Elicitability and Identifiability of Systemic Risk Measures and other Set-Valued Functionals

Tobias Fissler∗ Jana Hlavinová† Birgit Rudloff†

July 3, 2019

Abstract. This paper is concerned with a two-fold objective. Firstly, we establish elicitability and identifiability results for systemic risk measures introduced in Feinstein, Rudloff, and Weber (2017). Specifying the entire set of capital allocations adequate to render a financial system acceptable, these systemic risk measures are examples of set-valued functionals. A functional is elicitable (identifiable) if it is the unique minimiser (zero) of an expected scoring function (identification function). Elicitability and identifiability are essential for forecast ranking and validation, M- and Z-estimation, both possibly in a regression framework. To account for the set-valued nature of the systemic risk measures mentioned above, we secondly introduce a theoretical framework of elicitability and identifiability of set-valued functionals. It 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. Uncovering the structural relation between the two corresponding notions of elicitability and identifiability, we establish that a set-valued functional can be either selectively elicitable or exhaustively elicitable. Notably, selections of quantiles such as the lower quantile turn out not to be elicitable in general. Applying these structural results to systemic risk measures, we construct oriented selective identification functions, which induce a family of strictly consistent exhaustive elementary scoring functions. We discuss equivariance properties of these scores. We demonstrate their applicability in a simulation study considering comparative backtests of Diebold-Mariano type with a pointwise traffic-light illustration of Murphy diagrams.

∗Imperial College London, Department of Mathematics, 180 Queen’s Gate, London SW7 2AZ, United Kingdom, e-mail: t.fissler@imperial.ac.uk
†Vienna University of Economics and Business, Institute for Statistics and Mathematics, Welthandelsplatz 1, 1020 Vienna, Austria, e-mails: jana.hlavinova@wu.ac.at and birgit.rudloff@wu.ac.at
Keywords: Consistency; Convex level sets; Diebold-Mariano tests; Identification functions; Forecast evaluation; M-estimation; Murphy diagrams; Quantiles; Scoring functions

AMS 2010 Subject Classification: 62F07; 62F10; 91G70

Contents

1. Introduction 3
   1.1. Systemic risk measures ........................................ 3
   1.2. Elicitability and identifiability ................................ 4
   1.3. Novel contributions and structure of the paper .............. 6

2. Two modes of elicitability and identifiability 7
   2.1. Consistent scoring functions for point-valued functionals .... 7
   2.2. Selective and exhaustive scoring functions .................... 9
   2.3. Selective and exhaustive identification functions ........... 11
   2.4. Structural results .............................................. 12
   2.5. Literature on forecast evaluation for set-valued functionals .... 17
      2.5.1. Statistical forecast evaluation .............................. 17
      2.5.2. Statistical theory and risk measurement .................. 17
      2.5.3. Spatial statistics ......................................... 18
      2.5.4. Regression and Machine Learning ......................... 19
      2.5.5. Philosophy .................................................. 19

3. Measures of systemic risk 19

4. Main results 23
   4.1. Identifiability results .......................................... 23
   4.2. Elicitability results and mixture representation ............... 26
      4.2.1. Comparison with one-dimensional case ..................... 27
      4.2.2. Integrability .............................................. 28
      4.2.3. Normalisation ............................................ 29
      4.2.4. Characterisation of all consistent scoring functions .... 29
      4.2.5. Order-sensitivity ......................................... 30
      4.2.6. Forecast dominance and Murphy diagrams .................. 30

5. Homogeneous and translation invariant scoring functions 31

6. Elicitability of systemic risk measures based on Expected Shortfall 34
   6.1. Identifiability results .......................................... 35
   6.2. Elicitability results ............................................ 36

7. Examples and simulations 37
   7.1. Consistency of exhaustive scoring function for $R$ ............ 37
1. Introduction

1.1. Systemic risk measures

In the financial mathematics literature, there is a great interest in various types of risk and, in particular, its quantitative measurement. The quantitative assessment of risk connected to a particular financial position dates back to Artzner, Delbaen, Eber, and Heath (1999) and has since then been discussed, from several points of view, in many further works, see e.g. Föllmer and Schied (2002), Artzner, Delbaen, and Koch-Medina (2009), Föllmer and Weber (2015). For a thorough overview of risk measures we refer the reader to the textbook Föllmer and Schied (2004).

The financial crises of 2007–2009 and its aftermaths in the last decade have starkly underpinned the need to quantitatively assess the risk of an entire financial system rather than merely its individual entities. One of the first academic works on systemic risk is the seminal paper Eisenberg and Noe (2001). The focus of this work, however, lies on modeling the financial system rather than the quantitative measurement of systemic risk. Since then, financial mathematicians have developed a rich strand of literature, encompassing different approaches and emphasising different aspects of systemic risk. The model of Eisenberg and Noe (2001) has been generalized in different ways, for instance by considering illiquidity (Rogers & Veraart, 2013) or central clearing (Amini, Filipovic, & Minca, 2015). One strand of literature defines systemic risk measures by applying a scalar risk measure to the distribution of the total profits and losses of all firms in the system (Acharya, Pedersen, Philippon, & Richardson, 2016; Adrian & Brunnermeier, 2016). Recognizing the drawbacks of treating the economy as a portfolio, Chen, Iyengar, and Moallemi (2013) introduce an axiomatic approach to measuring systemic risk, further extended by Kromer, Overbeck, and Zilch (2016) and Hoffmann, Meyer-Brandis, and Svindland (2016). The axiomatic approach of Chen et al. (2013) is widely used and amounts to systemic risk measures of the form \( \rho(\Lambda(Y)) \), where \( Y \) is a \( d \)-dimensional random vector representing the financial system, \( \rho \) is a scalar risk measure and \( \Lambda: \mathbb{R}^d \to \mathbb{R} \) a non-decreasing aggregation function. However, this approach of aggregating first and
then adding a total capital requirement of the system, has the drawback that it results in the measurement of bailout costs rather than capital requirements that prevent a financial crisis. These types of risk measures are also called insensitive as they do not take into account the impact capital regulation has onto the system.

As an alternative, the so called sensitive systemic risk measures have been introduced by Feinstein et al. (2017); see also Biagini, Fouque, Frittelli, and Meyer-Brandis (2019) and Armenti, Crépey, Drapeau, and Papapantoleon (2018) for related approaches. There, one first adds the capital requirements to the financial institutions and then applies an aggregation function, that is, one considers systemic risk measures of the form

$$R(Y) = \{ k \in \mathbb{R}^d \mid \rho(\Lambda(Y + k)) \leq 0 \}.$$  

(1.1)

Thus, the impact of regulation onto the system is taken into account.

In this paper, we will mainly focus on this type of systemic risk measures as introduced in Feinstein et al. (2017); see Section 3 for more details. $R(Y)$ specifies the set of all capital allocations $k \in \mathbb{R}^d$ such that the new system $Y + k$ is deemed acceptable with respect to $\rho$ after being aggregated via $\Lambda$. As such, $R$ takes an ex ante perspective prescribing the injections to (and withdrawals from) each financial firm adequate to prevent the system $Y$ from a crises, whereas $\rho(\Lambda(Y))$ as described above can be interpreted as the bailout costs of the system after a systemic event has occurred.

1.2. Elicitability and identifiability

The field of quantitative risk management has seen a lively debate about which scalar risk measure is most appropriate in practice; see Embrechts, Puccetti, Rüschendorf, Wang, and Beleraj (2014) and Emmer, Kratz, and Tasche (2015) for detailed academic discussions and Bank for International Settlements (2014) for a regulatory perspective in banking. Besides differences in axiomatic properties such as coherence (Artzner et al., 1999) and convexity (Föllmer & Schied, 2002) of risk measures, the debate has also considered more statistical aspects of risk measures. The two most widely discussed statistical desiderata are robustness in the sense of Hampel (1971)—cf. Cont, Deguest, and Scandolo (2010), Krätzschmer, Schied, and Zähle (2012, 2014)—and elicitation of risk measures.

