{M}_1 -consistent) selective scoring functions for T . If $\mathcal{S}_{\mathcal{M}} = \mathcal{S}_{\mathcal{M}_1}$, then $T': \mathcal{M} \rightarrow A$ is not elicitable.

Proof. Let $\mathcal{S}'_{\mathcal{M}}$ ($\mathcal{S}'_{\mathcal{M}_1}$) be the class of strictly \mathcal{M} -consistent (\mathcal{M}_1 -consistent) scoring functions for T' . If $\mathcal{S}_{\mathcal{M}_1} = \mathcal{S}_{\mathcal{M}}$, it holds that $\mathcal{S}'_{\mathcal{M}} \subseteq \mathcal{S}'_{\mathcal{M}_1} = \mathcal{S}_{\mathcal{M}_1} = \mathcal{S}_{\mathcal{M}}$. However, any $S \in \mathcal{S}_{\mathcal{M}}$ fails to be strictly \mathcal{M} -consistent for T' . Hence, $\mathcal{S}'_{\mathcal{M}} = \emptyset$. \square

A common problem when applying Proposition 3.13 for practical purposes is that most characterisation results concerning the class of strictly consistent scoring functions, if known, typically assume regularity conditions on the scoring functions such as continuity or differentiability; cf. Table 1 in Gneiting (2011b) or Osband's Principle (Osband, 1985; Fissler and Ziegel, 2016). Interestingly, an argument similar to the one used in the proof of Theorem 3.7 leads to a result which rules out the elicibility of specifications under very weak conditions on the functional. In particular, it dispenses with regularity conditions on scoring functions.

In line with Bellini and Bignozzi (2015) we call a functional T' from a convex class of distributions \mathcal{M} to some topological space A mixture-continuous if for any $F_0, F_1 \in \mathcal{M}$ the map $[0, 1] \ni \lambda \mapsto T'((1 - \lambda)F_0 + \lambda F_1) \in A$ is continuous.

Proposition 3.14. *Let \mathcal{M} be convex and $T: \mathcal{M} \rightarrow 2^A$, $T \neq \emptyset$, satisfy the selective CxLS* property. Suppose there are distributions $F, G, H \in \mathcal{M}$ such that*

$$T(F) \cap T(G) = \{t_1\}, \quad T(F) \cap T(H) = \{t_2\}, \quad \text{with } t_1 \neq t_2. \quad (3.2)$$

Then, any specification $T': \mathcal{M} \rightarrow \mathbf{A}$ of T is neither elicitable nor identifiable. Moreover, if \mathbf{A} is a space with a Fréchet topology, that is, if for any $a, b \in \mathbf{A}$ with $a \neq b$ there is an open set $U \subseteq \mathbf{A}$ such that $a \in U$ and $b \notin U$, then any specification T' fails to be mixture-continuous.

Proof. Let $T': \mathcal{F} \rightarrow \mathbf{A}$ be a specification of T and suppose $S': \mathbf{A} \times \mathbf{O} \rightarrow \mathbb{R}^*$ is a strictly \mathcal{M} -consistent scoring function for T' . Let $F, G, H \in \mathcal{M}$ satisfy (3.2) with $t_1 \neq t_2$ specified there. The selective CxLS* property implies that for any $\lambda \in (0, 1)$ we have that $t_1 = T'((1 - \lambda)F + \lambda G)$ and $t_2 = T'((1 - \lambda)F + \lambda H)$. Then, for $t_0 = T'(F)$ we have that $t_0 \neq t_1$ or $t_0 \neq t_2$. Without loss of generality, assume $t_0 \neq t_1$. Therefore, the map $\gamma: [0, 1] \rightarrow \mathbf{A}$, $\lambda \mapsto T'((1 - \lambda)F + \lambda G)$ is neither injective nor constant, such that Lemma B.1 in Fissler and Ziegel (2019) implies that T' is not identifiable.

Using the fact that $t_0 = T'(F)$ and $t_1 = T'((1 - \lambda)F + \lambda G)$, $t_0 \neq t_1$ for all $\lambda \in (0, 1)$, we can directly use Brehmer and Strokorob (2019, Theorem 3.3) to conclude that there is no strictly \mathcal{M} -consistent (and \mathcal{M} -finite) score for T .

Finally, if \mathbf{A} has a Fréchet topology, γ is not continuous, which shows that T' is not mixture-continuous. Indeed, let $U \subset \mathbf{A}$ be an open set such that $t_0 \in U$, but $t_1 \notin U$. Then $\gamma^{-1}(U) = \{0\}$, which is not open in $[0, 1]$. \square

We would like to emphasise that the mere failure of mixture-continuity of T' does not rule out its elicibility. Indeed, Proposition 2.2 in Fissler and Ziegel (2019) (cf. Proposition 3.4 in Bellini and Bignozzi (2015)) only rules out the existence of a *continuous* strictly consistent scoring function for T' .

Corollary 3.15. *Let \mathcal{M} be a convex class of distributions such that (3.2) is satisfied for the α -quantile, $\alpha \in (0, 1)$. Then no specification of the α -quantile is elicitable.*

Proof. Since the α -quantile is selectively elicitable (see e.g. Gneiting (2011b)), the claim is directly implied by Proposition 3.14. \square

Note that (3.2) is satisfied for the α -quantile e.g. if \mathcal{M} contains all distributions with finite support. Corollary 3.15 thus rules out the elicibility of the lower quantile, known as Value at Risk in quantitative risk management, the upper quantile (note that the upper quantile is a max-functional in the sense of Brehmer and Strokorob (2019) such that the non-elicibility follows with Corollary 3.4 therein), or the specification introduced in the recent preprint Aronow and Lee (2018), relative to such classes.

Notably, Proposition 3.14 rules out the elicibility with an \mathcal{M} -finite score, which also translates to Corollary 3.15. Relaxing this condition and looking at the 0- and 1-quantile, we arrive at functionals which are both selectively and exhaustively elicitable, which is the content of Subsection 4.4.

4. Prediction intervals

A common task for the statistical forecaster is to report an interval $[a, b] \subseteq \mathbb{R}$ into which future observations of a given real-valued random variable Y will fall with

at least a specified coverage probability $\alpha \in (0, 1]$, that is, $\mathbf{P}(Y \in [a, b]) \geq \alpha$. Thereby, the inherent uncertainty of the actual outcome is captured. Any such interval will be referred to as an α -prediction interval.

The literature on evaluating prediction intervals considers reports for these functionals typically in the exhaustive sense, meaning that an interval is reported rather than a single point.¹ Gneiting and Raftery (2007, Sections 6.2 and 9.3) consider consistent exhaustive scores for the central α -prediction interval or ‘equal-tailed’ α -prediction intervals; cf. Greenberg (2018) for a discussion of these scores and Bracher et al. (2020) for a timely application of interval forecasts in the context of epidemiology. This basically amounts to a prediction of a pair of quantiles at the $(1 - \alpha)/2$ - and $(1 + (1 - \alpha)/2)$ -level. If one fixes a certain coverage of, say, α , this ansatz can be generalised to construct consistent scoring functions for a non-central α -prediction interval of which the endpoints are specified in terms of quantiles at level β and $\beta + \alpha$, where $\beta \in (0, 1 - \alpha)$. Schlag and van der Weele (2015) also consider exhaustive scoring functions for interval-valued predictions. However, they start with a certain scoring function of appeal to them and do not thoroughly characterise the functional which is elicited by this scoring function. We refer to Askanazi et al. (2018) for an overview of interval forecasts, in which, however, mostly impossibility results are presented.

There is typically a whole class of α -prediction intervals for Y , resulting in a collection of subsets of \mathbb{R} . In Section 4.2 we show that this whole class of α -prediction intervals is exhaustively elicitable, subject to sensible conditions on the class of distributions. As a direct consequence of Theorem 3.9, it is not selectively elicitable. This fact imposes a substantial challenge to the sound evaluation of single arbitrary α -prediction intervals without imposing any further restrictions. On the other hand, imposing such further restrictions, it is well known that an α -prediction interval given by two quantiles as its endpoints can be elicited due to the elicibility of the individual quantiles, if the quantiles are singletons on the respective class of probability distributions. Such an interval is a particular specification of the class of α -prediction intervals, and one might wonder about the elicibility (identifiability) of other specifications. In Section 4.3, we discuss the elicibility and identifiability of several specifications of the class of α -prediction intervals, with largely negative results.

Our results are nicely complemented by the very recent and independently developed preprint Brehmer and Gneiting (2020). They essentially study the subclass of α -prediction intervals with exact coverage α , and show that this subclass fails to be selectively elicitable; see Remark 4.8 for details. Furthermore, they establish properties on homogeneous and translation invariant scores for the central α -prediction interval (or ‘equal-tailed’ α -prediction interval) and

¹Reporting a single point in an α -prediction interval does not bear much information such that selective forecasts are of no practical interest. Concerning exhaustive forecasts, any interval can be identified with its endpoints and therefore the exhaustive elicibility (identifiability) is equivalent to the elicibility (identifiability) of a single vector, making use of the so called *revelation principle*, originating from Osband (1985, p. 9); see Gneiting (2011a, Theorem 4).

show some complementary impossibility results on the shortest α -prediction interval.

4.1. Notation

Let \mathcal{M}_0 be the class of Borel probability distributions on \mathbb{R} where we deliberately overload notation and identify the corresponding Borel measures with their cumulative distribution functions. Let $\bar{U} := \{(a, b)^\top \in \mathbb{R}^2 \mid a \leq b\}$, where $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$, and $U := \{(a, b)^\top \in \mathbb{R}^2 \mid a < b\}$. For any $F \in \mathcal{M}_0$ and for any $(a, b)^\top \in \bar{U}$, we set $F([a, b]) := F([a, b] \cap \mathbb{R}) = F(b) - F(a-)$ where $F(b) := F((-\infty, b])$ if $b < \infty$ and $F(\infty) := F(\mathbb{R}) = 1$, and $F(a-) := F((-\infty, a))$. For $\beta \in [0, 1]$ and $F \in \mathcal{M}_0$, recall the definition of the β -quantile, $q_\beta(F)$, and the lower β -quantile, $q_\beta^-(F)$, of F as

$$q_\beta(F) := \{t \in \bar{\mathbb{R}} \mid F((-\infty, t)) \leq \beta \leq F((-\infty, t])\},$$

$$q_\beta^-(F) := \inf q_\beta(F) = \inf\{t \in \bar{\mathbb{R}} \mid \beta \leq F((-\infty, t])\}.$$

We introduce the following subclasses of \mathcal{M}_0 : Let \mathcal{M}_{inc} be the class of strictly increasing distribution functions, $\mathcal{M}_{\text{cont}}$ the class of continuous distributions, and $\mathcal{M}_{\text{inc,cont}} := \mathcal{M}_{\text{inc}} \cap \mathcal{M}_{\text{cont}}$. Naturally, the fact that $F \in \mathcal{M}_{\text{inc}}$ implies that the support of F is whole \mathbb{R} . However, to allow for the treatment of distributions with bounded support, we define the class of α -pseudo-increasing distributions $\mathcal{M}_{\alpha,\text{inc}}$ for $\alpha \in (0, 1]$. For any $F \in \mathcal{M}_0$ we say that $F \in \mathcal{M}_{\alpha,\text{inc}}$ if and only if

$$\#q_\alpha(F) = \#q_{1-\alpha}(F) = 1, \quad \text{and} \tag{4.1}$$

$$\forall a \in \mathbb{R} \text{ s.t. } F(a-) < 1 - \alpha : \#q_{\alpha+F(a-)}(F) > 1 \implies q_{F(a-)}(F) = \{a\}, \tag{4.2}$$

where the notation $\#A$ denotes the cardinality of a set A . This means that $F \in \mathcal{M}_{\alpha,\text{inc}}$ if and only if its α - and $(1 - \alpha)$ -quantiles are singletons, and if for any $\beta \in (0, 1 - \alpha)$ the β -quantile or the $(\beta + \alpha)$ -quantile are singletons.

For any $\alpha \in (0, 1]$ and upon identifying any non-empty interval $[a, b] \subseteq \mathbb{R}$ with the vector of its endpoints $(a, b)^\top \in \bar{U}$, we formally introduce the class of α -prediction intervals for a distribution $F \in \mathcal{M}_0$ as

$$\mathcal{I}_\alpha(F) := \{(a, b)^\top \in \bar{U} \mid F([a, b]) \geq \alpha\}. \tag{4.3}$$

Clearly, $(a, b)^\top \in \mathcal{I}_\alpha(F)$ implies that $(a - x_1, b + x_2)^\top \in \mathcal{I}_\alpha(F)$ for all $x_1, x_2 \geq 0$. That is, $\mathcal{I}_\alpha(F)$ is an upper set with respect to the ordering cone $C := (-\infty, 0] \times [0, \infty)$. Moreover, $\mathcal{I}_\alpha(F)$ is non-empty since $F(\mathbb{R}) = 1 \geq \alpha$, implying that $(-\infty, \infty)^\top \in \mathcal{I}_\alpha(F)$. Therefore, we introduce the natural maximal exhaustive action domain for reports for \mathcal{I}_α as $\mathcal{U} = \{\emptyset \neq A \subseteq \bar{U} \mid A = A + C\}$, with the usual definition of the Minkowski sum. Moreover, for any $F \in \mathcal{M}_0$, we introduce several functions closely connected to $\mathcal{I}_\alpha(F)$: First, the function

$$\Gamma_\alpha(F) := \{a \in \bar{\mathbb{R}} \mid F(a-) \leq 1 - \alpha\} \rightarrow (-\infty, \infty],$$

$$a \mapsto \Gamma_\alpha(F)(a) = \inf\{b \geq a \mid F([a, b]) \geq \alpha\} = q_{\alpha+F(a-)}^-(F) \tag{4.4}$$

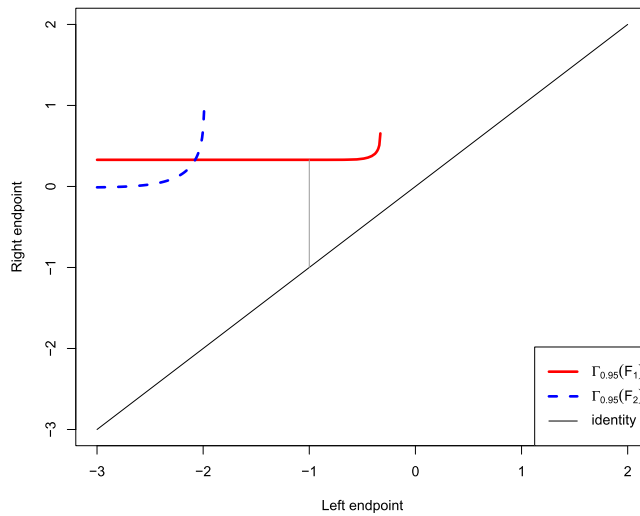


FIG 2. Graphs of $\Gamma_{0.95}(F_1)$ (red solid), $F_1 = \mathcal{N}(0, 0.2^2)$, and $\Gamma_{0.95}(F_2)$ (blue dashed), $F_2 = \mathcal{N}(-1, 0.6^2)$. In light grey, the shortest 0.95-prediction interval of F_1 with lower endpoint -1 is depicted.

gives the upper endpoint of the shortest α -prediction interval with lower endpoint a ; see Figure 2 for an illustration.

- Lemma 4.1.** (i) For all $F \in \mathcal{M}_0$ and for all $a \in \bar{\mathbb{R}}$ such that $F(a-) \leq 1 - \alpha$ it holds that $F([a, \Gamma_\alpha(F)(a)]) \geq \alpha$.
 (ii) For all $F \in \mathcal{M}_0$ the function $\Gamma_\alpha(F)$ is increasing and left-continuous.
 (iii) For $F \in \mathcal{M}_{\text{inc, cont}}$ the function $\Gamma_\alpha(F)$ is strictly increasing and continuous.
 (iv) For $F \in \mathcal{M}_{\alpha, \text{inc}}$ it holds that $\lim_{a \rightarrow -\infty} \Gamma_\alpha(F)(a) = q_\alpha^-(F)$, and that $\Gamma_\alpha(F)^{-1}(\{\infty\}) \in \{\emptyset, q_{1-\alpha}(F)\}$, which means that ∞ is attained at most once.