A real-valued law-invariant risk measure $\rho$—or, more generally, a functional $T$—is called elicitable if it can be written as the unique minimiser of an expected loss or scoring function $S: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, that is, $T(F) = \arg \min_x \int S(x, y) \, dF(y)$ (Lambert, Pennock, & Shoham, 2008; Osband, 1985). Such a scoring function, which incentivises truthful forecast from a risk-neutral forecaster, is called strictly consistent; see Subsection 2.1 for precise definitions. As such, the elicitation of a functional opens the way to meaningful forecast comparison (Gneiting, 2011a) which is closely related to comparative backtests in finance (Fissler, Ziegel, & Gneiting, 2016; Nolde & Ziegel, 2017). Similarly, the elicitation of a functional is crucial for $M$-estimation and regression, such as quantile regression (Koenker, 2005; Koenker & Basset, 1978) or expectile regression (Newey & Powell, 1987). The families of quantiles and expectiles are elicitable, and their most
prominent members, the median and the mean, have the absolute error, $S(x, y) = |x - y|$, and the squared error, $S(x, y) = (x - y)^2$, $x, y \in \mathbb{R}$, as strictly consistent scoring functions, respectively. Therefore, the prominent risk measure Value at Risk (VaR), basically corresponding to a quantile, is elicitable. On the other hand, Expected Shortfall, a tail expectation beyond a certain quantile and the most common counterpart in risk management and regulation, fails to have convex level sets (Gneiting, 2011a; Weber, 2006) which is a necessary condition for elicitability (Osband, 1985). This rules out the existence of a strictly consistent scoring function for Expected Shortfall (ES), imposing a major challenge to meaningfully compare ES forecasts. This issue can be overcome at the cost of additionally reporting VaR at the same level. More to the point, Fissler and Ziegel (2016) and Acerbi and Szekely (2014) showed that the pair (VaR, ES) is elicitable despite ES’s failure to have a strictly consistent scoring function on its own. A similar phenomenon was already known in the case of the elicitable pair (mean, variance), where variance clearly fails to have convex level sets rendering it non-elicitable. The key to establish the elicitation of (mean, variance) is that it is a bijection of the first two moments, which makes it an instance of the revelation principle (Gneiting, 2011a; Osband, 1985) establishing the elicitation of any bijection of an elicitable functional. In contrast, there is no known bijection of (VaR, ES) to a functional with elicitable components only. On the other hand, the positive elicitation result of (VaR, ES) has triggered a novel structural insight: the Bayes risk, or the minimal expected score, is jointly elicitable with the corresponding minimiser; see Frongillo and Kash (2018). The formalisation of how many auxiliary components are needed to render a functional $T$ a part of a higher dimensional elicitable $T'$ has led to the notion of forecast complexity; see again Frongillo and Kash (2018) and the discussion in Fissler and Ziegel (2016). For example, the risk measure Range Value at Risk has an elicitation complexity of 3 (Fissler & Ziegel, 2019a), while the mode functional generally possesses a corresponding complexity of $\infty$ (Dearborn & Frongillo, 2019; Heinrich, 2014). This means one needs to report the entire probability distribution, which amounts to considering the infinite dimensional identity functional (Brier, 1950; Good, 1952; Matheson & Winkler, 1976). We refer the reader to Gneiting and Raftery (2007) for a comprehensive overview of probabilistic forecast evaluation.

Closely related to the notion of elicitability is the concept of identifiability. While the former is useful for forecast comparison or model selection, the latter aims at model and forecast validation or verification. For real-valued forecasts, a (strict) identification function $V$ typically maps a forecast-observation pair $(x, y)$ to the real number $V(x, y)$ such that the expected identification function $\int V(x, y) dF(y)$ vanishes (only) at the correctly specified forecast $x = T(F)$. Typical examples are $V(x, y) = x - y$ for the mean or $V(x, y) = \mathbb{1}\{y \leq x\} - \alpha$, $x, y \in \mathbb{R}$, for the $\alpha$-quantile. In statistics and econometrics, identification functions are often known as moment functions and give rise to the (generalised) method of moments (Newey & McFadden, 1994). A functional is called identifiable if it possess a strict identification function. Steinwart, Pasin, Williamson, and Zhang (2014) showed that, under appropriate regularity conditions, the identifiability of a real-valued functional is equivalent to its elicitation. For probabilistic forecasts,
considering identification functions is akin to a goodness-of-fit analysis and, in particular, to assess various notions of calibration with tools such as the probability integral transform (Gneiting, Balabdaoui, & Raftery, 2007). For a discussion of identifiability and calibration in the context of assessing risk measures, we refer the reader to Davis (2016) and Nolde and Ziegel (2017).

1.3. Novel contributions and structure of the paper

The aim of the paper is two-fold: The primary goal is to establish elicitability and identifiability results for systemic risk measures of the form at (1.1), taking the form of subsets of $\mathbb{R}^d$. As outlined above, the literature on forecast evaluation has mainly focused on real-valued and vector-valued point forecasts as well as on probabilistic forecasts. While we could find various areas where forecasts of set-valued functionals are of interest (see Subsection 2.5 for a review), we are not aware of a comprehensive theoretical framework for the assessment of forecasts for set-valued functionals. The establishment of such a framework is therefore our secondary goal, which naturally precedes the presentation of the main results.

Section 2 formally introduces the notion of elicitability and identifiability for set-valued functionals. The key observation is that one should thoroughly distinguish whether one is interested in correctly specifying single points in the corresponding set-valued functional or if one is more ambitious and aims at correctly specifying the entire set. Mnemonically, we will label the former situation with the adjective selective, and we will use the term exhaustive to refer to the latter situation. Along with this distinction come two corresponding modes of elicitability and identifiability. Refining the classical result of the necessity of the convex level sets property for the elicitability of a functional (Proposition 2.10) leads the way to Corollary 2.13, the main result of Section 2: The two modes of elicitability are mutually exclusive—a set-valued functional is either selectively elicitable or exhaustively elicitable or not elicitable at all, subject to mild regularity conditions. Besides these structural results gathered in Subsection 2.4, Subsection 2.5 gives an overview of the existing literature on forecast evaluation for set-valued quantities, covering the fields of (spatial) statistics, machine learning, engineering, climatology and meteorology, and philosophy.

In Section 3, we discuss measures of systemic risk of the form at (1.1), gathering basic properties and assumptions. We also consider derived quantities, notably efficient cash-invariant allocation rules (EARs) introduced in Feinstein et al. (2017).

The main results about the identifiability and elicitability of systemic risk measures and derived quantities are gathered in Section 4. The most notable ones are Theorem 4.1 asserting the existence of oriented selective identification functions for $R_0(Y) = \{k \in \mathbb{R}^d | \Lambda(\rho(Y+k)) = 0\}$, and Theorem 4.9, which uses these identification functions to construct strictly consistent exhaustive scoring functions for $R$. That means while the first argument of the selective identification functions is a single capital allocation $k \in \mathbb{R}^d$ deemed suitable to exactly eliminate the risk of the financial system $Y$ under $R$, the first argument of the exhaustive scoring functions is a subset $A \subseteq \mathbb{R}^d$, consisting of all
capital allocations $k \in \mathbb{R}^d$ deemed appropriate to render $Y$ acceptable under $R$. Interestingly, these scoring functions arise as an integral construction of elementary scoring functions, exploiting the orientation of the identification function. This can be considered a higher-dimensional analogon to the mixture representation of scoring functions for one-dimensional forecasts established in the seminal paper Ehm, Gneiting, Jordan, and Krüger (2016). Similarly, this gives rise to the diagnostic tool of Murphy diagrams facilitating the assessment of forecast dominance; see Subsection 4.2.6. Characterisation results of all consistent scoring functions are discussed and order-sensitivity results of these consistent scoring functions are established. Concerning EARs mentioned above, Proposition 4.6 establishes strict selective identification functions for EARs, interestingly mapping to a function space.

Since systemic risk measures $R$ of the form at (1.1) are translation equivariant in the sense that $R(Y + k) = R(Y) - k$ for all $k \in \mathbb{R}^d$ and—under mild assumptions—homogeneous in that $R(cY) = c R(Y)$ for all $c > 0$ (Lemma 5.2), it makes sense to determine the subclasses of translation invariant or positively homogeneous consistent scoring functions for $R$, which is the content of Section 5.