Proof. (i) follows from the right continuity of F . (ii) is implied by the fact that the functions $a \mapsto F(a-)$ and $\beta \mapsto q_\beta^-(F)$ are increasing and left-continuous. For (iii) recall that for $F \in \mathcal{M}_{\text{inc, cont}}$, both $a \mapsto F(a-)$ and $\beta \mapsto q_\beta^-(F)$ are strictly increasing and continuous. Finally, for (iv) note that for $F \in \mathcal{M}_{\alpha, \text{inc}}$, $\beta \mapsto q_\beta^-(F)$ is continuous at α and $1 - \alpha$. Since $\lim_{a \rightarrow -\infty} F(a-) = 0$, the first assertion follows. For the latter, if there is no $a \in \mathbb{R}$ such that $F(a-) = 1 - \alpha$, the function $\Gamma_\alpha(F)$ is clearly finite. If there is some $a_0 \in \mathbb{R}$ such that $F(a_0-) = 1 - \alpha$ then $a_0 \in q_{1-\alpha}(F) = \{q_{1-\alpha}^-(F)\}$ such that it is unique. $\Gamma_\alpha(F)(q_{1-\alpha}^-(F)) = q_1^-(F) = \infty$ if and only if the support of F is unbounded from above. As $q_{1-\alpha}(F)$ is a singleton, for all $b < a_0$ it holds that $F(b-) < F(a_0-)$. Then $\alpha + F(b-) < 1$ and therefore $\Gamma_\alpha(F)(b)$, being the lower $(\alpha + F(b-))$ -quantile of F , is finite. \square

Similarly to $\Gamma_\alpha(F)$, we introduce for $F \in \mathcal{M}_0$ the map $d_\alpha(F): \mathbb{R} \rightarrow [0, \infty]$,

$$m \mapsto d_\alpha(F)(m) = \inf\{c \geq 0 \mid F([m - c, m + c]) \geq \alpha\}, \quad (4.5)$$

which gives (half of) the length of the shortest α -prediction interval of F , centred at m . Again, a continuity argument yields that the infimum is attained such that $F([m - d_\alpha(F)(m), m + d_\alpha(F)(m)]) \geq \alpha$. Moreover, for $\alpha < 1$, $d_\alpha(F)$ takes finite values only.

Note that $\mathcal{I}_\alpha(F)$ corresponds to the epigraph of $\Gamma_\alpha(F)$, given by

$$\text{epi } \Gamma_\alpha(F) := \{(a, b)^\top \in \bar{U} \mid b \geq \Gamma_\alpha(F)(a), F(a-) \leq 1 - \alpha\}, \quad F \in \mathcal{M}_0,$$

which also guarantees the (Borel-) measurability of the set $\mathcal{I}_\alpha(F)$. We also introduce the graph of $\Gamma_\alpha(F)$ as

$$\text{graph } \Gamma_\alpha(F) := \{(a, \Gamma_\alpha(F)(a))^\top \in \bar{U} \mid F(a-) \leq 1 - \alpha\}, \quad F \in \mathcal{M}_0.$$

Finally, we introduce the subclass $\mathcal{U}^* \subseteq \mathcal{U}$ of sets which can be written in form of epigraphs of left-continuous functions $\gamma: [-\infty, b] \rightarrow (-\infty, \infty]$, for some $b \in \mathbb{R}$ such that $\gamma^{-1}(\{\infty\}) \in \{\emptyset, \{b\}\}$ and such that $\lim_{a \rightarrow -\infty} \gamma(a) = \gamma(-\infty)$.

4.2. Elicitability and identifiability of the class of α -prediction intervals

One of the main results of this paper is as follows.

Theorem 4.2. *For $\alpha \in (0, 1]$ the following assertions hold:*

- (i) *The functional $\mathcal{M}_0 \ni F \mapsto \text{graph } \Gamma_\alpha(F)$ is selectively identifiable on $\mathcal{M}_{\text{inc, cont}}$ with the strict selective $\mathcal{M}_{\text{inc, cont}}$ -identification function*

$$V_{\text{sel}}: U \times \mathbb{R} \rightarrow \mathbb{R}, \quad (x, y) \mapsto V_{\text{sel}}(x, y) = \mathbf{1}\{y \in [x_1, x_2]\} - \alpha.$$

Moreover on \mathcal{M}_0 , V_{sel} is still a selective \mathcal{M}_0 -identification function for $\text{graph } \Gamma_\alpha(\cdot)$ and it is oriented in the sense that for all $F \in \mathcal{M}_0$ it holds that $\bar{V}(x, F) \geq 0$ if and only if $x \in \mathcal{I}_\alpha(F)$.

- (ii) *Let μ be a finite, σ -additive, nonnegative measure on U . The function $S_{\text{exh}}: \mathcal{U} \times \mathbb{R} \rightarrow \mathbb{R}$*

$$S_{\text{exh}}(A, y) = - \int_{A \cap U} V_{\text{sel}}(x, y) \, d\mu(x) = \alpha\mu(A) - \mu(((-\infty, y] \times [y, \infty)) \cap A) \quad (4.6)$$

is an \mathcal{M}_0 -consistent exhaustive scoring function for \mathcal{I}_α .

- (iii) *If additionally μ is positive on U ,² then the restriction of S_{exh} to $\mathcal{U}^* \times \mathbb{R}$ is strictly $\mathcal{M}_{\alpha, \text{inc}}$ -consistent for \mathcal{I}_α , rendering the class of α -prediction intervals exhaustively elicitable on $\mathcal{M}_{\alpha, \text{inc}}$.*

²That means any nonempty open subset of U has positive measure under μ .

Proof. Part (i) follows from Lemma 4.1 and standard arguments. For part (ii) let $F \in \mathcal{M}_0$, $A^* = \mathcal{I}_\alpha(F)$ and $A \in \mathcal{U}$. Then, using a Fubini argument, we obtain

$$\begin{aligned} \bar{S}_{\text{exh}}(A, F) - \bar{S}_{\text{exh}}(A^*, F) = \\ \int_{(A^* \setminus A) \cap U} \bar{V}_{\text{sel}}(x, F) \, d\mu(x) - \int_{(A \setminus A^*) \cap U} \bar{V}_{\text{sel}}(x, F) \, d\mu(x) \geq 0. \end{aligned} \quad (4.7)$$

The inequality is easily established by recalling that $\bar{V}_{\text{sel}}(x, F) \geq 0$ if and only if $x \in A^*$. The proof of part (iii) is deferred to Appendix C. \square

Note that in part (iii), the fact that $q_{F(x_1-)}(F)$ is a singleton whenever $\#q_{\alpha+F(x_1-)}(F) > 1$, implied by $F \in \mathcal{M}_{\alpha, \text{inc}}$, plays an important role. If this were not the case, we would obtain rectangles of points $x \in \mathcal{I}_\alpha(F)$ with $\bar{V}_{\text{sel}}(x, F) = 0$ with positive measure under μ , namely $q_{F(x_1-)}(F) \times q_{\alpha+F(x_1-)}(F)$. This would mean that our exhaustive scoring function fails to distinguish between the correct forecast and one that does not contain some of the points within this rectangle. The reasoning behind why the α - and $(1 - \alpha)$ -quantiles are required to be singletons is similar.

For $\alpha = 1$, note that $\mathcal{M}_{\alpha, \text{inc}}$ only contains distributions with support \mathbb{R} . But in that case, \mathcal{I}_α is constant, namely $\{\mathbb{R}\}$ for all distributions, and thus not interesting. We will therefore exclude the case $\alpha = 1$ from further discussion.

Remark 4.3. Imposing the normalisation condition $S(A, y) \geq 0$ with equality if and only if $A = \mathcal{I}_\alpha(\delta_y) = [-\infty, y] \times [y, \infty]$, it is straightforward to construct a score equivalent to the one in (4.6) given by

$$\begin{aligned} S_\mu(A, y) &= S_{\text{exh}}(A, y) - S_{\text{exh}}([-\infty, y] \times [y, \infty], y) \\ &= (1 - \alpha)\mu((-\infty, y] \times [y, \infty)) + \alpha\mu(A) - \mu(((-\infty, y] \times [y, \infty)) \cap A) \\ &= (1 - \alpha)\mu(((-\infty, y] \times [y, \infty)) \setminus A) + \alpha\mu(A \setminus ((-\infty, y] \times [y, \infty))). \end{aligned} \quad (4.8)$$

From this stage, one can easily construct a family of elementary scores, $S_u := S_{\delta_u}$, $u \in U$, given by (4.8). As a consequence of Theorem 4.2, these elementary scores are $\mathcal{M}_{\alpha, \text{inc}}$ -consistent for \mathcal{I}_α . Clearly, $S_\mu(A, y) = \int S_u(A, y)\mu(du)$ which is a mixture representation in the spirit of Ehm et al. (2016). This opens the way to the powerful tool of Murphy diagrams $u \mapsto S_u(A, y)$ discussed there as well. In order to avoid the necessity of choosing a measure μ , one instead considers the elementary scores in (4.8) over different values of the parameter $u \in U$. In the one-dimensional case discussed in Ehm et al. (2016) as well as in the case of the class of α -prediction intervals, one can easily visualise the values of the expected score differences graphically. With the possibly increasing dimensionality of the space u comes from, the illustrative accessibility of this approach gets more involved. We discuss an example with possibly higher dimension in Section 5. For an illustration of 2-dimensional Murphy diagrams, we refer the reader to Fissler, Hlavínová and Rudloff (2021).

Intuitively, the class of α -prediction intervals, $\mathcal{I}_\alpha(F)$, of a distribution F contains a great deal of information about F itself. So one might wonder if

it is possible to recover F , knowing $\mathcal{I}_\alpha(F)$. If so, this would mean that \mathcal{I}_α actually constitutes a bijection. And consequently, the exhaustive elicibility of \mathcal{I}_α would directly follow from the existence of strictly proper scoring rules for probabilistic forecasts (Gneiting and Raftery, 2007), invoking the *revelation principle* (Osband, 1985; Gneiting, 2011a). The following proposition asserts that \mathcal{I}_α is not a bijection, which underlines the novelty of Theorem 4.2.

Proposition 4.4. *For $\alpha \in (0, 1)$ the functional \mathcal{I}_α is not injective on $\mathcal{M}_{\alpha, \text{inc}}$.*

Proof. Assume first that $\alpha = 1/n$ for some integer $n > 1$. For any $a \in [0, 1]$ define the distribution $F_a \in \mathcal{M}_{\alpha, \text{inc}}$ with density $f_a(y) = 1 - a \cos(2\pi ny)$ for $y \in [0, 1]$ and $f_a(y) = 0$ otherwise. For all $a \in [0, 1]$ and $x \in [0, 1 - \alpha]$, observe that $\int_x^{x+\alpha} f_a(y) dy = \alpha - a \int_x^{x+1/n} \cos(2\pi ny) dy = \alpha$. Hence, we have $\mathcal{I}_\alpha(F_a) = \{(x, x + \alpha)^\top : x \in [0, 1 - \alpha]\} \cup \{(z, \alpha)^\top : z \leq 0\} \cup \{(1 - \alpha, z)^\top : z \geq 1\}$ for all $a \in [0, 1]$, violating injectivity of \mathcal{I}_α .

Otherwise, let $\beta := 1\% \alpha > 0$, where $a\%b := a - b[a/b]$, the real analog of the modulus for $a, b \in \mathbb{R}$, $a, b > 0$. For all $0 \leq a \leq \beta/(\alpha - \beta)$, define the probability density

$$f_a(y) = \begin{cases} 0 & y \notin [0, 1] \\ 1 - a(\alpha/\beta - 1) & y\% \alpha \leq \beta \\ 1 + a & y\% \alpha > \beta. \end{cases}$$

Thus, f_0 is the uniform density on $[0, 1]$, and for $a > 0$, f_a raises and lowers the density according to where y falls modulo α . Letting $F_a \in \mathcal{M}_{\alpha, \text{inc}}$ be the corresponding probability measure, we will show that $\mathcal{I}_\alpha(F_a) = \mathcal{I}_\alpha(F_0)$ for all $a > 0$. We again see that $\mathcal{I}_\alpha(F_0) = \{(x, x + \alpha)^\top : x \in [0, 1 - \alpha]\} \cup \{(z, \alpha)^\top : z \leq 0\} \cup \{(1 - \alpha, z)^\top : z \geq 1\}$. For F_a , note that the Lebesgue measure of the set $\{y \in [x, x + \alpha] : y\% \alpha \leq \beta\}$ is exactly β for all x . Thus, when $x \in [0, 1 - \alpha]$, we have $F_a([x, x + \alpha]) = \beta(1 - a(\alpha/\beta - 1)) + (\alpha - \beta)(1 + a) = \beta - a(\alpha - \beta) + (\alpha - \beta) + a(\alpha - \beta) = \alpha$, as desired. The cases $[z, \alpha]$ and $[1 - \alpha, z]$ follow immediately. \square

Remark 4.5. Variants of prediction intervals other than connected intervals might also be natural to consider, e.g., wrapped intervals (allowing intervals of the form $(-\infty, b] \cup [a, \infty)$ where $b < a$), unions of intervals, and most generally, any measurable prediction set. In Appendix B, we show that most of these generalisations are indeed bijective with F . That means their exhaustive elicibility follows directly from the existence of strictly proper scoring rules for probabilistic reports and the revelation principle. One exception is the case of wrapped intervals when α is rational, as the construction in the first case of Proposition 4.4 applies, and injectivity fails. (When α is irrational, repeatedly wrapping intervals corresponds to an irrational rotation, from which one can compute a dense set of quantiles such that one can again invoke the revelation principle to obtain exhaustive elicibility.) We claim that the class of wrapped prediction intervals with a rational α is exhaustively elicitable under

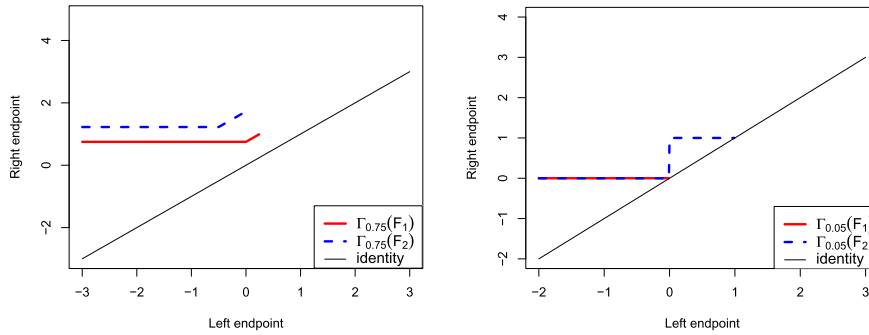


FIG 3. Illustration of the proper-subset property of \mathcal{I}_α . Left panel: Graphs of $\Gamma_{0.75}(F_1)$ (red solid), $F_1 = \text{Unif}([0, 1])$, and $\Gamma_{0.75}(F_2)$ (blue dashed), $F_2 = \text{Unif}([-0.5, 1.8])$. Right panel: Graphs of $\Gamma_{0.05}(F_1)$ (red solid), $F_1 = \delta_0$, and $\Gamma_{0.05}(F_2)$ (blue dashed), $F_2 = \delta_0/2 + \delta_1/2$.

mild assumptions on the underlying class of distributions, using a similar integral construction as the one in Theorem 4.2.

In order to use Theorems 3.9 and 4.2 to conclude that \mathcal{I}_α is not selectively elicitable on $\mathcal{M}_{\alpha,\text{inc}}$ it is essential to show that \mathcal{I}_α satisfies the proper-subset property on $\mathcal{M}_{\alpha,\text{inc}}$.

Lemma 4.6. \mathcal{I}_α satisfies the proper-subset property on $\mathcal{M}_{\alpha,\text{inc}}$ for $\alpha \in (0, 1)$.

Proof. We first show the claim for $\alpha \in [1/2, 1)$. We determine $\mathcal{I}_\alpha(F)$ for $F = \text{Unif}([b, c])$, a uniform distribution on $[b, c]$, $b < c$. One easily verifies that $\Gamma_\alpha(\text{Unif}([b, c]))(a) = \max(a, b) + \alpha(c - b)$ for $a \leq c - \alpha(c - b)$. For larger a , $\Gamma_\alpha(\text{Unif}([b, c]))$ is not defined. A straightforward calculation shows that for any $b < 0$ and any $1 - b(1 - \alpha)/\alpha \leq c \leq 1 - b\alpha/(1 - \alpha)$, the domain of $\Gamma_\alpha(\text{Unif}([b, c]))$ is contained in the domain of $\Gamma_\alpha(\text{Unif}([0, 1]))$ and $\Gamma_\alpha(\text{Unif}([b, c])) \geq \Gamma_\alpha(\text{Unif}([0, 1]))$ where the two functions do not coincide. As a result $\emptyset \neq \mathcal{I}_\alpha(\text{Unif}([b, c])) \subsetneq \mathcal{I}_\alpha(\text{Unif}([0, 1]))$; see the left panel of Figure 3. Moreover, any convex mixture of two uniform distributions is an element of $\mathcal{M}_{\alpha,\text{inc}}$.