The elicitability results of $R$ rely on the identifiability / elicitability of the underlying scalar risk measure $\rho$. This spells doom for the elicitability of systemic risk measures induced by ES as a scalar risk measure. Section 6 outlines this issue and establishes a solution to this challenge at the cost of a higher forecast complexity. Similarly to the scalar case, considering a pair of $R$ based on ES with a VaR-related quantity leads to selective identifiability and exhaustive elicitability results (Proposition 6.1 and Theorem 6.2).

The practical applicability of our results is demonstrated in terms of a simulation study, being the content of Section 7. Employing Diebold-Mariano tests, we examine how well the strictly consistent scores are able to distinguish different forecast performances. We also graphically illustrate the diagnostic tool of Murphy diagrams in a simulation example.

Section 8 closes the paper with a discussion and outlook of possible applications of our results and avenues of future research.

Results concerning risk measures insensitive with respect to capital allocations are collected in Appendix A. All proofs and purely technical results are deferred to Appendix B, while Appendix C contains some further graphics of simulation results in the context of Subsection 7.2.

2. Two modes of elicitability and identifiability

2.1. Consistent scoring functions for point-valued functionals

We use the decision-theoretic framework described for example in Gneiting (2011a); cf. Savage (1971), Osband (1985), Lambert et al. (2008), Fissler and Ziegel (2016, 2019c). Let $(\Omega, \mathcal{F}, P)$ be some complete, atomless probability space rich enough to accommodate all random elements mentioned in the sequel. We consider a time series $(Y_t)_{t \in \mathbb{N}}$ of obser-
vations of interest \( Y_t \) taking values in some measurable space \((O, \mathcal{O})\), called observation domain. Suppose there are \( m \in \mathbb{N} \) different forecasters, where, at time point \( t - 1 \), each forecaster \( i \in \{1, \ldots, m\} \) issues a one-step ahead forecast \( X_{i,t} \) for time point \( t \in \mathbb{N} \), taking values in an action domain \( A \), equipped with some \( \sigma \)-algebra \( A \). Describing the evolving information of each forecaster \( i \in \{1, \ldots, m\} \) in terms of a filtration \((\mathcal{G}_{i,t})_{t \in \mathbb{N}_0}\) and taking into account that the forecasts are non-anticipating, the time series \((X_{i,t})_{t \in \mathbb{N}}\) is \((\mathcal{G}_{i,t})_{t \in \mathbb{N}_0}\)-predictable while \((Y_t)_{t \in \mathbb{N}}\) is \((\mathcal{G}_{i,t})_{t \in \mathbb{N}_0}\)-adapted; cf. Strähl and Ziegel (2017).

In order to rank and compare the different sequences of forecasts, one commonly evaluates the corresponding prediction-observation sequences \((X_{i,t}, Y_t)_{t \in \mathbb{N}}\) in terms of a loss or scoring function. This is a measurable map \( S: A \times O \to \mathbb{R}^* := (-\infty, \infty] \) such that the prediction-observation pair \((x, y) \in A \times O\) is assigned the penalty \( S(x, y) \in \mathbb{R}^* \). For \( A = O = \mathbb{R} \), standard examples are the squared loss \( S(x, y) = (x - y)^2 \) or the absolute loss \( S(x, y) = |x - y| \). After \( N \geq 1 \) time steps, the forecasters are ranked in terms of their realised scores

\[
S_{i,N} = \frac{1}{N} \sum_{t=1}^{N} S(X_{i,t}, Y_t),
\]

\( i \in \{1, \ldots, m\} \). Invoking either a utility maximisation argument or a law of large numbers argument (under suitable mixing assumptions), each forecaster \( i \) has an incentive to minimise their expected realised score \( \mathbb{E}[S_{i,N}] \) over \((X_1, \ldots, X_N)\) such that \( X_{i,t} \) is \( \mathcal{G}_{i,t-1}\)-measurable (denoted by \( X_{i,t} \in \mathcal{G}_{i,t-1} \)). Hence, the optimal action, or Bayes-act, at each time point is given by

\[
X_{i,t}^* = \arg \min_{X_{i,t} \in \mathcal{G}_{i,t-1}} \mathbb{E}[S(X_{i,t}, Y_t)] = \arg \min_{X_{i,t} \in \mathcal{G}_{i,t-1}} \mathbb{E}[S(X_{i,t}, Y_t) | \mathcal{G}_{i,t-1}],
\]

if the (conditional) expectations exist.

The choice of the scoring function might be justified by an economically meaningful interpretation as a cost function. Alternatively, a directive for an ideal forecast might be given in terms of some statistical property of \( F_{Y_t|\mathcal{G}_{i,t-1}} \) (the regular version of) the conditional distribution of \( Y_t \) given \( \mathcal{G}_{i,t-1} \). Standard examples for such a property are the mean, some quantile, a risk-measure, or the probability distribution itself, giving rise to probabilistic forecasts. Mathematically speaking, the directive is given in terms of some functional \( T: \mathcal{M} \to A \), which is commonly a map from \( \mathcal{M} \), a space of probability distributions or distribution functions containing the conditional distributions of the form \( F_{Y_t|\mathcal{G}_{i,t-1}} \), to the action domain \( A \). In the latter setting, it is widely argued that the scoring function should incentivise truthful forecasts (Engelberg, Manski, & Williams, 2009; Murphy & Daan, 1985) in that

\[
\int S(T(F), y) dF(y) \leq \int S(x, y) dF(y)
\]  

\( ^1 \)Often in the literature, one considers only scoring functions that take finite values and which are assumed to have finite expectation for any argument \( x \). Including the value \( +\infty \) into the codomain however facilitates the technical treatment in the present context. Moreover, it also avoids problems for the certain widely used scoring functions such as the log-score for probabilistic forecasts; see Gneiting and Raftery (2007) and the discussion therein.
for all distributions $F \in \mathcal{M}$, for all $x \in \mathcal{A}$, and where equality implies that $x = T(F)$. In line with Fissler and Ziegel (2016) we call a scoring function $S$ strictly $\mathcal{M}$-consistent for $T$: $\mathcal{M} \to \mathcal{A}$ if it satisfies (2.2), where we implicitly assume that $\mathbb{E}_F[S(x, Y)] := \int S(x, y) dF(y)$ exists for all $x \in \mathcal{A}$, $F \in \mathcal{M}$ and takes a value in $\mathbb{R}^+$. For the sake of brevity, we shall henceforth use the shorthand $\bar{S}(x, F) := \mathbb{E}_F[S(x, Y)]$. Following Lambert et al. (2008) and Gneiting (2011a), we call a functional $T$: $\mathcal{M} \to \mathcal{A}$ elicitable if it possesses a strictly $\mathcal{M}$-consistent scoring function.

### 2.2. Selective and exhaustive scoring functions

Many statistical functionals, such as the mean-functional or variance, are point-valued, taking values in $\mathbb{R}$ for univariate observations (or in $\mathbb{R}^d$ and $\mathbb{R}^{d \times d}$ for multivariate observations). In such a situation, possible actions take the form of points, such that $\mathcal{A} \subseteq \mathbb{R}$ (or of vectors or matrices, respectively, such that $\mathcal{A} \subseteq \mathbb{R}^d$ or $\mathcal{A} \subseteq \mathbb{R}^{d \times d}$). Then the functional can be considered as a map $T$: $\mathcal{M} \to \mathcal{A}$, and in particular, there is a unique best action, namely the functional value $T(F) \in \mathcal{A}$ for $F \in \mathcal{M}$. Moreover, there is no ‘severe ambiguity’ of how to choose the action domain $\mathcal{A}$. That is, any reasonable choice of $\mathcal{A}$ will be that of a subset from the same space. E.g. if one considers mean forecasts for a real-valued observation, then $\mathcal{A} \subseteq \mathbb{R}$. Maybe it makes sense to consider $\mathcal{A} \subseteq [0, \infty)$ if one knows that the observation is non-negative, but it would not be reasonable at all to choose $\mathcal{A} = \mathbb{R}^2$.