For $\alpha \in (0, 1/2)$, it holds that $\mathcal{I}_\alpha(\delta_0) = [-\infty, 0] \times [0, \infty]$ while $\mathcal{I}_\alpha(\delta_0/2 + \delta_1/2) = [-\infty, 0] \times [0, \infty] \cup ([-\infty, 1] \times [1, \infty])$; see the right panel of Figure 3. Note that even though there are $\lambda \in (0, 1)$ such that $(1 - \lambda)\delta_0 + \lambda(\delta_0/2 + \delta_1/2) \notin \mathcal{M}_{\alpha,\text{inc}}$, the proper-subset property is still satisfied. \square

Corollary 4.7. For $\alpha \in (0, 1)$, the class \mathcal{I}_α of α -prediction intervals is not selectively elicitable on $\mathcal{M}_{\alpha,\text{inc}}$.

Proof. This is a direct combination of Theorems 3.9, 4.2, and Lemma 4.6. \square

Remark 4.8. Corollary 4.7 is related to the impossibility result established in Brehmer and Gneiting (2020, Section 3.1), that the ‘guaranteed coverage interval at level α ’ (GCI_α) is not selectively elicitable; see also Lambert and Shoham (2009, Proposition 7.6) for a related result. Roughly speaking, GCI_α coincides

with the topological boundary of $\mathcal{I}_{1-\alpha}$, or graph $\Gamma_{1-\alpha}$. Even though there is an obvious bijection between GCI_α and $\mathcal{I}_{1-\alpha} = \text{epi } \Gamma_{1-\alpha}$, one cannot invoke the revelation principle here. It would only hold for exhaustive elicibility and not for selective elicibility. Therefore, even though the results are closely related, they are complementary.

4.3. Prediction interval with an endpoint or the midpoint given by an identifiable functional

In some situations, one might be interested in prediction intervals with one endpoint or the midpoint specified as some (identifiable) functional. The simplest situation arises if the midpoint or an endpoint is simply a constant; we defer the discussion of this simple situation to Appendix A. Apart from constants, the most natural such functionals appear to be the mean or the median for the midpoint, while also other quantiles or expectiles might be interesting. If one endpoint is specified in terms of some quantile, the other endpoint must be a quantile itself and the elicibility of the vector is obvious and well known (if the quantiles are both singletons); see e.g. Gneiting and Raftery (2007) or Proposition 4.9, which recalls this result for the sake of completeness. On the other hand, we can show that there are no twice continuously differentiable strictly consistent exhaustive scoring functions (see Propositions 4.10 and 4.12) for other functionals under mild conditions. In the case of the midpoint given by an identifiable functional, this even holds for the quantile. This gives rise to the conjecture that such intervals are in general not elicitable. Despite their failure of being (smoothly) elicitable, these functionals are still identifiable, therefore possessing the CxLS property. This leads to the novel observation that, in the multivariate setting, the equivalence of the CxLS property with identifiability and elicibility established for one-dimensional functionals in Steinwart et al. (2014) fails to hold. Moreover, it is an interesting manifestation of the general gap between the classes of strict identification functions and strictly consistent scoring functions for multivariate functionals, which has been discussed in detail in Dimitriadis, Fissler and Ziegel (2020). This gap arises from classical integrability conditions: Identification functions may fail to possess an antiderivative.

We only address the case of the left endpoint given by an identifiable functional and remark that the right endpoint case works *mutatis mutandis*.

Proposition 4.9. *Let $QI_{\alpha,\beta}: \mathcal{M} \rightarrow U$ be a prediction interval given by two lower quantiles, i.e. $QI_{\alpha,\beta}(F) = (q_\beta^-(F), q_{\alpha+\beta}^-(F))^\top \in U$ with $\beta \in (0, 1 - \alpha)$. The following assertions hold:*

- (i) *$QI_{\alpha,\beta}$ is identifiable on any subclass \mathcal{M} of \mathcal{M}_0 such that for all $F \in \mathcal{M}$ and for $\gamma \in \{\beta, \alpha + \beta\}$ it holds that $q_\gamma^-(F) = \{q_\gamma^-(F)\}$ and $F(q_\gamma^-(F)) = \gamma$. A strict \mathcal{M} -identification function is given by*

$$V: U \times \mathbb{R} \rightarrow \mathbb{R}^2, \quad V(z_1, z_2, y) = (\mathbb{1}\{y \leq z_1\} - \beta, \mathbb{1}\{y \leq z_2\} - \alpha - \beta)^\top.$$

- (ii) $QI_{\alpha,\beta}$ is elicitable on any subclass \mathcal{M} of \mathcal{M}_0 such that for all $F \in \mathcal{M}$ and for $\gamma \in \{\beta, \alpha + \beta\}$ it holds that $q_\gamma(F) = \{q_\gamma^-(F)\}$. Any sum of two strictly \mathcal{M} -consistent scoring functions for the respective quantiles is a strictly \mathcal{M} -consistent scoring function for $QI_{\alpha,\beta}$.

Note that, in fact, essentially any strictly consistent scoring function for $QI_{\alpha,\beta}$ is a sum of two strictly consistent scoring functions for the respective quantiles; see Fissler and Ziegel (2016, Proposition 4.2); cf. Fissler and Ziegel (2021). Recently, Brehmer and Gneiting (2020, Theorem 3.1) characterised all translation invariant or positively homogeneous consistent scores for the central α -prediction interval.

Choosing $\beta = 0$, reporting $QI_{\alpha,\beta}$ would boil down to reporting the lower α -quantile, such that elicibility holds if the α -quantile is a singleton and identifiability holds if additionally $F(q_\alpha^-(F)) = \alpha$. For the case $\beta = 1 - \alpha$, the second component of $QI_{\alpha,\beta}$ is the essential supremum. We refer to Subsection 4.4 for elicibility results.

Interestingly, Proposition 4.9 together with the mutual exclusivity result of Theorem 3.9 implies that there cannot be a scoring function $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ such that the expected score is minimised on an interval between two quantiles, subject to very mild conditions on the class of distributions \mathcal{M} (such that the proper-subset property is satisfied for $QI_{\alpha,\beta}$).

In Proposition 4.9, we ensured the existence of the α -prediction interval by restricting the range of β . Similarly, one has to restrict the class of probability distributions suitably to ensure the existence of an interval with the demanded coverage when the left endpoint is given by some general identifiable functional $l: \mathcal{M} \rightarrow \mathbb{R}$ where \mathcal{M} is some subclass of \mathcal{M}_0 . To assure that there is enough mass above $l(F)$, we write $\mathcal{M}_l = \{F \in \mathcal{M} \mid F(l(F)-) \leq 1 - \alpha\}$. For a midpoint specification in terms of an identifiable functional $m: \mathcal{M} \rightarrow \mathbb{R}$, such a restriction is not necessary.

Proposition 4.10. *Let $\alpha \in (0, 1)$. Let $l: \mathcal{M} \rightarrow \mathbb{R}$ be an identifiable functional with a strict \mathcal{M} -identification function $V_l: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$. Set $b: \mathcal{M}_l \rightarrow [0, \infty)$, defined as $b(F) = (\Gamma_\alpha(F)(l(F)) - l(F))/2$, which is half of the length of the shortest α -prediction interval with lower endpoint l , and set $T_l := (l, b)^\top: \mathcal{M}_l \rightarrow \mathbb{R} \times [0, \infty)$. Then the following assertions hold:*

- (i) T_l is identifiable on $\mathcal{M}_l \cap \mathcal{M}_{\text{inc,cont}}$ with a strict identification function

$$V: \mathbb{R} \times [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}^2, \quad V(z_1, z_2, y) = (V_l(z_1, y), \mathbf{1}\{y \in [z_1, z_1 + 2z_2]\} - \alpha)^\top.$$

- (ii) Assume that \mathcal{M}_l is such that

- (a) $\mathcal{M}_l \cap \mathcal{M}_{\text{inc,cont}}$ is convex;
- (b) for any $z \in \mathbb{R} \times (0, \infty)$ there are $F_1, F_2, F_3 \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc,cont}}$ such that 0 is in the interior of the convex hull of the set $\{\bar{V}(z, F_i) \mid i \in \{1, 2, 3\}\} \subset \mathbb{R}^2$;
- (c) $\bar{V}(\cdot, F)$ is continuously differentiable on $\mathbb{R} \times (0, \infty)$ for all $F \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc,cont}}$;

(d) for any $(l^*, b^*)^\top \in \mathbb{R} \times (0, \infty)$ there are distribution functions $F_i \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$ with densities f_i , $i \in \{1, 2, 3, 4\}$, such that

$$\begin{aligned} (l(F_i), b(F_i))^\top &= (l^*, b^*)^\top \quad \forall i \in \{1, 2, 3, 4\}, \\ \bar{V}'_l(l^*, F_1) &= \bar{V}'_l(l^*, F_2) = \bar{V}'_l(l^*, F_3), \\ f_1(l^* + 2b^*) &= f_2(l^* + 2b^*), \\ f_1(l^* + 2b^*) &\neq f_3(l^* + 2b^*), \\ f_1(l^*) &\neq f_2(l^*), \\ \bar{V}'_l(l^*, F_4) &\neq 0. \end{aligned}$$

Then there is no strictly $\mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$ -consistent scoring function S for T_l such that $\bar{S}(\cdot, F)$ is twice continuously differentiable on $\mathbb{R} \times (0, \infty)$ for any $F \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$.

Proof. See Appendix C. □

Points (a), (b) and (d) are basically richness assumptions on the class $\mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$, which are needed to establish necessary conditions on the shape of possible strictly consistent scoring functions via Osband's principle; see Fissler and Ziegel (2016, Theorem 3.2). In particular, (b) and (d) in combination with the convexity stipulated under (a) constitute a surjectivity condition where (b) additionally assumes that the expected identification function may vary enough. Point (c) is a pure smoothness assumption which is needed since the proof exploits first and second order conditions. In concrete situations, e.g. when $\mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$ is the class of finite Gaussian mixtures and l is the mean functional, these conditions can be verified by straightforward calculations.

Remark 4.11. If l is a lower β -quantile with $\beta < 1 - \alpha$, one can choose $V_l(x, y) = \mathbb{1}\{y \leq x\} - \beta$. In this case $\bar{V}_l(x, F) = F(x) - \beta$ and $\bar{V}'_l(x, F) = f(x)$, thus there cannot be two distribution functions $F_1, F_2 \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$ with $\bar{V}'_l(l^*, F_1) = \bar{V}'_l(l^*, F_2)$ and $f_1(l^*) \neq f_2(l^*)$. Hence, the elicibility of an interval given by two lower quantiles does not contradict Proposition 4.10.

Proposition 4.12. Let $\alpha \in (0, 1)$. Let $m: \mathcal{M} \rightarrow \mathbb{R}$ be an identifiable functional with a strict \mathcal{M} -identification function $V_m: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$. Set $b: \mathcal{M} \rightarrow [0, \infty)$, defined as $b(F) = d_\alpha(F)(m(F))$, which is half of the length of the shortest α -prediction interval with midpoint $m(F)$, and set $T_m = (m, b)^\top: \mathcal{M} \rightarrow \mathbb{A} := \mathbb{R} \times [0, \infty)$. Then the following assertions hold:

(i) T_m is identifiable on $\mathcal{M} \cap \mathcal{M}_{\text{inc, cont}}$ with a strict identification function $V: \mathbb{R} \times [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}^2$,

$$V(z_1, z_2, y) = (V_m(z_1, y), \mathbb{1}\{y \in [z_1 - z_2, z_1 + z_2]\} - \alpha)^\top.$$

(ii) Assume that \mathcal{M} is such that assumptions (a), (b) and (c) from Proposition 4.10 hold, *mutatis mutandis*. Moreover, suppose that (d) for any $(m^*, b^*)^\top \in \mathbb{R} \times (0, \infty)$, there are distribution functions $F_i \in \mathcal{M} \cap \mathcal{M}_{\text{inc, cont}}$

with densities f_i , $i \in \{1, 2, 3, 4\}$, such that

$$\begin{aligned} (m(F_i), b(F_i))^\top &= (m^*, b^*)^\top \quad \forall i \in \{1, 2, 3, 4\}, \\ \bar{V}'_m(m^*, F_1) &= \bar{V}'_m(m^*, F_2) = \bar{V}'_m(m^*, F_3), \\ f_1(m^* + b^*) + f_1(m^* - b^*) &= f_2(m^* + b^*) + f_2(m^* - b^*), \\ f_1(m^* + b^*) + f_1(m^* - b^*) &\neq f_3(m^* + b^*) + f_3(m^* - b^*), \\ f_1(m^* + b^*) - f_1(m^* - b^*) &\neq f_2(m^* + b^*) - f_2(m^* - b^*), \\ \bar{V}'_m(m^*, F_4) &\neq 0. \end{aligned}$$

Then there is no strictly $\mathcal{M} \cap \mathcal{M}_{\text{inc, cont}}$ -consistent scoring function S for T_m such that $\bar{S}(\cdot, F)$ is twice continuously differentiable on $\mathbb{R} \times (0, \infty)$ for any $F \in \mathcal{M} \cap \mathcal{M}_{\text{inc, cont}}$.

Proof. See Appendix C. □

4.4. Shortest prediction intervals

In the context of probabilistic forecasts, Gneiting, Balabdaoui and Raftery (2007, p. 243) proposed the paradigm of “maximizing the sharpness of the predictive distribution subject to calibration”, continuing: “Calibration refers to the statistical consistency between the distributional forecasts and the observations and is a joint property of the predictions and the events that materialize. Sharpness refers to the concentration of the predictive distributions and is a property of the forecasts only.” Following this rationale, a particularly well-motivated restriction of \mathcal{I}_α is the shortest prediction interval SI_α , meaning a prediction interval of minimal length (sharp) subject to achieving a coverage of at least α (calibrated). This is in line with the decision-theoretic derivation of the ‘prescriptive optimal interval forecast’ given in Askanazi et al. (2018, Section 2.2): “restrict attention to correctly-calibrated intervals, and then pick the shortest (on average).” In this subsection, we will study the elicibility of SI_α .

Let us first consider the case $\alpha = 1$, where the shortest α -prediction interval of $F \in \mathcal{M}$ is $SI_1(F) = ((\text{ess inf}(F), \text{ess sup}(F)))^\top$, which is possibly of infinite length. Here, ess inf and ess sup are the essential infimum and supremum, respectively, defined by $\sup q_0$ and $\inf q_1$, where q_α is the quantile functional. Thus, to understand the elicibility of SI_1 , it suffices to study the elicibility of ess inf and ess sup .

To this end, let $g: \mathbb{R} \rightarrow \mathbb{R}$ be an increasing and bounded function, and set $g(\pm\infty) = \lim_{x \rightarrow \pm\infty} g(x)$. Recall that for $\alpha \in (0, 1)$ an \mathcal{M}_0 -consistent selective score for the α -quantile is given by $S_\alpha(x, y) = (\mathbb{1}\{y \leq x\} - \alpha)(g(x) - g(y))$. If q_α is surjective on \mathcal{M} in the sense that for any $x \in \mathbb{R}$ there exists an $F \in \mathcal{M}$ such that $x \in q_\alpha(F)$, then S_α becomes strictly \mathcal{M} -consistent if and only if g is strictly increasing. Now consider the following generalisations of S_α for

$\alpha \in \{0, 1\}$, clearly failing to be \mathcal{M} -finite in general:

$$S_0(x, y) = \infty \cdot \mathbf{1}\{y < x\} + g(y) - g(x), \quad (4.9)$$

$$S_1(x, y) = \infty \cdot \mathbf{1}\{y > x\} + g(x) - g(y). \quad (4.10)$$

Interestingly, if g is constant, S_0 becomes a strictly \mathcal{M}_0 -consistent selective scoring function for q_0 , and S_1 for q_1 . On the other hand, if g is strictly increasing, they become strictly \mathcal{M}_0 -consistent for the essential infimum and essential supremum, respectively, and the elicibility of SI_1 then follows.

Proposition 4.13. *SI_1 can be elicited on \mathcal{M}_0 with non \mathcal{M}_0 -finite, strictly \mathcal{M}_0 -consistent score $S((a, b)^\top, y) = \infty \cdot \mathbf{1}\{y \notin [a, b]\} + g(b) - g(a)$ where $g: \mathbb{R} \rightarrow \mathbb{R}$ is strictly increasing and bounded.*

Proof. We have $\bar{S}((a, b)^\top, F) = \infty$ if $F([a, b]) < 1$, and $g(b) - g(a)$ otherwise. Clearly, this is the sum of the strictly consistent functions for the essential supremum and infimum given in (4.9) and (4.10). \square

Remark 4.14. Since $q_0 = (-\infty, \sup q_0]$ and $q_1 = [\inf q_1, \infty)$, the scores S_0 and S_1 can be directly used to construct strictly consistent *exhaustive* scoring functions for q_0 and q_1 , respectively, by invoking the revelation principle. Thus, the 0-quantile and 1-quantile are both selectively and exhaustively elicitable. Theorem 3.9 does not apply here as Theorem 3.7 only holds for scoring functions whose expectation is always finite. Moreover, if we were to impose that all scores be finite in expectation, a common assumption in the literature (Fissler and Ziegel, 2016; Brehmer and Strokov, 2019; Wang and Wei, 2020), Proposition 3.14 would apply, implying the non-elicibility of ess inf and ess sup , recovering a result established in the proof of Ziegel (2016a, Corollary 4.3). Hence, while the result for SI_1 is positive, it is narrowly so, as it leans heavily on the ability to assign infinite expected scores.