On the other hand, there are also statistical functionals that are inherently set-valued. Examples for such a situation can be manifold, e.g. expectations of random sets (Molchanov, 2017), where real-world examples of random sets might stem 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). The main examples of set-valued functionals we have in mind are on the one hand law-invariant measures of systemic risk $R$: $\mathcal{M} \to 2^{\mathbb{R}^d}$ introduced in Feinstein et al. (2017). On the other the hand, we follow option (ii) of the interesting discussion provided in Mizera (2010, p. 170), and will consider quantiles to be set-valued. That is, for some $\alpha \in (0, 1)$, the $\alpha$-quantile $q_\alpha$ is defined as

$$q_\alpha: \mathcal{M} \to 2^\mathbb{R}, \quad q_\alpha(F) := \{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leq \alpha \leq F(x)\}.$$  

To unify notation, set-valued functionals take the form

$$T: \mathcal{M} \to 2^W,$$

where $W$ is some generic space. Now, there are two sensible choices for the action domain $\mathcal{A}$:

(i) $\mathcal{A} = \mathcal{A}_{sel} \subseteq W$: The elements of the action domain $\mathcal{A}_{sel}$ representing possible forecasts are points in the space $W$. Truthful reporting means there are generally multiple best actions, namely all selections $t \in T(F) \subseteq \mathcal{A}_{sel}$ for $F \in \mathcal{M}$. Mnemonically, we shall refer to $\mathcal{A}_{sel}$ as a selective action domain.


The elements of the action domain $A_{\text{exh}}$ representing possible forecasts are subsets of the space $W$. Truthful reporting means there is a unique best action, namely the exhaustive functional $T(F) \in A_{\text{exh}}$ for $F \in M$. Similarly, we shall refer to $A_{\text{exh}}$ as an exhaustive action domain.

The two different choices of action domains lay claim to different levels of precision and ambition of the forecasts. Whereas selective forecasts from $A_{\text{sel}}$ are content to specifying one single point in the set of interest—e.g. one location affected by a flood, one tumorous cell, one capital allocation making a financial system acceptable, or one point such that the distribution function attains a certain level $\alpha$—, exhaustive forecasts from $A_{\text{exh}}$ aim at specifying the entire set of interest simultaneously. E.g. they specify the entire expected region affected by a flood, the entire expected tumorous tissue, the entire set of acceptable capital allocations, or the whole set of values such that the distribution function attains a certain level $\alpha$.

For a certain function $T: M \to 2^W$, 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 one might have in mind is

$$A_{\text{sel}} = \bigcup_{B \in A_{\text{exh}}} B.$$ 

To allow for a rigorous treatment of forecast evaluation for set-valued functionals, we continue to use the dichotomy introduced above also for scoring functions evaluating forecasts for some set-valued functional $T: M \to 2^W$.

**Definition 2.1** (Consistency). (i) A selective scoring function $S_{\text{sel}}: A_{\text{sel}} \times O \to \mathbb{R}^*$ is $\mathcal{M}$-consistent for $T: M \to 2^{A_{\text{sel}}}$ 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 $S_{\text{sel}}$ is strictly $\mathcal{M}$-consistent for $T$ if it is $\mathcal{M}$-consistent for $T$ and if equality in (2.3) implies that $x \in T(F)$.

(ii) An exhaustive scoring function $S_{\text{exh}}: A_{\text{exh}} \times O \to \mathbb{R}^*$ is $\mathcal{M}$-consistent for $T: M \to A_{\text{exh}}$ 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 $S_{\text{exh}}$ is strictly $\mathcal{M}$-consistent for $T$ if it is $\mathcal{M}$-consistent for $T$ and if equality in (2.4) implies that $B = T(F)$.

Note that the strict consistency of a selective (exhaustive) scoring function $S_{\text{sel}}$ 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 $F \in \mathcal{M}$). Along with the definitions of the two modes of consistency come two ways of defining elicitability.

**Definition 2.2** (Elicitability). (i) A functional $T: M \to 2^{A_{\text{sel}}}$ is selectively elicitable if it possesses a strictly $\mathcal{M}$-consistent selective scoring function $S_{\text{sel}}: A_{\text{sel}} \times O \to \mathbb{R}^*$. \hfill 10
(ii) A functional $T: \mathcal{M} \to A_{\text{exh}}$ is \textit{exhaustively elicitable} if it possesses a strictly $\mathcal{M}$-consistent exhaustive scoring function $S_{\text{exh}}: A_{\text{exh}} \times \mathcal{O} \to \mathbb{R}^*$.

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 point-valued functionals such as the mean, the distinction between selective and exhaustive elicitation 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 identify a point-valued functional $T: \mathcal{M} \to A$ with the set-valued functional $T': \mathcal{M} \to A' = \{\{a\} | a \in A\}$ where $T'(F) = \{T(F)\}$. Then clearly, the exhaustive and selective elicitation of $T'$ are equivalent, and they are equivalent to the elicitation of $T$; see Lemma 2.6.

While we are only aware of contributions to the literature which either consider the selective or the exhaustive interpretation (see Subsection 2.5), one novelty in the present paper is that we thoroughly study and compare these two alternative notions (see Subsection 2.4).

### 2.3. Selective and exhaustive identification functions

The notion of identifiability is closely connected to the notion of elicitation. Similarly to scoring functions, an identification function is a measurable map $V: A \times \mathcal{O} \to \mathbb{R}$ where we again make the tacit assumption that $\bar{V}(x,F) := E[F][V(x,F)] := \int V(x,y)dF(y)$ exists for all $x \in A, F \in \mathcal{M}$ with the additional assumption that the expectation be finite. Expectations of strictly consistent scoring functions are minimised by the correctly specified forecast for the functional at hand. Likewise, the set of zeros of the expectation of a strict identification function coincides with the correctly specified forecast for the functional at hand. Again, we make the distinction between selective and exhaustive identification functions to allow for a rigorous treatment of set-valued functionals.

**Definition 2.3 (Identification function).**

(i) A map $V_{\text{sel}}: A_{\text{sel}} \times \mathcal{O} \to \mathbb{R}$ is a \textit{selective $\mathcal{M}$-identification function} for $T: \mathcal{M} \to 2^{A_{\text{sel}}}$ if $\bar{V}_{\text{sel}}(t,F) = 0$ for all $t \in T(F)$ and for all $F \in \mathcal{M}$. Moreover, $V_{\text{sel}}$ is a strict selective $\mathcal{M}$-identification function for $T$ if

\[
\bar{V}_{\text{sel}}(x,F) = 0 \iff x \in T(F), \quad \forall x \in A_{\text{sel}}, \quad \forall F \in \mathcal{M}. \quad (2.5)
\]

(ii) A map $V_{\text{exh}}: A_{\text{exh}} \times \mathcal{O} \to \mathbb{R}$ is an \textit{exhaustive $\mathcal{M}$-identification function} for $T: \mathcal{M} \to A_{\text{exh}}$ if $\bar{V}_{\text{exh}}(T(F),F) = 0$ for all $F \in \mathcal{M}$. Moreover, $V_{\text{exh}}$ is a strict exhaustive $\mathcal{M}$-identification function for $T$ if

\[
\bar{V}_{\text{exh}}(B,F) = 0 \iff B = T(F), \quad \forall B \in A_{\text{exh}}, \quad \forall F \in \mathcal{M}. \quad (2.6)
\]

\[\text{At the costs of a more involved technical treatment we could also consider identification functions mapping to } [-\infty, \infty]. \text{ However, we do not see the benefits in the present context which is why we omit this approach.}\]
Definition 2.4 (Identifiability). (i) A functional $T: \mathcal{M} \to \mathbb{2}^{\mathcal{A}_{\text{sel}}}$ is selectively identifiable if it possesses a strict selective $\mathcal{M}$-identification function.

(ii) A functional $T: \mathcal{M} \to \mathcal{A}_{\text{exh}}$ is exhaustively identifiable if it possesses a strict exhaustive $\mathcal{M}$-identification function.

In the literature about point-valued functionals, it has appeared to be appropriate to be a bit more flexible with the choice of the space the identification functions map to. In particular, Osband (1985), Frongillo and Kash (2015b) and Fissler and Ziegel (2016) suggest that the dimension of the identification function should coincide with the dimension of the forecasts. Indeed, statistical practice demands to evaluate the realised identification function, which can be seen as the counterpart of (2.1) upon replacing $S$ with $V$. To this end, $V$ needs to map to some linear space such as some functional space (Proposition 4.6). Moreover, note that in order to define a property in the spirit of (2.5) or (2.6), it is not essential that the expected identification function attains a 0 at the correctly specified forecast (or an element thereof in the selective setting), but rather that it attains some predefined particular value(s)—the important requirement being that this value be identifiable in the common sense; see Proposition 4.6 where we introduce a sort of generalised version of (selective) identifiability where the left-hand side of (2.5) is replaced by some more general statement.

Similarly to elicitability, when dealing with point-valued functionals, the distinction between selective and exhaustive identifiability is rather artificial and they essentially coincide; see Lemma 2.6.

According to Steinwart et al. (2014), we say that a strict identification function $V: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ for a real-valued functional $T: \mathcal{M} \to \mathbb{R}$ is called oriented if for all $x \in \mathbb{R}$ and for all $F \in \mathcal{M}$

$$V(x, F) \begin{cases} < 0, & \text{if } x < T(F) \\ = 0, & \text{if } x = T(F) \\ > 0, & \text{if } x > T(F). \end{cases}$$

They also show that under some continuity assumptions on the functional $T$ there exists an oriented identification function for $T$.

2.4. Structural results

In this subsection, we gather structural relations between the notions of selective and exhaustive elicitability and identifiability, respectively. The first lemma, of which the proof is standard, basically states that these notions are equivalent for point-valued functionals.

Lemma 2.5. Let $T: \mathcal{M} \to \mathcal{A}$ be some point-valued functional taking values in some space $\mathcal{A}$. Define the set-valued functional $T'(F) := \{T(F)\}$, $F \in \mathcal{M}$. Then $T': \mathcal{M} \to 2^{\mathcal{A}}$ is selectively elicitable (identifiable) if and only if $T': \mathcal{M} \to 2^{\mathcal{A}'} = \{\{a\} | a \in \mathcal{A}\}$ is exhaustively elicitable (identifiable). Moreover, the selective elicitability (identifiability) of $T': \mathcal{M} \to 2^{\mathcal{A}}$ is equivalent to the elicitability (identifiability) of $T$. 

12
The next lemma is concerned with selections of set-valued functionals. For a set \( W \neq \emptyset \) and a set-valued functional \( T': M \to 2^W \), a point-valued functional \( T: M \to W \) is a selection of \( T' \) if \( T(F) \in T'(F) \) for all \( F \in M \).

**Lemma 2.6.** Let \( T': M \to 2^A \) be a set-valued functional with \( T'(F) \neq \emptyset \) for all \( F \in M \), and let \( T: M \to A \) be a selection of \( T' \).

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

(ii) If \( V: A \times O \to \mathbb{R} \) is a strict selective \( M \)-identification function for \( T' \), then it is an \( M \)-identification function for \( T \).

Clearly, the scoring function \( S \) (identification function \( V \)) appearing in Lemma 2.6 is only strict for \( T \) if \( T' \) is a singleton on \( M \), that is, \( T'(F) = \{T(F)\} \) for all \( F \in M \). This suggests the question as to whether the selection can be elicitable (identifiable) at all, which the following proposition is concerned with.

**Proposition 2.7.** Let \( T': M \to 2^A \) be selectively elicitable and \( T: M \to A \) a selection of \( T' \). Let \( M_1 := \{F \in M | T'(F) = \{T(F)\}\} \) and suppose that \( M \setminus M_1 \neq \emptyset \).

Let \( S'_M, (S'_A) \) be the class of strictly \( M \)-consistent (\( M \)-consistent) selective scoring functions for \( T' \). If \( S'_M = S'_A \), then \( T: M \to A \) is not elicitable.

A common problem when applying Proposition 2.7 for practical purposes is that most characterisation results concerning the class of strictly consistent scoring functions assume regularity conditions on the scoring functions such as continuity or differentiability; cf. Table 1 in Gneiting (2011b) or Osband’s Principle (Fissler & Ziegel, 2016; Osband, 1985). Note that Proposition 2.16 establishes an alternative route to a similar result which does not rely on such a characterisation result of the class of strictly consistent scoring functions.

While for point-valued functionals mapping to \( A = \mathbb{R} \) and satisfying sufficient regularity conditions, identifiability and elicitation are equivalent (Steinwart et al., 2014), this is not always the case if one weakens the regularity conditions.\(^3\) A classical result dating back to the seminal work of Osband (1985) is that convex level sets (CxLS) of a functional are necessary both for elicitation and identifiability; see also Fissler and Ziegel (2019c) for some refinements. Steinwart et al. (2014) showed that for point-valued functionals, the CxLS property is even sufficient for elicitation under some additional regularity conditions. Continuing with the consequent distinction between selective and exhaustive elicitation (identifiability) for set-valued functionals, we state the corresponding CxLS properties.

**Definition 2.8.** Let \( T: M \to 2^W \) be a set-valued functional on some convex class of measures \( M \).\(^4\)

\(^3\)E.g., the lower \( \alpha \)-quantile is elicitable relative to the class of strictly increasing distribution functions. However, it fails to be identifiable if some of the distributions are discontinuous in their \( \alpha \)-quantile. On the other hand, the canonical scoring function \( S(x, y) = (1 \{y \leq x\} - \alpha)(x - y) \) is still strictly consistent on that class. The discontinuity only implies that the expected scores are not differentiable.

\(^4\)That is, for any \( F_0, F_1 \in M \) and for any \( \lambda \in (0, 1) \) it holds that \( (1 - \lambda)F_0 + \lambda F_1 \in M \).
(i) \( T \) has the selective CxLS property if for any \( F_0, F_1 \in \mathcal{M} \) and \( \lambda \in (0, 1) \)
\[ T(F_0) \cap T(F_1) \subseteq T((1 - \lambda)F_0 + \lambda F_1). \]

(ii) \( T \) has the selective CxLS* property if for any \( F_0, F_1 \in \mathcal{M} \) and \( \lambda \in (0, 1) \)
\[ T(F_0) \cap T(F_1) \neq \emptyset \implies T(F_0) \cap T(F_1) = T((1 - \lambda)F_0 + \lambda F_1). \]

(iii) \( T \) has the exhaustive CxLS property if for any \( F_0, F_1 \in \mathcal{M} \) and \( \lambda \in (0, 1) \)
\[ T(F_0) = T(F_1) \implies T(F_0) = T((1 - \lambda)F_0 + \lambda F_1). \]

Note that the exhaustive CxLS property is the most common one in the literature, and the one used for point-valued functionals (Bellini & Bignozzi, 2015; Delbaen, Bellini, Bignozzi, & Ziegel, 2016; Steinwart et al., 2014; Wang & Wei, 2018). 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 Strokorb (2019) introduced the notion of max-functionals. Using our notation, a real-valued functional \( T: \mathcal{M} \to \mathbb{R} \) 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 \( T: \mathcal{M} \to \mathbb{R} \) 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.

**Lemma 2.9.** Let \( T: \mathcal{M} \to 2^\mathcal{W} \) be a set-valued functional on some convex class of measures \( \mathcal{M} \).

(i) If \( T \) has the selective CxLS* property, then it also has the selective and the exhaustive CxLS property.

(ii) If \( T \) is singleton-valued, then the selective CxLS property, the exhaustive CxLS property and the selective CxLS* property are equivalent.

The second point of Lemma 2.9 underpins why the distinction of the CxLS properties is obsolete for the point-valued case.

It is classical knowledge originating from the seminal work of Osband (1985) that the exhaustive CxLS property is necessary for exhaustive elicitation and exhaustive identifiability, and that the selective CxLS property is necessary for selective elicitability and selective identifiability. What is novel is that the selective CxLS* property is necessary for selective elicitability.

**Proposition 2.10.** Let \( T: \mathcal{M} \to 2^{\mathcal{A}_{\text{sel}}} \) be a set-valued functional on some convex class of measures \( \mathcal{M} \). If \( T \) is selectively elicitable, then it satisfies the selective CxLS* property.

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.
Definition 2.11. A set-valued functional $T: \mathcal{M} \to 2^W$ has the proper-subset property if there are $F, G \in \mathcal{M}$ such that
$$\emptyset \neq T(G) \subset T(F).$$

Theorem 2.12. Let $T: \mathcal{M} \to A_{\text{exh}}$ be a set-valued functional on a convex class $\mathcal{M}$. If $T$ satisfies the proper-subset property and the selective $C\text{xLS}^*$ property, then it is not exhaustively elicitable.