Turning now to the case $\alpha \in (0, 1)$, we first observe that the shortest α -prediction interval is necessarily bounded: by a simple continuity argument for the probability measure $F \in \mathcal{M}$, there is some $C > 0$ (depending on F) such that $(-C, C)^\top \in \mathcal{I}_\alpha(F)$. For $F \in \mathcal{M}$ and $\alpha \in (0, 1)$, we may therefore define

$$SI_\alpha(F) = \{(a, b)^\top \in \mathcal{I}_\alpha(F) \mid b - a \leq d - c \text{ for all } (c, d)^\top \in \mathcal{I}_\alpha(F), c, d \in \mathbb{R}\}. \quad (4.11)$$

For many distributions F , such as the uniform distribution on $[0, 1]$, $SI_\alpha(F)$ contains more than a single element, so we again must formally distinguish between exhaustive and selective reports. Importantly, Lemma 4.15, proven in Appendix C, asserts that $SI_\alpha(F)$ is always non-empty, meaning that each distribution has at least one shortest α -prediction interval.

Lemma 4.15. *For all $\alpha \in (0, 1]$ and for all $F \in \mathcal{M}$ it holds that $SI_\alpha(F) \neq \emptyset$.*

The following theorem gives a comprehensive negative result of the elicibility of SI_α for $\alpha \in (0, 1)$.

Theorem 4.16 (Shortest α -prediction interval). (i) *For $\alpha \in (0, 1)$, the shortest α -prediction interval SI_α is not selectively elicitable on any class \mathcal{M}*

containing (a) all distributions with bounded Lebesgue densities, or (b) all distributions on \mathbb{N}_0 which are unimodal with mode k for some $k \geq 1$. The same assertion holds even if we remove the restriction that scores be \mathcal{M} -finite.

- (ii) For $\alpha \in (0, 1]$, the shortest α -prediction interval SI_α is not exhaustively elicitable (with \mathcal{M} -finite scores) on any class \mathcal{M} such that for some $x \neq y \in \mathbb{R}$, $\{(1 - \lambda)\delta_x + \lambda\delta_y \mid \lambda \in [0, 1]\} \subseteq \mathcal{M}$.

Proof. (i) If (b) holds, the assertion is an immediate consequence of Brehmer and Gneiting (2020, Theorem 3.5).

Suppose that (a) holds. Inspired by the argument given in Frongillo and Kash (2012, Section 4.2), let $\alpha \in (0, 1)$ and $G_1, G_2, G_3 \in \mathcal{M}$ be the uniform distributions on the intervals $[0, 1]$, $[1, 2]$, and $[3, 1 + 2/\alpha]$, respectively. Let $F_0 = \alpha G_1 + (1 - \alpha)G_3$ and $F_1 = \frac{1}{2}\alpha G_1 + \frac{1}{2}\alpha G_2 + (1 - \alpha)G_3$, and define $x_0 = (0, 1)^\top$, $x_1 = (0, 2)^\top$, so that we have $\{x_0\} = SI_\alpha(F_0)$ and $\{x_1\} = SI_\alpha(F_1)$. Now define $F_\lambda = (1 - \lambda)F_0 + \lambda F_1$; by construction, $\{x_1\} = SI_\alpha(F_\lambda)$ for all $0 < \lambda \leq 1$.

Now suppose for a contradiction that some scoring function S was strictly \mathcal{M} -consistent for SI_α . Here, we also dispense with the assumption that S is \mathcal{M} -finite. We first argue $\bar{S}(x_i, G_j) < \infty$ for all $i \in \{0, 1\}$ and $j \in \{1, 2, 3\}$. Let $F' = \alpha G_1 + \epsilon G_2 + (1 - \alpha - \epsilon)G_3$ for $\epsilon = \frac{1}{2} \min(\alpha, 1 - \alpha)$. The case $i = 0$ follows as $x_0 \in SI_\alpha(F')$, so $\bar{S}(x_0, F') = \alpha \bar{S}(x_0, G_1) + \epsilon \bar{S}(x_0, G_2) + (1 - \alpha - \epsilon) \bar{S}(x_0, G_3) < \infty$, and thus each term must be finite. For $i = 1$, observing $x_1 \in SI_\alpha(F_1)$, the same reasoning gives the result. We conclude that $\bar{S}(x_i, F_j) < \infty$ for $i, j \in \{0, 1\}$, as all constituent terms are finite.

Define the function $\gamma : [0, 1] \rightarrow \mathbb{R}$ by $\gamma(\lambda) = \bar{S}(x_0, F_\lambda) - \bar{S}(x_1, F_\lambda)$. Expanding by linearity of expectation, we have $\gamma(\lambda) = (1 - \lambda)(\bar{S}(x_0, F_0) - \bar{S}(x_1, F_0)) + \lambda(\bar{S}(x_0, F_1) - \bar{S}(x_1, F_1))$. We conclude that γ is a continuous function as all terms above are finite. By strict \mathcal{M} -consistency, we now have $\gamma(\lambda) > 0$ for $0 < \lambda \leq 1$ and $\gamma(0) < 0$, contradicting the continuity of γ . (Cf. Frongillo and Kash (2012, Corollary 4.12).)

- (ii) Without loss of generality, we assume that the mixtures $F_\lambda = (1 - \lambda)\delta_0 + \lambda\delta_1$ are in \mathcal{M} for all $\lambda \in [0, 1]$. Using again the convention to identify any interval $[a, b]$ with the vector of its two endpoints $(a, b)^\top$, we obtain for $\beta \in (0, 1/2]$ and $\gamma \in (1/2, 1]$

$$SI_\beta(F_\lambda) = \begin{cases} \{(0, 0)^\top\}, & \lambda \in [0, \beta), \\ \{(0, 0)^\top, (1, 1)^\top\}, & \lambda \in [\beta, 1 - \beta], \\ \{(1, 1)^\top\}, & \lambda \in (1 - \beta, 1], \end{cases}$$

$$SI_\gamma(F_\lambda) = \begin{cases} \{(0, 0)^\top\}, & \lambda \in [0, 1 - \gamma], \\ \{(0, 1)^\top\}, & \lambda \in (1 - \gamma, \gamma), \\ \{(1, 1)^\top\}, & \lambda \in [\gamma, 1]. \end{cases}$$

Clearly, for $\alpha \in (0, 1]$, SI_α is non-constant on $\{F_\lambda \mid \lambda \in [0, 1]\}$ and attains only finitely many values. Therefore, as a direct consequence of Brehmer and

Strokorb (2019, Corollary 3.5), SI_α there is no strictly \mathcal{M} -consistent and \mathcal{M} -finite exhaustive score for SI_α . \square

While Theorem 4.16 (i) considers the question of selective elicibility of SI_α and is in line with the findings of Brehmer and Gneiting (2020) (where our results complement theirs since the classes \mathcal{M} considered are different and since we can dispense with the assumption of \mathcal{M} -finiteness), there is no counterpart to part (ii) in the literature, dedicated to the exhaustive elicibility of SI_α .

Remark 4.17. The classes \mathcal{M} specified in (i) and (ii) of Theorem 4.16 are not contained in $\mathcal{M}_{\alpha,\text{inc}}$, which makes it hard to thoroughly compare the results of Theorem 4.16 with Theorem 4.2. The elicibility of SI_α on $\mathcal{M}_{\alpha,\text{inc}}$ remains an open problem, though we conjecture a negative result.

This should also be compared with the discussion of Brehmer and Gneiting (2020, Condition 3.7 and Theorem 3.8). The class considered in their Theorem 3.8 is also not contained in $\mathcal{M}_{\alpha,\text{inc}}$. In particular, they also leave open the problem of elicibility on classes of distributions with strictly positive Lebesgue densities.

Remark 4.18. On any class \mathcal{M} satisfying the conditions of Theorem 4.16 (i) and (ii) SI_α fails to be selectively elicitable and exhaustively elicitable. This yields an interesting set-valued functional which fails to be elicitable in either sense.

Remark 4.19. One may also consider general prediction *regions* rather than merely intervals, in which case a natural object to study is the α -prediction region with smallest Lebesgue-measure. One can employ a very similar argument to the one used in the proof of Theorem 4.16 (ii) to rule out the exhaustive elicibility (with an \mathcal{M} -finite score) of the class of α -prediction regions with minimal Lebesgue measure, denoted by SR_α . Indeed, again writing $F_\lambda = (1 - \lambda)\delta_0 + \lambda\delta_1$, we obtain for $\beta \in (0, 1/2]$ and $\gamma \in (1/2, 1]$

$$SI_\beta(F_\lambda) = \begin{cases} \{\{0\}\}, & \lambda \in [0, \beta), \\ \{\{0\}, \{1\}\}, & \lambda \in [\beta, 1 - \beta], \\ \{\{1\}\}, & \lambda \in (1 - \beta, 1], \end{cases} \quad SI_\gamma(F_\lambda) = \begin{cases} \{\{0\}\}, & \lambda \in [0, 1 - \gamma], \\ \{\{0, 1\}\}, & \lambda \in (1 - \gamma, \gamma), \\ \{\{1\}\}, & \lambda \in [\gamma, 1]. \end{cases}$$

The rest of the argument follows the lines of the proof.

Proposition 4.20. *The shortest α -prediction interval is selectively identifiable on $\mathcal{M}_{\text{inc,cont}}$ in the following sense. Let $[-1, 1]^\mathbb{R}$ denote the space of all functions $\mathbb{R} \rightarrow [-1, 1]$, and define the function-valued identification function $V: U \times \mathbb{R} \rightarrow [-1, 1]^\mathbb{R}$ by*

$$\mathbb{R} \ni a \mapsto V((x_1, x_2)^\top, y)(a) = \mathbb{1}\{y \in [x_1 + a, x_2 + a]\} - \alpha,$$

for $(x_1, x_2)^\top \in U$ and $y \in \mathbb{R}$. Then for any $F \in \mathcal{M}_{\text{inc,cont}}$ and any $(x_1, x_2)^\top \in U$ it holds that $(x_1, x_2)^\top \in SI_\alpha(F)$ if and only if

$$\bar{V}((x_1, x_2)^\top, F)(a) \leq 0 \quad \forall a \in \mathbb{R} \quad \text{and} \quad \bar{V}((x_1, x_2)^\top, F)(0) = 0. \quad (4.12)$$

Proof. Let $F \in \mathcal{M}_{\text{inc, cont}}$ and $(x_1, x_2)^\top \in U$ such that (4.12) holds. The equality $\bar{V}((x_1, x_2)^\top, F)(0) = 0$ implies that $(x_1, x_2)^\top \in \mathcal{I}_\alpha(F)$ and $x_2 = \Gamma_\alpha(F)(x_1)$ such that there is no shorter α -prediction interval for F with a lower endpoint of x_1 . The condition $\bar{V}((x_1, x_2)^\top, F)(a) \leq 0$ for all $a \in \mathbb{R}$ means that all other intervals $[x_1 + a, x_2 + a]$ with the same length either fail to be an α -prediction interval (corresponding to a strict inequality), or they are also an α -prediction interval (corresponding to equality), but with the same logic as above, there cannot be a shorter one with the same lower endpoint $x_1 + a$. Hence, we can conclude that $(x_1, x_2)^\top \in SI_\alpha(F)$. Vice versa, if $(x_1, x_2)^\top \in SI_\alpha(F)$, then (4.12) is immediate. \square

In closing, we would like to remark that for multivariate observations, a generalisation from prediction intervals to prediction regions is mandatory. If we do not impose any restrictions other than measurability, say, we can still obtain a selective identifiability result in the spirit of Theorem 4.2 (i). For other similar extensions of our results, Remark 4.19 points into a negative direction for the smallest prediction regions. For considerations analogue to the ones in Subsection 4.3 one would need to impose further restrictions on the shape of the regions (e.g. one might consider balls with a certain centre and radius) to ask sensible questions. This is beyond the scope of the current project. On the other hand, the following section elaborates on a complementary direction of Vorob'ev quantiles, which only become interesting in a multivariate / spatial setting.

5. Vorob'ev quantiles

As Azzimonti et al. (2021) point out, the “problem of estimating the set of inputs that leads a system to a particular behavior is common in many applications”, and they explicitly mention the fields of reliability engineering and climatology (see references therein). In such a context, the quantity of interest is a *random set* \mathbf{Y} . This set could specify the region of a blackout in a country, the area affected by an avalanche in the mountains or tumorous tissue in the human body. In many situations such as extreme weather events, e.g. floods, storms or heatwaves, the random set \mathbf{Y} is specified in terms of an excursion set $\{z \in \mathbb{R}^d \mid \xi_z \geq t\}$, $t \in \mathbb{R}$, of some random field $(\xi_z)_{z \in \mathbb{R}^d}$. One of the main tasks in mathematical statistics is to construct confidence intervals or confidence regions in \mathbb{R}^d from a random sample. Consequently, such confidence regions may also be considered as random sets. Functionals of interest are various expectations of \mathbf{Y} as described in the comprehensive textbook Molchanov (2017), notably, the Vorob'ev expectation (Chevalier et al., 2013), the distance average expectation (Azzimonti et al., 2016) and conservative estimates based on Vorob'ev quantiles (Azzimonti et al., 2021).

In this section, we shall focus on Vorob'ev quantiles and shall notably establish exhaustive elicibility results and related selective identifiability results under reasonable conditions. In that respect, it generalises and extends the known result that the symmetric difference in measure is an exhaustive consis-

tent scoring function for the median; see Proposition 2.2.8 in Molchanov (2017) and below for details.

To settle some notation, we work again on some suitable complete, atomless probability space $(\Omega, \mathfrak{F}, \mathbf{P})$. Let E be some generic separable Banach space equipped with its Borel σ -algebra, with the Euclidean space as a leading example. Let μ be some σ -finite non-negative reference measure on E and let \mathfrak{U} be the family of closed subsets of E . We shall use the convention to denote any subset of E with a capital latin letter, with the additional distinction that a random set will be denoted with a bold capital letter.

Definition 5.1 (Random closed set). $\mathbf{Y}: \Omega \rightarrow \mathfrak{U}$ is called a *random closed set* if for all compact sets $K \subseteq E$

$$\{\omega \mid \mathbf{Y}(\omega) \cap K \neq \emptyset\} \in \mathfrak{F}.$$

In decision-theoretic terminology, that means that our observation domain \mathcal{O} coincides with \mathfrak{U} . In line with Definition 5.1 and following Molchanov (2017, Chapter 1), we equip \mathfrak{U} with the σ -algebra generated by the family $\mathfrak{B}(\mathfrak{U}) := \{U \in \mathfrak{U}: U \cap K \neq \emptyset, K \in \mathfrak{K}\}$ where \mathfrak{K} is the collection of all compact subsets of E . Consequently, we shall identify the distribution $F_{\mathbf{Y}}$ of a random closed set \mathbf{Y} with its *capacity functional*. That is, we set $F_{\mathbf{Y}}: \mathfrak{K} \rightarrow [0, 1]$, $F_{\mathbf{Y}}(K) = \mathbf{P}(\mathbf{Y} \cap K \neq \emptyset)$. As before, let \mathcal{M} denote some generic class of distributions $\mathfrak{K} \rightarrow [0, 1]$.

While $F_{\mathbf{Y}}$ characterises the whole (joint) distribution of a random closed set \mathbf{Y} , its restriction to singletons, in some sense, specifies the marginal distributions of \mathbf{Y} . This restriction is called *coverage function* $p_{\mathbf{Y}}: E \rightarrow [0, 1]$ and is formally defined as

$$p_{\mathbf{Y}}(u) := F_{\mathbf{Y}}(\{u\}) = \mathbf{P}(u \in \mathbf{Y}).$$

Finally, we can define the Vorob'ev quantiles of closed random sets.

Definition 5.2 (Vorob'ev quantile). The upper excursion set of $p_{\mathbf{Y}}$ at level $\alpha \in [0, 1]$,

$$Q_{\alpha}(\mathbf{Y}) := \{u \in E \mid p_{\mathbf{Y}}(u) \geq \alpha\},$$

is called the *Vorob'ev α -quantile* of \mathbf{Y} .