Proposition 2.10 and Theorem 2.12 establish that the notion of selective and exhaustive elicitability are mutually exclusive under the mild proper-subset property.

Corollary 2.13 (Mutual exclusivity). Let $\mathcal{M}$ be a convex class of distributions and $T: \mathcal{M} \to A_{\text{exh}} \subseteq 2^{A_{\text{sel}}}$ a set-valued functional with the proper-subset property. Then it holds that
(i) if $T$ is selectively elicitable, it is not exhaustively elicitable;
(ii) if $T$ is exhaustively elicitable, it is not selectively elicitable.

This structural insight is an entirely novel and interesting result. It basically establishes that there are only three pairwise disjoint classes 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.

Some examples are in order.

Example 2.14. (i) Any $\alpha$-quantile is selectively elicitable. If the class $\mathcal{M}$ is reasonably large (e.g. it contains all measures with finite support), then the $\alpha$-quantile clearly satisfies the proper-subset property. Hence, they fail to be exhaustively elicitable.

(ii) If $\mathcal{M}$ is the class of distributions on $\mathbb{R}$ with finite support, then the mode functional is selectively elicitable on $\mathcal{M}$ with the strictly $\mathcal{M}$-consistent selective scoring function $S(x, y) = I\{x \neq y\}$ (Gneiting, 2017; Heinrich, 2014). Since the mode functional satisfies the proper-subset property on $\mathcal{M}$, it also fails to be exhaustively elicitable on $\mathcal{M}$.

(iii) In Theorem 4.9 we establish the exhaustive elicitability of the set-valued systemic risk measure $R$ defined at (3.1). The cash-invariance property of $R$ implies that it satisfies the proper-subset property. This means $R$ cannot be selectively elicitable.

(iv) Any elicitable real-valued functional $T: \mathcal{M} \to \mathbb{R}$ induces trivial set-valued functionals $T^-(F) := (-\infty, T(F)]$ and $T^+(F) = [T(F), \infty)$. Clearly, the elicitability of $T$ is equivalent to the exhaustive elicitability of $T^-$ and $T^+$ considered as maps to $A^- = \{(-\infty, x] \mid x \in \mathbb{R}\}$ and $A^+ = \{(x, \infty) \mid x \in \mathbb{R}\}$, e.g. by invoking the revelation principle (Fissler, 2017; Gneiting, 2011a; Osband, 1985). If $T$ is not constant on $\mathcal{M}$, then $T^-$ and $T^+$ also satisfy the proper-subset property, which means they violate the selective $C\text{xLS}^*$ 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 Strokorb (2019) such that $T$ (and $-T$) is not elicitable unless it is constant. This recovers Theorem 3.3 in Brehmer and Strokorb (2019) and shows that it is, indeed, a special case of Theorem 2.12.

Remark 2.15. It is worth noting the structural difference between elicitation and identifiability. While Theorem 2.12 carries over to exhaustive identifiability with an easy adaption of the proof, it does not seem to be possible to establish an analogon of Proposition 2.10 for selective identifiability due to possible cancellation effects. One can merely establish that selective identifiability implies the selective CxLS property. Therefore, it remains open if selective and exhaustive identifiability are mutually exclusive in the sense of Corollary 2.13.

Let us take another look at selections of selectively elicitable functionals $T$, such as the lower quantile if $T$ is the quantile functional. Proposition 2.7 rules out the elicitation of such selections. However, the practical applicability of this result seems to be somewhat limited since one needs to know the class of all strictly consistent scoring functions for $T$ and the corresponding selection. And these characterisation results typically impose regularity conditions on the scoring functions—if they are known at all. Interestingly, an argument similar to the one used in the proof of Theorem 2.12 leads to a result which rules out the elicitation of selections 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 to some topological space $A$ mixture-continuous if for any $F_0,F_1 \in \mathcal{F}$ the map $[0,1] \ni \lambda \mapsto T((1-\lambda)F_0 + \lambda F_1) \in A$ is continuous.

Proposition 2.16. Let $\mathcal{M}$ be a convex class of distributions and $T: \mathcal{M} \rightarrow 2^A$, $T \neq \emptyset$, a functional satisfying the selective CxLS* property. Suppose there are distributions $F,G,H \in \mathcal{M}$ such that

$$T(F) \cap T(G) = \{t_1\}, \quad T(F) \cap T(H) = \{t_2\}, \quad \text{with } t_1 \neq t_2. \quad (2.7)$$

Then, any selection $T_{sel}: \mathcal{M} \rightarrow A$ of $T$ fails to be elicitable and identifiable. In particular, if $A$ is a space with a Fréchet topology, then any selection $T_{sel}$ fails to be mixture-continuous.

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

\footnote{For $\emptyset \neq T(G) \subseteq T(F)$ and a strict exhaustive $\mathcal{M}$-identification function $V$ for $T$ the CxLS* property implies that

$$0 = \bar{V}(T(G),(1-\lambda)F + \lambda G) = (1-\lambda)\bar{V}(T(G),F) + \lambda \bar{V}(T(G),G) = (1-\lambda)\bar{V}(T(G),F) \neq 0$$

for any $\lambda \in (0,1)$, which is a contradiction.}

\footnote{That is, if for any $a,b \in A$ with $a \neq b$ there is an open set $U \subseteq A$ such that $a \in U$ and $b \notin U$.}
We would like to remark that the $\alpha$-quantile satisfies the condition at (2.7) if the class $\mathcal{M}$ is reasonably large, e.g. contains all distributions with finite support. Therefore, Proposition 2.16 rules out the elicitability of any selection of the quantile on such classes, e.g. the lower quantile or the selection introduced in the recent preprint Aronow and Lee (2018).

2.5. Literature on forecast evaluation for set-valued functionals

2.5.1. Statistical forecast evaluation

While Lambert et al. (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. Bellini and Bignozzi (2015). However, only the concept of selective scoring functions with the corresponding notion of (strict) consistency and elicitability 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 (Gneiting, 2011b; Koenker, 2005; Komunjer, 2005), in which they are elicitable. The reason for not considering them in the exhaustive sense might lie in the impossibility of establishing corresponding elicitability results, of which the first formal proof—to the best of our knowledge—is given in this paper.

On the other hand, 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. Gneiting and Raftery (2007, Sections 6.2 and 9.3) consider consistent exhaustive scores for the central $(1 - \alpha)$-prediction interval. This basically amounts to a prediction for a pair of quantiles at the $\alpha/2$- and $(1 - \alpha/2)$-level. If one fixes a certain coverage of, say, $1 - \alpha$, this ansatz can be generalised to construct consistent scoring functions for a non-central $(1 - \alpha)$-prediction interval of which the endpoints are specified in terms of quantiles at level $\beta$ and $\beta + 1 - \alpha$, where $\beta \in (0, \alpha)$. 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. Finally, we would like to refer the reader to Askanazi, Diebold, Schorheide, and Shin (2018) for a good overview of interval forecasts, who, however, mostly present impossibility results. While the complexity of reporting interval forecasts is quite modest and actually amounts to specifying a two-dimensional vector, our results, and in particular the mutual exclusivity result of Corollary 2.13, provide a novel insight in 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.

2.5.2. Statistical theory and risk measurement

Quantiles and expectiles (Newey & Powell, 1987) of univariate distributions are well known (selectively) elicitable functionals. In the risk measure literature, 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.

2.5.3. Spatial statistics

As Azzimonti, Ginsbourger, Chevalier, Bect, and Richet (2018) 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). That means, the quantity of interest \( Y \) is a random set which is often specified as an excursion set \( \{ z \in \mathbb{R}^d \mid \xi_z \geq t \} \), \( t \in \mathbb{R} \), of some random field \( (\xi_z)_{z \in \mathbb{R}^d} \). Functionals of interest are often various expectations of random sets as described in the comprehensive textbook Molchanov (2017), notably, the Vorob’ev expectation (Chevalier, Ginsbourger, Bect, & Molchanov, 2013), the distance average expectation (Azzimonti, Bect, Chevalier, & Ginsbourger, 2016) and conservative estimates based on Vorob’ev quantiles (Azzimonti et al., 2018). The most common evaluation metric—or scoring function—seems to be the symmetric distance in measure \( S(X,Y) = \mu(X \triangle Y) \), where \( \mu \) is some measure on \( (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \), but also (pseudo-)metrics such as the Hausdorff distance and others (Molchanov, 2017, pp. 286–290) might make sense. Interestingly, the distance in measure admits a straightforward interpretation as the sum of false positive and false negative events via the identity \( \mu(X \triangle Y) = \mu(X \setminus Y) + \mu(Y \setminus X) \), which suggests also asymmetric versions of this measure, taking the form \( (1 - \alpha)\mu(X \setminus Y) + \alpha \mu(Y \setminus X) \), \( \alpha \in [0, 1] \).