The Vorob'ev α -quantile plays a special role in the context of confidence regions. Suppose $\mathbf{Y} = g_{\alpha}(Y_1, \dots, Y_n)$ where Y_1, \dots, Y_n are i.i.d. random vectors in \mathbb{R}^m following some parametric distribution $F(\theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$, and $g_{\alpha}: (\mathbb{R}^m)^n \rightarrow \mathfrak{U}$ is a measurable map. In this context, E clearly corresponds to Θ . Then one can say that \mathbf{Y} constitutes an α -confidence region for the parameter θ if $\theta \in Q_{\alpha}(\mathbf{Y})$ for all $\theta \in \Theta$.

As $p_{\mathbf{Y}}$ is an upper semicontinuous function (Molchanov, 2017), $Q_{\alpha}(\mathbf{Y})$ is a closed set in E . Therefore, in a decision-theoretic terminology, we set the exhaustive action domain to be \mathfrak{U} and the selective action domain to be E . Moreover, for further reference, define the sets

$$Q_{\alpha}^{>}(\mathbf{Y}) = \{u \in E \mid p_{\mathbf{Y}}(u) > \alpha\}, \quad Q_{\alpha}^{\leq}(\mathbf{Y}) = \{u \in E \mid p_{\mathbf{Y}}(u) = \alpha\}.$$

Note that the measurability of these sets is implied by the upper-semicontinuity (and thus measurability) of $p_{\mathbf{Y}}$. It goes without saying that the quantities Q_α , $Q_\alpha^>$ and Q_α^- are law-invariant functionals in that they only depend on the distribution $F_{\mathbf{Y}}$ of a random closed set \mathbf{Y} , and, *a fortiori*, on its coverage function $p_{\mathbf{Y}}$. Therefore, we shall consider them as maps defined on some generic class of distributions \mathcal{M} .

Proposition 2.2.8 in Molchanov (2017) establishes that the symmetric difference in measure

$$S_{1/2}: \mathfrak{U} \times \mathfrak{U} \rightarrow \mathbb{R}, \quad S_{1/2}(X, Y) = \frac{1}{2}\mu(X \Delta Y) \tag{5.1}$$

is an \mathcal{M} -consistent exhaustive scoring function for the Vorob'ev median $Q_{1/2}(\mathbf{Y}): \mathcal{M} \rightarrow \mathfrak{U}$. Other Vorob'ev quantiles solve a restricted minimisation problem; see Proposition 4 in Azzimonti et al. (2021). More precisely, for $\alpha \in [0, 1]$, $Q_\alpha = Q_\alpha(\mathbf{Y})$ it holds that

$$\mathbf{E}\left[\frac{1}{2}\mu(Q_\alpha \Delta \mathbf{Y})\right] \leq \mathbf{E}\left[\frac{1}{2}\mu(M \Delta \mathbf{Y})\right]$$

for all measurable sets $M \subseteq E$ such that $\mu(M) = \mu(Q_\alpha)$. To arrive at a consistent scoring function for a general $\alpha \in [0, 1]$, we first introduce a strict selective \mathcal{M} -identification function for Q_α^- .

Proposition 5.3. *For $\alpha \in [0, 1]$, the function $V_\alpha: E \times \mathfrak{U} \rightarrow \mathbb{R}$, $V_\alpha(u, Y) = \mathbb{1}_Y(u) - \alpha$, is a strict selective \mathcal{M} -identification function for Q_α^- . Moreover, V_α is oriented in the sense that for all $u \in E$ and for all $F \in \mathcal{M}$*

$$\bar{V}_\alpha(u, F) \begin{cases} > 0, & u \in Q_\alpha^>(F) \\ = 0, & u \in Q_\alpha^-(F) \\ < 0, & u \notin Q_\alpha(F). \end{cases}$$

Proof. The proof follows directly from the definition of $p_{\mathbf{Y}}$. □

This oriented strict \mathcal{M} -identification function for Q_α^- turns out to be the main building block in the construction of an exhaustive \mathcal{M} -consistent scoring function for Q_α . The rationale is akin to the ones presented for the scalar case in Ziegel (2016b) or Dawid (2016), and the multivariate case in Fissler, Hlavinová and Rudloff (2021, Section 3.2).

Proposition 5.4. *Let $\mathfrak{U}_0 := \{M \in \mathfrak{U} \mid \mu(M) < \infty\}$. For any $\alpha \in [0, 1]$ with $Q_\alpha(F) \in \mathfrak{U}_0$ for all $F \in \mathcal{M}$ the function $\tilde{S}_\alpha: \mathfrak{U}_0 \times \mathfrak{U} \rightarrow \mathbb{R}$*

$$\tilde{S}_\alpha(X, Y) = - \int_X V_\alpha(u, Y) \mu(du) = \alpha\mu(X) - \mu(Y \cap X) \tag{5.2}$$

is an \mathcal{M} -consistent exhaustive scoring function for Q_α . More precisely, it holds that for all $F \in \mathcal{M}$

$$\begin{aligned} & \arg \min_{X \in \mathfrak{U}} \mathbf{E}_F \left[\tilde{S}_\alpha(X, \mathbf{Y}) \right] \\ & = \{X \in \mathfrak{U}_0 \mid \exists D \subseteq E \text{ measurable: } Q_\alpha^>(F) \subseteq D \subseteq Q_\alpha(F) \text{ and } \mu(X \Delta D) = 0\}. \end{aligned} \tag{5.3}$$

Proof. First note that—if we extend \tilde{S}_α to the family of measurable subsets of E with finite measure—it holds that for any such $D \subseteq E$ we have $\tilde{S}_\alpha(X, Y) = \tilde{S}_\alpha(D, Y)$ whenever $\mu(X \Delta D) = 0$. Now, let $X \in \mathfrak{U}_0$ such that $\mu(X \Delta D) = 0$ for some measurable $Q_\alpha^>(F) \subseteq D \subseteq Q_\alpha(F)$. Then, invoking Robbin’s Theorem (Molchanov, 2017, Theorem 1.5.16), it holds that for any $M \in \mathfrak{U}_0$ and any $F \in \mathcal{M}$

$$\begin{aligned} & \mathbf{E}_F \left[\tilde{S}_\alpha(M, \mathbf{Y}) - \tilde{S}_\alpha(X, \mathbf{Y}) \right] \\ &= \mathbf{E}_F \left[\tilde{S}_\alpha(M, \mathbf{Y}) - \tilde{S}_\alpha(D, \mathbf{Y}) \right] \\ &= \int_D \mathbf{E}_F [V_\alpha(u, \mathbf{Y})] \mu(du) - \int_M \mathbf{E}_F [V_\alpha(u, \mathbf{Y})] \mu(du) \\ &= \int_{Q_\alpha^>(F) \setminus M} \bar{V}_\alpha(u, F) \mu(du) - \int_{M \setminus Q_\alpha(F)} \bar{V}_\alpha(u, F) \mu(du) \geq 0, \end{aligned}$$

where the last inequality follows from the orientation of V_α . Moreover, the inequality is strict if and only if $\mu(Q_\alpha^>(F) \setminus M) + \mu(M \setminus Q_\alpha(F)) > 0$, which establishes equality in (5.3). Indeed, for any $M \in \mathfrak{U}$ with $\mu(Q_\alpha^>(F) \setminus M) + \mu(M \setminus Q_\alpha(F)) = 0$ choose $D = (M \cap Q_\alpha(F)) \cup Q_\alpha^>(F)$. Then D is measurable and it can be easily verified that $Q_\alpha^>(F) \subseteq D \subseteq Q_\alpha(F)$. Moreover, $D \setminus M = Q_\alpha^>(F) \setminus M$ and $M \setminus D = M \setminus Q_\alpha(F)$, so that $\mu(M \Delta D) = 0$. \square

Proposition 5.4 and in particular the equality in (5.3) exactly quantify by how much the score \tilde{S}_α fails to be strictly consistent for Q_α . Moreover, in contrast to the symmetric difference in measure in (5.1), the score \tilde{S}_α in (5.2) assumes both negative and positive values in general. Imposing the normalisation condition that $S_\alpha(Y, Y) = 0$ for all $Y \in \mathfrak{U}$, which implies the non-negativity of S_α , the score \tilde{S}_α in (5.2) is equivalent to

$$S_\alpha(X, Y) = \tilde{S}_\alpha(X, Y) + (1 - \alpha)\mu(Y) = \alpha\mu(X \setminus Y) + (1 - \alpha)\mu(Y \setminus X).$$

Moreover, one can see that one really retrieves the symmetric difference in measure for $\alpha = 1/2$. The following theorem states conditions for the *strict* consistency of S_α . In the sequel we denote the closure of any set $M \subseteq E$ with $\text{cl}(M)$ and its interior with $\text{int}(M)$.

Theorem 5.5. *Let $\alpha \in [0, 1]$.*

- (i) *For any $u \in E$ the elementary score $S_{\alpha,u} : \mathfrak{U} \times \mathfrak{U} \rightarrow [0, \infty)$,*

$$S_{\alpha,u}(X, Y) = \alpha \mathbf{1}_{X \setminus Y}(u) + (1 - \alpha) \mathbf{1}_{Y \setminus X}(u), \tag{5.4}$$

is a non-negative exhaustive \mathcal{M} -consistent scoring function for Q_α .

- (ii) *Let π be a σ -finite non-negative measure on E . Then the map $S_{\alpha,\pi} : \mathfrak{U} \times \mathfrak{U} \rightarrow [0, \infty]$,*

$$S_{\alpha,\pi}(X, Y) = \int S_{\alpha,u}(X, Y) \pi(du) = \alpha \pi(X \setminus Y) + (1 - \alpha) \pi(Y \setminus X) \tag{5.5}$$

is a non-negative exhaustive \mathcal{M} -consistent scoring function for Q_α .

(iii) If \mathcal{M} is such that $Q_\alpha(F) = \text{cl}(Q_\alpha^>(F))$ and $Q_\alpha(F) = \text{cl}(\text{int}(Q_\alpha(F)))$ for all $F \in \mathcal{M}$, then Q_α is exhaustively elicitable on \mathcal{M} .

Moreover, for any σ -finite positive measure π (that is, π assigns positive mass to all open non-empty sets) on E such that $\mathbf{E}_F[\pi(\mathbf{Y})] < \infty$ and $\pi(Q_\alpha(F)) < \infty$ for all $F \in \mathcal{M}$, the restriction of $S_{\alpha,\pi}$ defined in (5.5) to the family $\mathfrak{U}' := \{U \in \mathfrak{U} \mid U = \text{cl}(\text{int}(U))\}$ is a strictly \mathcal{M} -consistent exhaustive scoring function for Q_α .

To prove this theorem, we will need an auxiliary result that we introduce now.

Lemma 5.6. *If for two sets $A, B \subseteq E$ it holds that $A = \text{cl}(\text{int}(A))$ and $B = \text{cl}(\text{int}(B))$, then $A \Delta B \neq \emptyset$ implies $\text{int}(A \Delta B) \neq \emptyset$.*

Proof. Let A, B as above and assume that there is some $x \in A \Delta B$. Without loss of generality assume $x \in A \setminus B$. Since $x \in A$, there is a sequence $(a_n)_{n \in \mathbb{N}} \subseteq \text{int}(A)$ converging to x . Moreover, since B is closed and $x \notin B$, there is some $m \in \mathbb{N}$ such that for all $n \geq m$ it holds that $a_n \notin B$. Thus, for all $n \geq m$ we have $a_n \in \text{int}(A) \setminus B = \text{int}(A) \cap B^c \subseteq A \cap B^c$. Since the interior of a set is the union of all of its open subsets, $\emptyset \neq \text{int}(A) \cap B^c \subseteq \text{int}(A \cap B^c)$. \square

Proof of Theorem 5.5. The proof of (i) follows along the lines of the proof of Proposition 5.4 upon setting $\mu = \delta_u$. Note that with this choice of μ , any set is of finite measure. Point (ii) is a direct consequence of the nonnegativity and consistency of $S_{\alpha,u}(X, Y)$.

For (iii), let $F \in \mathcal{M}$ such that $\mathbf{E}_F[\pi(\mathbf{Y})] < \infty$ and note that for any $M \in \mathfrak{U}'$ with $\pi(M) = \infty$, we have $\bar{S}_{\alpha,\pi}(M, F) = \infty$. Therefore it suffices to consider $M \in \mathfrak{U}'$ with $\pi(M) < \infty$ and one can invoke the equality in (5.3). If $Q_\alpha(F)$ is the topological closure of $Q_\alpha^>(F)$, then $X = Q_\alpha(F)$ is the only closed set such that $Q_\alpha^>(F) \subseteq X \subseteq Q_\alpha(F)$. For any other closed set $M \in \mathfrak{U}'$ we therefore obtain that $X \Delta M \neq \emptyset$. This implies that $\text{int}(X \Delta M) \neq \emptyset$ and therefore, since π is positive, $\pi(X \Delta M) > 0$. \square

The orientation of the selective identification function V_α directly implies order-sensitivity in the sense of Nau (1985) or Fissler and Ziegel (2019) with respect to the partial order induced by the subset relation.

Proposition 5.7. *Let $\alpha \in [0, 1]$. Then any exhaustive \mathcal{M} -consistent scoring function $S_{\alpha,\pi}$ for Q_α of the form (5.5) is order-sensitive. That means for any $F \in \mathcal{M}$ and for any $A, B \in \mathfrak{U}$ such that $Q_\alpha(F) \subseteq A \subseteq B$ or $B \subseteq A \subseteq Q_\alpha(F)$ it holds that $\bar{S}_{\alpha,\pi}(A, F) \leq \bar{S}_{\alpha,\pi}(B, F)$.*

It is worth to explore further connections between mixture representation of consistent scoring functions established for Vorob'ev quantiles in Theorem 5.5 and the corresponding mixture representation in the one-dimensional case, which was introduced and discussed in Ehm et al. (2016). Indeed, the elementary scores introduced there,

$$S_{\alpha,\theta}^Q(x, y) = (\mathbb{1}\{y < x\} - \alpha) (\mathbb{1}\{\theta < x\} - \mathbb{1}\{(\theta < y)\}), \quad x, y, \theta \in \mathbb{R}$$

can be rewritten as $\alpha \mathbb{1}_{[x, \infty) \setminus [y, \infty)}(\theta) + (1 - \alpha) \mathbb{1}_{[y, \infty) \setminus [x, \infty)}(\theta)$. Of course, we can identify the reals x, y with the corresponding sets $X = [x, \infty)$ and $Y = [y, \infty)$, which shows that we end up with the form given in (5.4). For further analogy, let Z be a real-valued random variable. This induces a random closed set $\mathbf{Z} = [Z, \infty)$. Then it holds that $Q_\alpha(\mathbf{Z}) = [q_\alpha^-(Z), \infty)$. As discussed in Example 3.10 (iv), the elicibility of $q_\alpha^-(Z)$ is equivalent to the exhaustive elicibility of $[q_\alpha^-(Z), \infty)$. One can easily check that for a positive measure H on \mathbb{R} ,

$$S(x, y) = \int S_{\alpha, \theta}^Q(x, y) H(d\theta)$$

is a strictly consistent exhaustive scoring function for $[q_\alpha^-(Z), \infty)$ if and only if the closure of $Q_\alpha^>(\mathbf{Z}) = (q_\alpha^+(Z), \infty)$, where $q_\alpha^+(Z)$ is the upper α quantile of Z , corresponds to $Q_\alpha(\mathbf{Z}) = [q_\alpha^-(Z), \infty)$. That is, if and only if the α -quantile of Z is unique. This retrieves the first condition in part (iii) of Theorem 5.5. Note that in the case of a one-dimensional quantile, the second condition is equivalent to $q_\alpha^-(Z) = q_\alpha^+(Z)$, too. However, in the case of Vorob'ev quantiles it is more involved and does not follow from the first condition in general. This structural difference also highlights the importance of a thorough framework for dealing with set-valued functionals.

Remark 5.8. While we use a ‘bottom-up’ approach to construct our exhaustive scoring functions, exploiting oriented identification functions to come up with an integral construction or mixture-representation, Meng et al. (2020) pursue a complementary ‘top-down’ ansatz: They ‘disintegrate’ scoring rules for multivariate probabilistic forecasts, such as energy scores, to come up with exhaustive scoring functions for level sets (or excursion sets). For particular situations, the two approaches lead to similar results: E.g. their equation (24) turns out to be a special case of (5.2) upon considering the random set $\mathbf{Y} = \{z \in \mathbb{R}^d \mid z \geq \mathbf{y}\}$, where \mathbf{y} is an \mathbb{R}^d -valued random vector.

6. Connections to forecast evaluation in the literature

We would like to close the paper with a comprehensive literature review of different practices of treating forecasts for set-valued functionals. We think that these various perspectives illustrate the advantage our unified theoretical framework on set-valued forecast evaluation, with the thorough distinction between a selective and an exhaustive mode, entails. At the same time, these perspectives offer numerous starting points for further research projects to uncover their behaviour in terms of the classification into selectively elicitable functionals, exhaustively elicitable functionals, and functionals failing to be elicitable at all.

6.1. Statistical forecast evaluation

While Lambert, Pennock and Shoham (2008) only consider real-valued functionals where the distinction between selective and exhaustive scoring functions is

superfluous, the influential paper Gneiting (2011a) treats functionals as potentially set-valued; cf. Lambert and Shoham (2009); Bellini and Bignozzi (2015). However, only the concept of selective scoring functions with the corresponding notion of (strict) consistency and elicibility are given. Presumably, the motivation for doing so was induced by the quantile-functional as one of the most prominent examples of a set-valued functional. To the best of our knowledge, forecasts for the quantile are exclusively considered in the selective sense (Koenker, 2005; Komunjer, 2005; Gneiting, 2011b), in which they are elicitable. The reason for not considering them in the exhaustive sense might lie in the impossibility of establishing corresponding elicibility results, of which the first formal proof—to the best of our knowledge—is given in this paper.

Buansing, Golan and Ullah (2020) construct interval-valued forecasts for interval-valued observations. Yet, their directive is not explicitly specified in terms of a functional. Also, they evaluate the accuracy of their forecasts in terms of various loss functions without providing a discussion of which one is best suited for such an assessment.

The recent preprint Brehmer and Gneiting (2020) considers elicibility for the class of predictive intervals and certain specifications thereof through the lens of the selective notion.

6.2. Statistical theory and risk measurement

Quantiles and expectiles (Newey and Powell, 1987) of univariate distributions are well known (selectively) elicitable functionals. In the literature on quantitative risk management, they are also common scalar risk measures. There are different competing attempts to generalise them to a multivariate setting. We refer the reader to two recent and insightful papers and the corresponding references therein: Hamel and Kostner (2018) introduce multivariate quantiles taking the form of convex sets, and Daouia and Paindaveine (2019) introduce hyperplane-valued multivariate M -quantiles with a particular focus on hyperplane-valued multivariate expectiles. For both approaches, it remains an intriguing open question whether these functionals are selectively elicitable, exhaustively elicitable or not elicitable at all.

The newly established framework has been applied in Fissler, Hlavinová and Rudloff (2021), providing exhaustive elicibility results and selective identifiability results for set-valued systemic risk measures introduced in Feinstein, Rudloff and Weber (2017).

6.3. Spatial statistics

As mentioned at the beginning of Section 5, estimating set-valued quantities is a common endeavour in spatial statistics. In that context, forecasts and estimates are commonly considered with what we call an exhaustive angle. Interesting open theoretical questions besides Vorob'ev quantiles are to consider other functionals, notably expectations, of random sets presented in the book Molchanov (2017).

One area of particular interest in spatial statistics is meteorology and climatology. In these disciplines, forecast evaluation is more commonly known under the term *forecast verification*. We refer the reader to the comprehensive overview paper Dorninger et al. (2018). Besides simply comparing a set-valued forecast and a set-valued observation as outlined above, there are also more involved situations covered. E.g. acknowledging the spatio-temporal structure of many processes such as precipitation, one might evaluate probabilistic forecasts for the marginal distributions of the random field of interest at certain grid points, using the *neighbourhood* method (see Dorninger et al. (2018) for references). Assessing the entire joint distribution of the random field seems extremely ambitious and we are unaware of any verification method at the moment.

6.4. Regression and machine learning

Recent literature on isotonic regression embraces the idea of explicitly modelling functionals as set-valued; see Jordan, Mühlemann and Ziegel (2019) and Mösching and Dümbgen (2020), where the two papers consider these functionals in the selective sense.

Kivaranovic, Johnson and Leeb (2020) examine how to obtain prediction intervals with deep neural networks. In the area of machine learning, the recent paper Gao et al. (2019) considers set-valued regression as well, however, considering finite sets only. The observations (or response variables) Y_t are finite subsets of some label space S , which is assumed to be at most countably infinite. Denoting the regressors with $X_t \in \mathbb{R}^p$ then they are interested in finding a function $m: \mathbb{R}^p \rightarrow \{I \mid I \subseteq S, |I| < \infty\}$ such that $m(X_t)$ is reasonably close to Y_t . However, they do not explicitly specify the loss function they use for the regression problem. In an orthogonal direction, Zaheer et al. (2017) consider the case of set-valued regressors rather than set-valued responses, which does not lead to the question of an appropriate choice of loss function with set-valued arguments.

6.5. Philosophy

Within a more philosophical strand of literature about *credences*, i.e., subjective probabilities of degrees of belief, Mayo-Wilson and Wheeler (2016) argue that imprecise credences about the probability of a binary event can be represented as subsets of the unit interval $[0, 1]$; cf. Seidenfeld, Schervish and Kadane (2012). They consider numerical accuracy measures, being functions of the set-valued credence and the binary outcome. In this regard, they consider scoring functions taking sets as arguments. However, this ansatz is distinct from our focus since we consider forecasts for functionals which are inherently set-valued and dispense with a discussion of subjective probabilities, whereas they consider set-valued forecasts for a functional which is actually real-valued, namely the probability of a binary event.

Appendix A: Prediction intervals with a fixed endpoint or midpoint

We consider α -prediction intervals where one endpoint or the midpoint of the interval is fixed *a priori* at some point $a \in \mathbb{R}$ or $c \in \mathbb{R}$, respectively. We only address the case of the left endpoint being fixed—the case of the right endpoint follows *mutatis mutandis*. Clearly, if the position of the interval is specified in this sense, reporting such an interval boils down to reporting a one-dimensional quantity, e.g., the other endpoint or the length of the interval. For a fixed endpoint, however, the existence of such a prediction interval is in general no longer guaranteed (there might not be enough mass on the right of the lower endpoint), and one needs to specify the class of distribution functions more carefully. Therefore, for some fixed $\alpha \in (0, 1)$ and using the notation from Section 4, we introduce $\mathcal{M}_{a,0} = \{F \in \mathcal{M}_0 \mid F(a-) \leq 1 - \alpha\}$ for some fixed $a \in \mathbb{R}$. Moreover, let $\mathcal{M}_{a,\alpha,\text{inc}}$ be the subset of distributions in $\mathcal{M}_{a,0}$ such that $q_{\alpha+F(a-)}(F)$ is a singleton (with $\Gamma_\alpha(F)(a)$ being its unique element). We start with an endpoint specification.

Proposition A.1. (i) *On $\mathcal{M}_{a,0} \cap \mathcal{M}_{\text{inc,cont}}$ the functional $F \mapsto \Gamma_\alpha(F)(a) \in [a, \infty]$ is identifiable with a strict $\mathcal{M}_{a,0} \cap \mathcal{M}_{\text{inc,cont}}$ -identification function*

$$V: [a, \infty] \times \mathbb{R} \rightarrow \mathbb{R}, \quad V(x, y) = \mathbb{1}\{y \in [a, x]\} - \alpha.$$

(ii) *On $\mathcal{M}_{a,\alpha,\text{inc}}$ the functional $F \mapsto \Gamma_\alpha(F)(a) \in [a, \infty]$ is elicitable. An $\mathcal{M}_{a,0}$ -consistent scoring function is given by $S: [a, \infty] \times \mathbb{R} \rightarrow \mathbb{R}$*

$$S(x, y) = \int_{[x, \infty)} V(z, y) d\mu(z) = \alpha\mu([x, \infty)) - \mu([x, \infty) \cap [y, \infty)) \mathbb{1}\{y \geq a\} \tag{A.1}$$

where μ is a non-negative finite measure on \mathbb{R} . If moreover μ is positive, S is also strictly $\mathcal{M}_{a,\alpha,\text{inc}}$ -consistent.

Proof. The proof follows along the lines of the proof of Theorem 4.2. \square

Remark A.2. A non-negative equivalent version of the score in (A.1) is given via

$$\begin{aligned} \tilde{S}(x, y) &= \mathbb{1}\{y \geq a\} \left((1 - \alpha)\mu([y, \infty) \setminus [x, \infty)) + \alpha\mu([x, \infty) \setminus [y, \infty)) \right) \\ &\quad + \mathbb{1}\{y < a\} \alpha\mu([x, \infty)) \\ &= \mathbb{1}\{y \geq a\} \left((\mathbb{1}\{y \leq x\} - \alpha)(h(y) - h(x)) \right) + \mathbb{1}\{y < a\} \alpha h(x), \end{aligned} \tag{A.2}$$

where $h: (-\infty, \infty] \rightarrow \mathbb{R}$ is a decreasing function given by $h(t) = \mu([t, \infty))$ such that $h(+\infty) = 0$. Besides the obvious interpretation of the first line of (A.2) in the context of mixture representation, it is remarkable to see the structural similarity to a standard quantile score in the second line of (A.2), which is of the form $(\mathbb{1}\{y \leq x\} - \alpha)(h(y) - h(x))$. In fact, if the whole support of F lies above

a , the right endpoint of the resulting interval is the α -quantile. If F assigns positive mass to $(-\infty, a)$, there is a correction term accounting for the fact that the α -quantile would not be sufficient to achieve the required coverage anymore. Note that without the correction term, one is in the setting of Theorem 5 in Gneiting (2011a) with $w(y) = \mathbf{1}\{y \geq a\}$. This would correspond to forecasting the α -quantile of F truncated at a , i.e., loosely speaking the point under which $\alpha \cdot 100\%$ of the mass above a is located. The aim, however, is to report a point such that $\alpha \cdot 100\%$ of the *whole mass* is between a and the reported point.

Proposition A.3. *For some $m \in \mathbb{R}$ consider the functional $b_m: F \mapsto d_\alpha(F)(m) \in [0, \infty)$, specifying half of the length of the shortest α -prediction interval with midpoint m . Then the following assertions hold:*

(i) b_m is identifiable on $\mathcal{M}_{\text{inc, cont}}$ with a strict identification function

$$V: [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}, \quad (x, y) \mapsto V(x, y) = \mathbf{1}\{y \in [m - x, m + x]\} - \alpha.$$

(ii) b_m is elicitable on $\mathcal{M}_{\text{inc, cont}}$. If μ is a finite non-negative measure on $[0, \infty)$, then $S: [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$

$$S(x, y) = \int_{[0, x]} V(z, y) d\mu(z) = \mu([0, x] \cap [|y - m|, \infty)) - \alpha\mu([0, x]) \quad (\text{A.3})$$

is an $\mathcal{M}_{\text{inc, cont}}$ -consistent scoring function for b_m . It is strictly $\mathcal{M}_{\text{inc, cont}}$ -consistent if μ is positive.

Proof. The proof follows along the lines of the proof of Theorem 4.2. \square

Remark A.4. A non-negative equivalent version of the score in (A.3) is given via

$$\begin{aligned} \tilde{S}(x, y) &= \mu([0, x] \cap [|y - m|, \infty)) - \alpha\mu([0, x]) + \alpha\mu([0, |y - m|)) \\ &= (\mathbf{1}\{|y - m| \leq x\} - \alpha)(g(x) - g(|y - m|)), \end{aligned}$$

where $g: [0, \infty) \rightarrow \mathbb{R}$ is an increasing function given by $g(t) = \mu([0, t))$ such that $g(0) = 0$. Again we see the structural similarity to a standard quantile score in the second line. In particular, we see that b_m corresponds to the α -quantile of the distribution of $|Y - m|$.

Appendix B: Injectivity results for prediction interval variants

Again, we use the notation from Section 4. For simplicity, let us consider the class $\mathcal{M}^c \subseteq \mathcal{M}_0$ of probability measures with single-valued quantiles in the range $(0, 1)$, i.e., supported on an interval (potentially all of \mathbb{R}) and whose CDFs are strictly increasing on that interval. We first observe that if a functional value $T(F)$ uniquely determines a dense set of quantiles for F , then T must be injective.

Lemma B.1. *For some set W , let $T : \mathcal{M}^c \rightarrow 2^W$ and let $Q \subseteq (0, 1)$ be dense. If for all $F \in \mathcal{M}^c$, the value of $T(F)$ uniquely determines the values of $q_\beta(F)$ for all $\beta \in Q$, then T is injective.*

Proof. By definition of \mathcal{M}^c , we have $q_\beta(F) = F^{-1}(\beta)$ for all $F \in \mathcal{M}^c$ and $\beta \in (0, 1)$. As F is continuous and strictly monotone, its inverse is also continuous on $(0, 1)$ and strictly monotone. Thus, specifying the values of F^{-1} on a dense subset of $(0, 1)$ uniquely specifies F^{-1} and thus F . \square

For $\alpha \in (0, 1]$ we will now define the collection $\mathcal{C}_\alpha(F)$ of all α -prediction sets of F , as well as unions of two prediction intervals $\mathcal{I}_\alpha^2(F)$ and wrapped intervals $\mathcal{I}_\alpha^w(F)$. Recall the definition $\bar{U} := \{(a, b)^\top \in \bar{\mathbb{R}}^2 \mid a \leq b\}$, where $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$. In what follows, we will overload notation and interpret $I \in \bar{U}$ as a closed interval, so for example if $I = (a, b)^\top$ we have $F(I) = F([a, b] \cap \mathbb{R})$.

$$\begin{aligned} \mathcal{C}_\alpha : \mathcal{M}^c &\rightarrow 2^{\mathcal{B}(\mathbb{R})}, & F &\mapsto \{B \in \mathcal{B}(\mathbb{R}) : F(B) \geq \alpha\}, \\ \mathcal{I}_\alpha^2 : \mathcal{M}^c &\rightarrow 2^{\mathcal{B}(\mathbb{R})}, & F &\mapsto \{(I_1, I_2) \in \bar{U}^2 : F(I_1 \cup I_2) \geq \alpha\}, \\ \mathcal{I}_\alpha^w : \mathcal{M}^c &\rightarrow 2^{\mathcal{B}(\mathbb{R})}, & F &\mapsto \mathcal{I}_\alpha(F) \cup \{(I_1, I_2) \in \mathcal{I}_\alpha^2(F) : I_1 = (a, \infty)^\top, \\ & & & I_2 = (-\infty, b)^\top, a \geq b\}. \end{aligned}$$

We first show injectivity of \mathcal{I}_α^2 , and thus \mathcal{C}_α . We will routinely rely on the bijection between the above \mathcal{I}_α^2 and the functional $\mathcal{I}_{=\alpha}^2 : \mathcal{M}^c \rightarrow 2^{\mathcal{B}(\mathbb{R})}$ defined by $\mathcal{I}_{=\alpha}^2(F) = \{(I_1, I_2) \in \mathcal{I}_\alpha^2 : F(I_1 \cup I_2) = \alpha\}$. It is clear that $\mathcal{I}_\alpha^2(F)$ can be constructed from $\mathcal{I}_{=\alpha}^2(F)$ and vice versa, by adding or removing nested intervals. In particular, \mathcal{I}_α^2 is injective if and only if $\mathcal{I}_{=\alpha}^2$ is.

Proposition B.2. *\mathcal{I}_α^2 is injective for all $\alpha \in (0, 1)$.*

Proof. We will instead show injectivity of the functional $\mathcal{I}_{=\alpha}^2$. Let $F \in \mathcal{M}^c$, and consider first the case $\alpha \leq 1/2$. Given $\mathcal{I}_{=\alpha}^2(F)$, we will show how to compute the quantiles $q_{k,n} := F^{-1}(k\alpha/2^n)$ for all $k \in \mathbb{N}$, $n \in \mathbb{N}_0$ such that $k\alpha/2^n \in (0, 1)$, at which point the result will follow from Lemma B.1. We first show the result for $k \leq 2^n$; the other values will follow from the observation that $F((q_{k,n}, q_{k+2^n,n}]) = \alpha$.

As a base case, consider $n = 0$ and $k = 1$. The value $q_{1,0}$ defines the unique interval of the form $(-\infty, q_{1,0}]$ such that $F((-\infty, q_{1,0}]) = \alpha$. Thus, we may take any $(I_1, I_2) \in \mathcal{I}_{=\alpha}^2(F)$ with $I_1 = (-\infty, a)^\top$ and $I_2 = (a, b)^\top$ and set $q_{1,0} = b$.