While we are unaware of a thorough discussion of exhaustive elicitability of functionals of random sets, we would like to mention that the Vorob’ev median (Molchanov, 2017, p. 285) possesses at least a consistent exhaustive scoring function with \( S(X,Y) = \mu(X \triangle Y) \) for some measure \( \mu \) on \( (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \) (Molchanov, 2017, Proposition 2.2.8),\(^7\) and that the Vorob’ev mean is a solution to a restricted minimisation problem with respect to the same scoring function (Molchanov, 2017, Theorem 2.2.6).

Ascending to a higher level, the collection of all Fréchet means—as a collection of sets (Molchanov, 2017, Definition 2.2.18)—is selectively elicitable by definition.

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

\(^7\)This suggests a similar relation between the Vorob’ev quantile and the asymmetric distance in measure presented above.
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.

### 2.5.4. 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 (2019), where the two papers consider these functionals in the selective sense.

In the area of machine learning, the recent paper Gao, Chen, Chenthamarakshan, and Witbrock (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 finite. Denoting the regressors with $X_t \in \mathbb{R}^p$ then they are interested in finding a function $m: \mathbb{R}^p \to \{I | I \subseteq S, |I| < \infty\}$ 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.

### 2.5.5. 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.

### 3. Measures of systemic risk

We consider set-valued systemic risk measures studied in Feinstein et al. (2017). In particular, we concentrate on law-invariant risk measures $R$ that are induced by some law-invariant scalar risk measure $\rho$. To settle some notation, let $(\Omega, \mathcal{F}, \mathbb{P})$ be an atomless probability space. For some integer $d \geq 1$, let $\mathcal{Y}^d \subseteq L^0(\Omega; \mathbb{R}^d)$ be some subclass of $d$-dimensional random vectors. From a risk management perspective, a random vector $Y = (Y_1, \ldots, Y_d) \in \mathcal{Y}^d$ represents the respective gains and losses of a system of $d$ financial
firms. That is, positive values of the component $Y_i$ represent gains of firm $i$ and negative values correspond to losses. Let $\mathcal{M}^d$ be the class of probability measures of elements of $\mathcal{Y}^d$. Let $\Lambda: \mathbb{R}^d \to \mathbb{R}$ be an aggregation function meaning that it is non-decreasing with respect to the componentwise order. An aggregation function is typically, but not necessarily, assumed to be continuous or even concave. We introduce $\mathcal{Y} \subseteq L^0(\Omega; \mathbb{R})$ where $\{\Lambda(Y) \mid Y \in \mathcal{Y}^d\} \subseteq \mathcal{Y}$ and let $\mathcal{M}$ be the class of distributions of elements of $\mathcal{Y}$. Where convenient we will tacitly assume that $\mathcal{Y}^d$ and $\mathcal{Y}$ are closed under translation meaning that $X \in \mathcal{Y}, Y \in \mathcal{Y}^d$ implies that $X + m \in \mathcal{Y}$ and $Y + k \in \mathcal{Y}^d$ for all $m \in \mathbb{R}, k \in \mathbb{R}^d$.

We consider some scalar monetary law-invariant risk measure $\rho: \mathcal{Y} \to \mathbb{R}^8$ (Artzner et al., 1999). That means, we can alternatively consider $\rho$ as a map $\rho: \mathcal{M} \to \mathbb{R}$ such that for a random variable $X \in \mathcal{Y}$ with distribution $F_X \in \mathcal{M}$ we define $\rho(F_X) := \rho(X)$. We assume that $\rho$ is cash-invariant, that is, $\rho(X + m) = \rho(X) - m$ for all $m \in \mathbb{R}$ and all $X \in \mathcal{Y}$, and monotone, meaning $X \geq Z \; \mathbb{P}$-a.s. implies that $\rho(X) \leq \rho(Z)$ for all $X, Z \in \mathcal{Y}$. We often dispense with the usual normalisation assumption that $\rho(0) = 0$. We present the two most natural law-invariant set-valued measures of systemic risk that are based on $\rho$ and $\Lambda$, namely

\[ R: \mathcal{Y}^d \to 2^{\mathbb{R}^d}, \quad \mathcal{Y} \mapsto R(Y) = \{k \in \mathbb{R}^d \mid \rho(\Lambda(Y + k)) \leq 0\}; \tag{3.1} \]
\[ R^{ins}: \mathcal{Y}^d \to 2^{\mathbb{R}^d}, \quad \mathcal{Y} \mapsto R^{ins}(Y) = \{k \in \mathbb{R}^d \mid \rho(\Lambda(Y) + k) \leq 0\}. \tag{3.2} \]

In (3.2) and later on, we have used the shorthand $\bar{k} := \sum_{i=1}^d k_i$ for some vector $k = (k_1, \ldots, k_d) \in \mathbb{R}^d$. Note the difference between $R$ and $R^{ins}$. The risk measure $R$ takes an \textit{ex ante} perspective in the sense that it specifies all capital allocations $k \in \mathbb{R}^d$ needed to be added to the system $Y$ to make the aggregated system $\Lambda(Y + k)$ acceptable under $\rho$. On the other hand, $R^{ins}$ takes an \textit{ex post} perspective on quantifying the risk of the system $Y$. That means it first considers the current aggregated system $\Lambda(Y)$ and then specifies the \textit{total} capital requirement $\bar{k}$ one needs to add to make the aggregated system acceptable, which amounts to specifying the \textit{bail-out costs} of the aggregated system $\Lambda(Y)$ under $\rho$. In particular, the risk measure $R^{ins}$ is \textit{insensitive} to the capital allocation to each financial firm, disregarding possible transaction costs or other dependence structures between the financial firms. This justifies the mnemonic terminology. We would like to remark that both risk measures, $R$ and $R^{ins}$, can be of interest in applications, taking into regard the different perspectives on systemic risk. However, the mathematical treatment and complexity differ considerably: Due to the cash-invariance of $\rho$, $R^{ins}$ takes the equivalent form

\[ R^{ins}(Y) = \{k \in \mathbb{R}^d \mid \rho(\Lambda(Y)) \leq \bar{k}\}. \]

This means that $R^{ins}$ is actually a bijection of the scalar risk measure $\rho \circ \Lambda: \mathcal{M}^d \to \mathbb{R}$ considered in Chen et al. (2013). Therefore, one has to evaluate the risk measure $\rho$ only once to determine $R^{ins}$. In contrast, such an appealing equivalent formulation is\footnote{Very often, the scalar risk measure is assumed to map to $\mathbb{R}^\star = (-\infty, \infty]$. We could also do that at the costs of a more technical treatment. However, to avoid unnecessary technicalities, we refrain from that and will assume throughout the paper that any scalar risk measure will attain real values only.}
generally not available for $R$, unless $\Lambda$ is additive, or is even the sum in which case $R$ and $R^{ins}$ coincide. Consequently, in general, one is bound to evaluate $\rho$ infinitely often to compute $R$; see also the discussion in Feinstein et al. (2017). The main focus of this paper are elicitation and identifiability results of systemic risk measures of the form at (3.1) and (3.2). However, since one can exploit the one-to-one relation between $R^{ins}$ and $\rho \circ \Lambda$ and make use of the revelation principle (Fissler, 2017; Gneiting, 2011a; Osband, 1985) to establish (exhaustive) elicitation and identifiability results, we do not present results about $R^{ins}$ in the main body of the paper, but rather gather some interesting observations in Appendix A.