Now assume the value of $q_{k,n} = F^{-1}(k\alpha/2^n)$ is known for all $1 \leq k \leq 2^n$; we will show how to compute $q_{k,n+1}$ for all $1 \leq k \leq 2^{n+1}$. We will show that there is a unique triple of intervals of the form $I_1 = (-\infty, x)^\top$, $I_2 = (x, q_{1,n})^\top$, $I_3 = (q_{1,0}, y)^\top$ such that $(I_1, I_3), (I_2, I_3) \in \mathcal{I}_{=\alpha}^2(F)$. For existence, take $x = q_{1,n+1}$ and $y = q_{2^{n+1}+1, n+1}$. For uniqueness, we have $F(I_1) + F(I_2) = \alpha/2^n$ by definition of $q_{1,n}$, and $F(I_1) + F(I_3) = F(I_2) + F(I_3) = \alpha$ by definition of $\mathcal{I}_{=\alpha}^2$, from which we conclude $F(I_1) = F(I_2) = \alpha/2^{n+1}$. Thus, we must have $x = q_{1, n+1}$, and the value of y follows. This triple therefore uniquely determines $q_{1, n+1}$. To recover the other values of k , observe that for all $k \leq 2^n$ we trivially have $q_{2k, n+1} = q_{k, n}$, and the unique $I'_k = (q_{k, n}, z)^\top$ with $(I'_k, I_3) \in \mathcal{I}_{=\alpha}^2(F)$ satisfies $z = q_{2k+1, n+1}$.

When $\alpha > 1/2$, we may proceed with the previous construction replacing α with $\beta = 1 - \alpha$, as follows. We now let $q_{k,n} := F^{-1}(k\beta/2^n)$ for all $k, n \in \mathbb{N}$ such that $k\beta/2^n \in (0, 1)$. Again, we first show how to compute these values for $1 \leq k \leq 2^n$, as the other values follow from the observation that $F((-\infty, q_{k,n}] \cup [q_{k+2^n, n}, \infty)) = \alpha$.

The base of the induction defines $q_{1,0} = x$ from the unique interval $I = (x, \infty)^\top$ such that $F(I) = \alpha$. To induct, we again ask for intervals $I_1 = (-\infty, x)^\top$, $I_2 = (x, q_{1,n})^\top$, $I_3 = (q_{1,0}, y)^\top$ such that $(I_1, I_3), (I_2, I_3) \in \mathcal{I}_{=\alpha}^2(F)$, with the same argument as before replacing α with β . Finally, for all $1 \leq k \leq 2^n$ we again have $q_{2k, n+1} = q_{k, n}$, and the unique $I'_k = (q_{k, n}, x)^\top$ with $(I'_k, I_3) \in \mathcal{I}_{=\alpha}^2(F)$ gives $x = q_{2k+1, n+1}$. \square

Corollary B.3. \mathcal{C}_α is injective for all $\alpha \in (0, 1)$.

Proposition B.4. Let $\alpha \in (0, 1)$. \mathcal{I}_α^w is injective if and only if α is irrational.

Proof. First, consider irrational $\alpha \in (0, 1)$. As in Proposition B.2, we will instead show injectivity of $\mathcal{I}_{=\alpha}^w : \mathcal{M}^c \rightarrow 2^{\mathcal{B}(\mathbb{R})}$, $F \mapsto \mathcal{I}_{=\alpha}(F) \cup \{(I_1, I_2) \in \mathcal{I}_{=\alpha}^2(F) : I_1 = (-\infty, b)^\top, I_2 = (a, \infty)^\top, b \leq a\}$, where naturally $\mathcal{I}_{=\alpha}(F) = \{I \in \mathcal{I}_\alpha(F) : F(I) = \alpha\}$.

For any $q \in (0, 1)$, suppose we have established $x = F^{-1}(q)$. If $q + \alpha \leq 1$, then by taking $I \in \mathcal{I}_{=\alpha}^w(F)$ with $I = (x, y)^\top$ for some $y \in \mathbb{R}$, we may conclude $F(y) = q + \alpha$. If $q + \alpha > 1$, then by taking $(I_1, I_2) \in \mathcal{I}_{=\alpha}^w(F)$ with $I_1 = (x, \infty)^\top$, $I_2 = (-\infty, y)^\top$ for some $y \in \mathbb{R}$, we may conclude $F(y) = q + \alpha - 1 \in (0, 1)$. In both cases, therefore, the values of $F^{-1}(q)$ and $\mathcal{I}_{=\alpha}^w(F)$ together determine the value of $F^{-1}((q + \alpha)\%1)$, where $a\%b := a - b[a/b]$ is the real modulus (see Proposition 4.4).

Let $x_1 \in \mathbb{R}$ be defined by $(-\infty, x_1)^\top \in \mathcal{I}_{=\alpha}^w(F)$, so that $F(x_1) = \alpha$. Proceeding as above, for all $k \in \mathbb{N}$, we may determine the values $x_k \in \mathbb{R}$ by $F(x_k) = k\alpha\%1$. As α is irrational, the set $\{k\alpha\%1 : k \in \mathbb{N}\}$ is dense in $(0, 1)$, concluding the proof.

Now consider rational $\alpha \in (0, 1)$, so that $\alpha = k/n$ for some $k, n \in \mathbb{N}$, $k < n$. For any $0 \leq b < 1/(2\pi n)$ define the cumulative distribution function F_b by $F_b(y) = y + b \sin(2\pi n y)$ for $y \in [0, 1]$. Then for all $y \in [0, 1 - \alpha]$, we have

$$F_b(y + \alpha) - F_b(y) = \alpha + b \sin(2\pi n(y + k/n)) - b \sin(2\pi n y) = \alpha.$$

For $1 - \alpha < y \leq 1$, let $z = y + \alpha - 1$. Then

$$\begin{aligned} F_b(1) - F_b(y) + F_b(z) - F_b(0) &= (1 - y) + b(\sin(2\pi n) - \sin(2\pi n y)) + z + b(\sin(2\pi n z) - \sin(0)) \\ &= (1 - y + z) + b(\sin(2\pi n z) - \sin(2\pi n y)) \\ &= \alpha + b(\sin(2\pi n(1 + z)) - \sin(2\pi n y)) \\ &= \alpha + b(\sin(2\pi n(y + k/n)) - \sin(2\pi n y)) = \alpha. \end{aligned}$$

We conclude, for all $0 \leq b < 1/(2\pi n)$, that

$$\begin{aligned} \mathcal{I}_{=\alpha}^w(F_b) &= \{(y, y + \alpha)^\top : y \in [0, 1 - \alpha]\} \cup \{(y, \alpha)^\top : y \leq 0\} \\ &\cup \{(1 - \alpha, y)^\top : y \geq 1\} \cup \{((y, \infty)^\top, (-\infty, y + \alpha - 1)^\top) : y \in [1 - \alpha, 1]\}. \end{aligned}$$

Since \mathcal{I}_α^w is in turn determined by $\mathcal{I}_{=\alpha}^w$, it also fails to be injective for rational α . \square

Appendix C: Omitted technical proofs

Proof of Theorem 4.2(iii). Let μ be positive, $F \in \mathcal{M}_{\alpha, \text{inc}}$, $A^* = \mathcal{I}_\alpha(F) \in \mathcal{U}^*$ and $A \in \mathcal{U}^*$ such that $A \neq A^*$. Let $\gamma: [-\infty, b] \rightarrow (-\infty, \infty]$ and $\gamma^*: [-\infty, b^*] \rightarrow (-\infty, \infty]$, $b, b^* \in \mathbb{R}$, such that $A = \text{epi } \gamma$ and $A^* = \text{epi } \gamma^*$. We first show that $\text{int}((A \Delta A^*) \cap U) \neq \emptyset$. If $b \neq b^*$, it is obvious that $\text{int}((A \Delta A^*) \cap U) \neq \emptyset$, due to the fact that γ and γ^* are left-continuous and since γ and γ^* assume infinity at most at b and b^* , respectively. Now assume that $b = b^*$. If $\gamma(-\infty) \neq \gamma^*(-\infty)$, then the right-continuity at $-\infty$ implies that there is some $a \in \mathbb{R}$ such that $\gamma(x) \neq \gamma^*(x)$ for all $x \in (-\infty, a)$. Hence $\text{int}((A \Delta A^*) \cap U) \neq \emptyset$. Finally, if there is some $a \in (-\infty, b]$ such that $\gamma(a) \neq \gamma^*(a)$, then the left-continuity implies that there is some $\varepsilon > 0$ such that $\gamma \neq \gamma^*$ on $(a - \varepsilon, a]$. Hence, again $\text{int}((A \Delta A^*) \cap U) \neq \emptyset$.

Since $\bar{V}(x, F) < 0$ for $x \notin A^*$, we obtain a strict inequality in (4.7) if $\text{int}((A \setminus A^*) \cap U) \neq \emptyset$. Otherwise, consider any $x \in \text{int}((A^* \setminus A) \cap U) \neq \emptyset$. If $\bar{V}(x, F) > 0$ for all such x , we are done. Suppose there is some $(x_1, x_2)^\top \in \text{int}((A^* \setminus A) \cap U)$ such that $\bar{V}(x_1, x_2, F) = 0$. We show that there exists some $\delta > 0$ such that $\bar{V}(\cdot, F) > 0$ on the open rectangle $(x_1 - \delta, x_1) \times (\gamma^*(x_1), x_2)$. First note that there is some $\delta > 0$ such that $(x_1 - \delta, x_1) \times (\gamma^*(x_1), x_2) \subset \text{int}((A^* \setminus A) \cap U)$, since the latter set is open and since both A and A^* are upper sets with ordering cone $(-\infty, 0] \times [0, \infty)$. Therefore, it is sufficient to show that for all $z_1 \in (x_1 - \delta, x_1)$ it holds that $\bar{V}(z_1, \gamma^*(x_1), F) = F([z_1, \gamma^*(x_1)]) - \alpha > 0$. If $\bar{V}(x_1, x_2, F) = 0$ that means that $F((\gamma^*(x_1), x_2]) = 0$. Therefore, $[\gamma^*(x_1), x_2] \subseteq q_{\alpha + F(x_1 -)}(F)$. Since $F \in \mathcal{M}_{\alpha, \text{inc}}$, that means that $q_{F(x_1 -)}(F) = \{x_1\}$. It is straightforward to see that for all $z_1 \in (x_1 - \delta, x_1)$ we have that $F(z_1 -) < F(x_1 -)$. Hence,

$$F([z_1, \gamma^*(x_1)]) - \alpha = F(\gamma^*(x_1)) - F(z_1 -) - \alpha > F(\gamma^*(x_1)) - F(x_1 -) - \alpha \geq 0.$$

The claim follows. \square

Proof of Proposition 4.10. Part (i) follows from the fact that V_l is a strict identification function for l and from the strict monotonicity and continuity of $F \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$.

For (ii) suppose S is a strictly $\mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$ -consistent scoring function for T_l such that $\bar{S}(\cdot, F)$ is twice differentiable on $\mathbb{R} \times (0, \infty)$ for all $F \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$. Assumptions (a), (b) and (c) and a slight adaptation of Fissler and Ziegel (2016, Theorem 3.2, Corollary 3.3) imply the existence of a function $h: \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}^{2 \times 2}$ with differentiable components h_{ij} such that

$$\nabla_x \bar{S}(x, F) = h(x) \bar{V}(x, F), \quad \text{for all } x \in \mathbb{R} \times (0, \infty), F \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}},$$

where V is the strict \mathcal{M}_l -identification function from part (i). For any $F \in \mathcal{M}_l$ and $x \in \mathbb{R} \times (0, \infty)$, the Hessian $\nabla_x^2 \bar{S}(x, F)$ must be symmetric and, for $x = T_l(F)$, it must be positive semidefinite. We obtain

$$\partial_i \bar{S}(x, F) = h_{i1}(x) \bar{V}_l(x_1, F) + h_{i2}(x)(F(x_1 + 2x_2) - F(x_1) - \alpha)$$

for $i = 1, 2$. From the symmetry of the Hessian it follows that

$$\begin{aligned} & \partial_2 h_{11}(x) \bar{V}_l(x_1, F) + \partial_2 h_{12}(x)(F(x_1 + 2x_2) - F(x_1) - \alpha) + 2h_{12}(x)f(x_1 + 2x_2) \\ &= \partial_1 h_{21}(x) \bar{V}_l(x, F) + h_{21}(x) \bar{V}_l'(x_1, F) + \partial_1 h_{22}(x)(F(x_1 + 2x_2) - F(x_1) - \alpha) \\ &+ h_{22}(x)(f(x_1 + 2x_2) - f(x_1)). \end{aligned}$$

At $x = T_l(F)$ this yields

$$\begin{aligned} & 2h_{12}(T_l(F))f(l(F) + 2b_l(F)) \\ &= h_{21}(T_l(F))\bar{V}_l'(l(F), F) + h_{22}(T_l(F))(f(l(F) + 2b_l(F)) - f(l(F))). \end{aligned}$$

The existence of $F_1, F_2 \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$ as in assumption (d) for any $(l^*, b^*)^\top \in \mathbb{R} \times (0, \infty)$ as well as the surjectivity of T_l implicitly implied via (a) and (b) yield that $h_{22} \equiv 0$. Furthermore, the existence of $F_3 \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$ as in assumption (d) for any $(l^*, b^*)^\top \in \mathbb{R} \times (0, \infty)$ together with the surjectivity of T implies that $h_{12} \equiv 0$. Finally, the existence of $F_4 \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$ as assumed implies $h_{21} \equiv 0$.

Now, for any $x \in \mathbb{R} \times (0, \infty)$ let $F \in \mathcal{M}_l \cap \mathcal{M}_{\text{inc, cont}}$ be such that $l(F) \neq x_1$. Since V_l is a strict identification function for l , $\bar{V}_l(x_1, F) \neq 0$ and we obtain $\partial_2 h_{11}(x) = 0$. Therefore, there is a function $g: \mathbb{R} \rightarrow \mathbb{R}$ such that $h_{11}(x) = g(x_1)$ for all $x \in \mathbb{R} \times (0, \infty)$. In summary, we obtain that

$$\nabla_x \bar{S}(x, F) = \begin{pmatrix} g(x_1) \bar{V}_l(x_1, F) \\ 0 \end{pmatrix},$$

which implies that $\bar{S}(\cdot, F)$ is constant in x_2 and S cannot be strictly consistent for T_l . \square

Proof of Proposition 4.12. Part (i) follows from the fact that V_m is a strict \mathcal{M} -identification function for m and from the strict monotonicity and continuity of $F \in \mathcal{M} \cap \mathcal{M}_{\text{inc, cont}}$.

For (ii) suppose S is a strictly $\mathcal{M} \cap \mathcal{M}_{\text{inc, cont}}$ -consistent scoring function for T_m such that $\bar{S}(\cdot, F)$ is twice differentiable on $\mathbb{R} \times (0, \infty)$ for all $F \in \mathcal{M} \cap \mathcal{M}_{\text{inc, cont}}$. Using exactly the same arguments as in the proof of Proposition 4.10, we can derive the existence of a function $h: \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}^{2 \times 2}$ with differentiable components h_{ij} such that

$$\nabla_x \bar{S}(x, F) = h(x) \bar{V}(x, F), \quad \text{for all } x \in \mathbb{R} \times (0, \infty), F \in \mathcal{M} \cap \mathcal{M}_{\text{inc, cont}},$$

where V is the strict $\mathcal{M} \cap \mathcal{M}_{\text{inc, cont}}$ -identification function from part (i). Again, for any $x \in \mathbb{R} \times (0, \infty)$ and $F \in \mathcal{M} \cap \mathcal{M}_{\text{inc, cont}}$, the Hessian $\nabla_x^2 \bar{S}(x, F)$ must

be symmetric and, for $x = T_m(F)$, it must be positive semidefinite. We obtain that

$$\partial_i \bar{S}(x, F) = h_{i1}(x) \bar{V}_i(x_1, F) + h_{i2}(x)(F(x_1 + x_2) - F(x_1 - x_2) - \alpha)$$

for $i = 1, 2$. From the symmetry of the Hessian it follows that

$$\begin{aligned} & \partial_2 h_{11}(x) \bar{V}_m(x_1, F) + \partial_2 h_{12}(x)(F(x_1 + x_2) - F(x_1 - x_2) - \alpha) \\ & \quad + h_{12}(x)(f(x_1 + x_2) + f(x_1 - x_2)) \\ & = \partial_1 h_{21}(x) \bar{V}_m(x_1, F) + \partial_1 h_{22}(x)(F(x_1 + x_2) - F(x_1 - x_2) - \alpha) \\ & \quad + h_{21}(x) \bar{V}'_m(x_1, F) + h_{22}(x)(f(x_1 + x_2) - f(x_1 - x_2)). \end{aligned}$$

At $x = T_m(F)$ this yields

$$\begin{aligned} & h_{12}(T_m(F))(f(m(F) + b_m(F)) + f(m(F) - b_m(F))) \\ & = h_{21}(T_m(F)) \bar{V}'_m(m(F), F) \\ & \quad + h_{22}(T_m(F))(f(m(F) + b_m(F)) - f(m(F) - b_m(F))). \end{aligned}$$

The existence of F_1, F_2 as in assumption (d) for any $(m^*, b^*)^\top \in \mathbb{R} \times (0, \infty)$ and the surjectivity of T_m implicitly given via (a) and (b) imply that $h_{22} \equiv 0$. Furthermore, the existence of F_3 as in assumption (d) together with the surjectivity of T_m implies that $h_{12} \equiv 0$. Finally, the existence of F_4 as assumed implies $h_{21} \equiv 0$. The rest of the argument follows as in the proof of Proposition 4.10. \square

Proof of Lemma 4.15. For $\alpha = 1$, note that $SI_1(F) = \{(\text{ess inf}(F), \text{ess sup}(F))^\top\}$. Now let $\alpha \in (0, 1)$. First note that $SI_\alpha(F) \neq \emptyset$ if and only if the function $h(x) := \Gamma_\alpha(F)(x) - x$ attains its infimum over the interval $P := \{x \in \mathbb{R} \mid F(x-) \leq 1 - \alpha\}$ where we note that P is closed, bounded from above and unbounded from below.

Assume that $\Gamma_\alpha(F)$ is continuous. Then h is continuous, too. Since $\Gamma_\alpha(F)(x) \geq x$, $m := \inf_{x \in P} h(x) \geq 0$. The tightness of F implies that $m < \infty$. From the definition of the infimum, there is a sequence $(x_n)_{n \in \mathbb{N}} \subseteq P$ with $h(x_n) \rightarrow m$. If this sequence is bounded from below, there is a convergent subsequence $(x_{n_k})_{k \in \mathbb{N}}$ with limit $x \in P$ and the continuity of h implies that $h(x_{n_k}) \rightarrow h(x)$, thus the infimum is attained. If $(x_n)_{n \in \mathbb{N}}$ is not bounded from below, there is a divergent subsequence $(x_{n_l})_{l \in \mathbb{N}}$. But then Lemma 4.1 (iv) implies that $h(x_{n_l}) \rightarrow \infty \neq m$, which is a contradiction.

Finally assume that $\Gamma_\alpha(F)$ fails to be right-continuous such that h is also not continuous. h is discontinuous at x if and only if $\Gamma_\alpha(F)$ has a jump at x . Jumps of $\Gamma_\alpha(F)$ can be caused by two situations, namely if F has jumps or if F has flat spots. Both of them can occur at most countably many times (see e.g. Theorem 2.1 in Shorack (2006)) which means that $\Gamma_\alpha(F)$ can have at most countably many jumps. Let $I = \{1, 2, \dots, n_0\}$ for some $n_0 \in \mathbb{N}$ or $I = \mathbb{N}$ be some index set such that $(a_i)_{i \in I}$ is the collection jump points of $\Gamma_\alpha(F)$ with corresponding jump sizes $(j_i)_{i \in I}$. For all $i \in I$ and any $\varepsilon \in (0, j_i/2]$ it holds that $\Gamma_\alpha(F)(a_i + \varepsilon) \geq \Gamma_\alpha(F)(a_i) + j_i$ and consequently that $h(a_i + \varepsilon) > h(a_i)$. Thus, if h attains its minimum, it is not in any of the intervals $(a_i, a_i + j_i/2)$, $i \in I$.

Now define the sequence of functions h_i in the following way: Set $h_0 := h$. For any $i \in I$, if h_{i-1} is continuous at a_i , set $h_i = h_{i-1}$, else

$$h_i(x) = \begin{cases} u_i x + v_i, & x \in (a_i, a_i + j_i/2) \\ h(x), & \text{otherwise,} \end{cases}$$

where $u_i = 2(h(a_i + j_i/2) - h(a_i))/j_i$ and $v_i = h(a_i) - u_i a_i$. It can easily be verified that h_i is continuous on $[a_i, a_i + j_i/2)$ and moreover that $h_i(x) > h_i(a_i)$ for all $x \in [a_i, a_i + j_i/2)$. Therefore if h_{i-1} attains its infimum at $x^* \in \mathbb{R}$, so does h_i and $h_{i-1}(x^*) = h_i(x^*)$. The pointwise limiting function h^* of $(h_i)_{i \in \mathbb{N}}$ is a continuous function that, by an earlier argument, attains its infimum over P . By the construction of the functions h_i , h also attains its infimum over P , at the same point as h^* . \square

Acknowledgements

We would like to express our sincere gratitude to Tilmann Gneiting and Johanna Ziegel for insightful discussions about the topic, to Alexander Jordan who helped to coin the terminology of *exhaustive* versus *selective* elicibility in a joint discussion, to Dario Azzimonti, Zied Ben Bouallègue, Seamus Bradley, Jonas Brehmer, Timo Dimitriadis, David Ginsbourger, Claudio Heinrich, Kory Johnson, Xiaochun Meng, Ilya Molchanov, Ruodu Wang, and Dominic Wrazidlo for valuable discussions and helpful references, and to Yuan Li for a careful proofreading of an earlier version of this paper.

References

- ARONOW, P. M. and LEE, D. K. K. (2018). A note on breaking ties among sample medians. *arXiv:1807.03462*.
- ASKANAZI, R., DIEBOLD, F. X., SCHORHEIDE, F. and SHIN, M. (2018). On the Comparison of Interval Forecasts. *J. Time Series Anal.* **39** 953–956. [MR3867133](#)
- AZZIMONTI, D., BECT, J., CHEVALIER, C. and GINSBOURGER, D. (2016). Quantifying Uncertainties on Excursion Sets Under a Gaussian Random Field Prior. *SIAM/ASA J. Uncertain. Quantif.* **4** 850–874. [MR3531737](#)
- AZZIMONTI, D., GINSBOURGER, D., CHEVALIER, C., BECT, J. and RICHEL, Y. (2021). Adaptive Design of Experiments for Conservative Estimation of Excursion Sets. *Technometrics* **63** 13–26. [MR4205688](#)
- BASSE, G. and BOJINOV, I. (2020). A general theory of identification. *arXiv:2002.06041*.
- BELLINI, F. and BIGNOZZI, V. (2015). On elicitable risk measures. *Quant. Finance* **15** 725–733. [MR3334566](#)
- BOLIN, D. and LINDGREN, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 85–106. [MR3299400](#)

- BRACHER, J., RAY, E. L., GNEITING, T. and REICH, N. G. (2020). Evaluating epidemic forecasts in an interval format. [arXiv:2005.12881](#).
- BREHMER, J. R. and GNEITING, T. (2020). Scoring Interval Forecasts: Equal-Tailed, Shortest, and Modal Interval. [arXiv:2007.05709](#).
- BREHMER, J. R. and STROKORB, K. (2019). Why scoring functions cannot assess tail properties. *Electron. J. Stat.* **13** 4015–4034. [MR4015787](#)
- BUANSING, T. S. T., GOLAN, A. and ULLAH, A. (2020). An information-theoretic approach for forecasting interval-valued SP500 daily returns. *Int. J. Forecast.* **36** 800–813.
- CHEVALIER, C., GINSBOURGER, D., BECT, J. and MOLCHANOV, I. (2013). Estimating and quantifying uncertainties on level sets using the Vorob'ev expectation and deviation with Gaussian process models. In *mODa 10 – Advances in Model-Oriented Design and Analysis*, (D. Uciński, A. Atkinson and C. Patan, eds.) **1** Springer International Publishing.
- DAOUIA, A. and PAINDAVEINE, D. (2019). From Halfspace M-depth to Multiple-output Expectile Regression. [arXiv:1905.12718](#).
- DAVIS, M. H. A. (2016). Verification of internal risk measure estimates. *Stat. Risk Model.* **33** 67–93. [MR3574946](#)
- DAWID, P. (2016). Contribution to the discussion of “Of quantiles and expectiles: Consistent scoring functions, Choquet representations and forecast rankings” by Ehm, W., Gneiting, T., Jordan, A. and Krüger, F. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 505–562. [MR3506792](#)
- DELBAEN, F., BELLINI, F., BIGNOZZI, V. and ZIEGEL, J. F. (2016). Risk measures with the CxLS property. *Finance Stoch.* **20** 433–453. [MR3479327](#)
- DIMITRIADIS, T., FISSLER, T. and ZIEGEL, J. F. (2020). The Efficiency Gap. [arXiv:2010.14146](#).
- DORNINGER, M., GILLELAND, E., CASATI, B., MITTERMAIER, M. P., EBERT, E. E., BROWN, B. G. and WILSON, L. J. (2018). The Setup of the MesoVICT Project. *Bull. Am. Meteorol. Soc.* **99** 1887–1906.
- EHM, W., GNEITING, T., JORDAN, A. and KRÜGER, F. (2016). Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 505–562. [MR3506792](#)
- ENGELBERG, J., MANSKI, C. F. and WILLIAMS, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *J. Bus. Econ. Stat.* **27** 30–41. [MR2484982](#)
- FEINSTEIN, Z., RUDLOFF, B. and WEBER, S. (2017). Measures of Systemic Risk. *SIAM J. Financial Math.* **8** 672–708. [MR3697160](#)
- FISSLER, T. (2017). On Higher Order Elicitability and Some Limit Theorems on the Poisson and Wiener Space, PhD thesis, University of Bern.
- FISSLER, T., HLAVINOVÁ, J. and RUDLOFF, B. (2021). Elicitability and identifiability of set-valued measures of systemic risk. *Finance Stoch* **25** 133–165. [MR4193814](#)
- FISSLER, T. and ZIEGEL, J. F. (2016). Higher order elicibility and Osband's principle. *Ann. Statist.* **44** 1680–1707. [MR3519937](#)
- FISSLER, T., ZIEGEL, J. F. and GNEITING, T. (2016). Expected shortfall is

- jointly elicitable with value-at-risk: implications for backtesting. *Risk Magazine* 58–61.
- FISSLER, T. and ZIEGEL, J. F. (2019). Order-Sensitivity and Equivariance of Scoring Functions. *Electron. J. Stat.* **13** 1166–1211. [MR3935847](#)
- FISSLER, T. and ZIEGEL, J. F. (2021). Correction note: Higher order elicibility and Osband’s principle. *Ann. Statist.* **49** 614. [MR4206694](#)
- FRONGILLO, R. and KASH, I. A. (2012). General Truthfulness Characterizations via Convex Analysis. [arXiv:1211.3043](#).
- FRONGILLO, R. and KASH, I. (2015). Vector-Valued Property Elicitation. *JMLR Workshop Conf. Proc.* **40** 1–18.
- GAO, T., CHEN, T., CHENTHAMARAKSHAN, V. and WITBROCK, M. (2019). A Sequential Set Generation Method for Predicting Set-Valued Outputs. [arXiv:1903.05153](#).
- GNEITING, T. (2011a). Making and Evaluating Point Forecasts. *J. Amer. Statist. Assoc.* **106** 746–762. [MR2847988](#)
- GNEITING, T. (2011b). Quantiles as optimal point forecasts. *Int. J. Forecast.* **27** 197–207.
- GNEITING, T. (2017). When is the mode functional the Bayes classifier? *Stat* **6** 204–206. [MR3671158](#)
- 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](#)
- GNEITING, T. and RAFTERY, A. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Amer. Statist. Assoc.* **102** 359–378. [MR2345548](#)
- GREENBERG, S. (2018). Calibration Scoring Rules for Practical Prediction Training. [arXiv:1808.07501](#).
- HAMEL, A. H. and KOSTNER, D. (2018). Cone distribution functions and quantiles for multivariate random variables. *J. Multivariate Anal.* **167** 97–113. [MR3830636](#)
- HEINRICH, C. (2014). The mode functional is not elicitable. *Biometrika* **101** 245–251. [MR3180670](#)
- HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 221–233. Berkeley: University of California Press. [MR0216620](#)
- HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, Second ed. John Wiley & Sons, Inc., Hoboken, New Jersey. [MR2488795](#)
- JORDAN, A., MÜHLEMANN, A. and ZIEGEL, J. F. (2019). Optimal solutions to the isotonic regression problem. [arXiv:1904.04761](#).
- KIVARANOVIC, D., JOHNSON, K. D. and LEEB, H. (2020). Adaptive, Distribution-Free Prediction Intervals for Deep Networks. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (S. CHIAPPA and R. CALANDRA, eds.). *Proceedings of Machine Learning Research* **108** 4346–4356. PMLR, Online.
- KOENKER, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge. [MR2268657](#)

- KOENKER, R. and BASSET, G. (1978). Regression Quantiles. *Econometrica* **46** 33–50. [MR0474644](#)
- KOMUNJER, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *J. Econometrics* **128** 137–164. [MR2022929](#)
- LAMBERT, N., PENNOCK, D. M. and SHOHAM, Y. (2008). Eliciting properties of probability distributions. In *EC'08: ACM Conference on Electronic Commerce, Chicago II, USA* 129–138. Association for Computing Machinery, New York, USA.
- LAMBERT, N. and SHOHAM, Y. (2009). Eliciting Truthful Answers to Multiple-Choice Questions. In *EC'09: ACM Conference on Electronic Commerce, Stanford, California, USA* 109–118. Association for Computing Machinery, New York, USA.
- MAYO-WILSON, C. and WHEELER, G. (2016). Scoring Imprecise Credences: A Mildly Immodest Proposal. *Philos. Phenomenol. Res.* **93** 55–78.
- MENG, X., TAYLOR, J. W., BEN TAIEB, S. and LI, S. (2020). Scoring Functions for Multivariate Distributions and Level Sets. [arXiv:2002.09578](#).
- MIZERA, I. (2010). *Qualitative robustness and weak continuity: the extreme uncertainty?* In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková. Collections Volume 7* 169–181. Institute of Mathematical Statistics, Beachwood, Ohio, USA. [MR2808377](#)
- MOLCHANOV, I. (2017). *Theory of Random Sets*, 2 ed. *Probability Theory and Stochastic Modelling*. Springer-Verlag London, London. [MR3751326](#)
- MÖSCHING, A. and DÜMBGEN, L. (2020). Monotone least squares and isotonic quantiles. *Electron. J. Statist.* **14** 24–49. [MR4047593](#)
- MURPHY, A. H. and DAAN, H. (1985). Forecast Evaluation. In *Probability, Statistics and Decision Making in the Atmospheric Sciences* (A. H. Murphy and R. W. Katz, eds.) 379–437. Westview Press, Boulder, Colorado.
- NAU, R. F. (1985). Should Scoring Rules Be ‘Effective’? *Manag. Sci.* **31** 527–535.
- NEWAY, W. K. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, (R. Engle and D. McFadden, eds.) **4** 36, 2111–2245. Elsevier. [MR1315971](#)
- NEWAY, W. K. and POWELL, J. L. (1987). Asymmetric Least Squares Estimation and Testing. *Econometrica* **55** 819–847. [MR0906565](#)
- NOLDE, N. and ZIEGEL, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *Ann. Appl. Stat.* **11** 1833–1874. [MR3743276](#)
- OSBAND, K. H. (1985). Providing Incentives for Better Cost Forecasting, PhD thesis, University of California, Berkeley.
- SAVAGE, L. J. (1971). Elicitation of Personal Probabilities and Expectations. *J. Amer. Statist. Assoc.* **66** 783–801. [MR0331571](#)
- SCHLAG, K. H. and VAN DER WEELE, J. J. (2015). A method to elicit beliefs as most likely intervals. *Judgm. Decis. Mak.* **10**.
- SEIDENFELD, T., SCHERVISH, M. J. and KADANE, J. B. (2012). Forecasting with imprecise probabilities. *Internat. J. Approx. Reason.* **53** 1248–1261. [MR2971869](#)

- SHORACK, G. R. (2006). *Probability for Statisticians. Springer Texts in Statistics*. Springer New York. [MR2125120](#)
- STEINWART, I., PASIN, C., WILLIAMSON, R. and ZHANG, S. (2014). Elicitation and Identification of Properties. *JMLR Workshop Conf. Proc.* **35** 1–45.
- WANG, R. and WEI, Y. (2020). Risk functionals with convex level sets. *Math. Finance* **30** 1337–1367. [MR4154772](#)
- ZAHHEER, M., KOTTUR, S., RAVANBAKHS, S., POZOS, B., SALAKHUTDINOV, R. and SMOLA, A. (2017). Deep Sets. *arXiv:1703.06114*.
- ZIEGEL, J. F. (2016a). Coherence and Elicitability. *Math. Finance* **26** 901–918. [MR3551510](#)
- ZIEGEL, J. F. (2016b). Contribution to the discussion of “Of quantiles and expectiles: Consistent scoring functions, Choquet representations and forecast rankings” by Ehm, W., Gneiting, T., Jordan, A. and Krüger, F. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 505–562. [MR3506792](#)