For the sake of completeness, we evoke the most important properties of $R$ presented in Feinstein et al. (2017). Due to the properties that $\rho$ is cash-invariant and that $\Lambda$ is increasing, we obtain that the values of $R$ defined at (3.1) are upper sets. That means for any $Y \in \mathcal{Y}^d$, $R(Y) = R(Y) + \mathbb{R}^d$, where $\mathbb{R}^d_+ := \{x \in \mathbb{R}^d \mid x_1, \ldots, x_d \geq 0\}$ and where for any two sets $A, B \subseteq \mathbb{R}^d$, $A + B := \{a + b \mid a \in A, b \in B\}$ denotes the usual Minkowski sum. Mutatis mutandis, the same is true for $R^{ins}$. Following the notation of Feinstein et al. (2017), we denote the collection of upper sets in $\mathbb{R}^d$ with ordering cone $\mathbb{R}^d_+$ as

$$\mathcal{P}(\mathbb{R}^d; \mathbb{R}^d_+) := \{B \subseteq \mathbb{R}^d \mid B = B + \mathbb{R}^d_+\}.$$ 

Note that both $\mathbb{R}^d$ and $\emptyset$ are elements of $\mathcal{P}(\mathbb{R}^d; \mathbb{R}^d_+)$. Moreover, $R$ defined at (3.1) can attain these values even if the underlying scalar risk measure $\rho$ maps to $\mathbb{R}$ only, e.g., when $\Lambda$ is bounded. While the case $R(Y) = \emptyset$ corresponds to the case that a scalar risk measure of a financial position is $+\infty$, meaning that the system $Y + k$ is deemed risky no matter how much capital is injected, the case $R(Y) = \mathbb{R}^d$ corresponds to $-\infty$ in the scalar case. The latter situation of “cash cows” with the possibility to withdraw any finite amount of money without rendering the position risky is usually deemed unrealistic and is excluded. Therefore, we shall usually only discuss the former case, but remark that a treatment of the latter were also possible for most results.

The cash-invariance and monotonicity carry over to $R$ ($R^{ins}$) in that for all $Y, Z \in \mathcal{Y}^d$ with $Y \geq Z$ $\mathbb{P}$-a.s. componentwise and all $k \in \mathbb{R}^d$, $R(Y) \supseteq R(Z)$, and $R(Y + k) = R(Y) - k$.

To shorten the notation, we also introduce further subclasses of $\mathcal{P}(\mathbb{R}^d; \mathbb{R}^d_+)$ where $\mathbb{B}(\mathbb{R}^d)$ denotes the Borel-$\sigma$-algebra.

**Definition 3.1.**
(i) The class of Borel-measurable upper subsets of $\mathbb{R}^d$ is denoted with $\hat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}^d_+) := (\mathcal{P}(\mathbb{R}^d; \mathbb{R}^d_+) \cap \mathbb{B}(\mathbb{R}^d)) \setminus \{\mathbb{R}^d\}$.

(ii) The class of closed upper subsets of $\mathbb{R}^d$ is denoted with $\mathcal{F}(\mathbb{R}^d; \mathbb{R}^d_+)$. Note that $\mathcal{F}(\mathbb{R}^d; \mathbb{R}^d_+) \subset \hat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}^d_+)$.

We shall regularly make use of the following assumptions.

**Assumption (1).** For all $Y \in \mathcal{Y}^d$, $R(Y) \in \hat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}^d_+)$. 

**Assumption (2).** For all $Y \in \mathcal{Y}^d$, $R(Y) \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}^d_+)$ and the set $\{k \in \mathbb{R}^d \mid \rho(\Lambda(Y + k)) = 0\}$ corresponds to the topological boundary $\partial R(Y)$ of $R(Y)$. 

21
If \( \Lambda : \mathbb{R}^d \to \mathbb{R} \) is continuous and \( \rho \) satisfies the Fatou property (which means it is lower-semicontinuous), the values of \( R \) are closed. Note that the law-invariance of \( \rho \) implies the Fatou property (Jouini, Schachermayer, & Touzi, 2006). If moreover the function \( \Lambda : \mathbb{R}^d \to \mathbb{R} \) is strictly increasing, the second part of Assumption (2) is satisfied as well. On the other hand, if \( \rho \) is convex (i.e., the corresponding acceptance set is convex) and \( \Lambda \) is concave, then \( R(Y) \) is convex (Feinstein et al., 2017).

Similarly to \( R \), we introduce the law-invariant map

\[
R_0 : \mathcal{Y}^d \to 2^\mathbb{R}^d, \quad Y \mapsto R_0(Y) = \{ k \in \mathbb{R}^d \mid \rho(\Lambda(Y + k)) = 0 \},
\]

(3.3)

Note that for \( R(Y) \neq \emptyset \), \( R_0(Y) \) is non-empty if the first part of Assumption (2) is satisfied. Since \( \Lambda \) is increasing and \( \rho \) is cash-invariant, one then obtains the relation

\[
R(Y) = R_0(Y) + \mathbb{R}^d_+.
\]

That means that the values of \( R_0 \) determine \( R \) completely. Moreover, if \( \Lambda \) is strictly increasing, then \( R_0(Y) \) can be characterised as the topological boundary of \( R(Y) \) which has the interpretation that \( R_0(Y) \) contains the efficient capital allocations that make \( Y \) acceptable under \( R \). That means for such situations, \( R \) and \( R_0 \) are connected via a one-to-one relation. Again, this means that exhaustive elicitability results for \( R \) (Theorem 4.9) carry over to \( R_0 \) for such situations, invoking the revelation principle.

Finally, we introduce an important scalarization of the systemic risk measure \( R \), called efficient cash-invariant allocation rule (EAR), as introduced in Feinstein et al. (2017). Under certain circumstances, an EAR can also be considered as a selection of \( R \), or alternatively, of \( R_0 \). Roughly speaking, for \( Y \in \mathcal{Y}^d \), the value of \( EAR(Y) \) gives the capital allocation(s) with minimal weighted costs of an allocation in \( R(Y) \). For simplicity, we shall confine attention to \( EARs \) with a fixed price or weight vector \( w \in \mathbb{R}^d_{++} := \{ x \in \mathbb{R}^d \mid x_1, \ldots, x_d > 0 \} \). To settle some notation, we introduce the following definition.

**Definition 3.2 (EAR).** An efficient cash-invariant allocation rule for a fixed price vector \( w \in \mathbb{R}^d_{++} \) is given by

\[
EAR_w(Y) = \arg \min_{k \in R(Y)} w^\top k.
\]

(3.4)

Note that \( EAR_w(Y) \) is well defined and non-empty for \( w \in \mathbb{R}^d_{++} \) if \( R(Y) \) is closed and there is a supporting hyperplane for \( R(Y) \) that is orthogonal to \( w \). \( EAR_w(Y) \) is then necessarily the intersection of \( R_0(Y) \) and this hyperplane. If \( R(Y) \) is not closed, the minimum in (3.4) might not be attained, resulting in \( EAR_w(Y) = \emptyset \). If, on the other hand, there is no supporting hyperplane for \( R(Y) \) orthogonal to \( w \), \( w^\top k \) for \( k \in R(Y) \) is unbounded from below and we have again \( EAR_w(Y) = \emptyset \).

As discussed in Feinstein et al. (2017), \( EAR_w(Y) \) is actually not necessarily a singleton. More precisely, for closed \( R(Y) \) it fails to be a singleton if and only if \( \partial R(Y) \) contains a line segment that is orthogonal to the price vector.

Since the scalar risk measure \( \rho \) is assumed to be law-invariant, also the derived quantities \( R, R^{ins}, R_0 \) and \( EAR_w \) are law-invariant. Therefore, we shall frequently abuse notation and write \( R(F_Y) := R(Y) \) for \( Y \in \mathcal{Y}^d \) with distribution \( F_Y \in \mathcal{M}^d \), with analogous conventions for the other law-invariant maps.
4. Main results

We present some of the main results of the paper in this section where we gather identifiability results in Subsection 4.1 and elicitability results are presented in Subsection 4.2. Theorem 4.1 establishes the selective identifiability of $R_0$. Remarkably, the fact that this selective identification function can usually be chosen to be oriented in an appropriate sense allows for an integral construction of strictly consistent exhaustive scoring functions for $R$ (Theorem 4.9). Moreover, the selective identifiability of $R_0$ also implies that efficient cash-invariant allocation rules (EARs) are selectively identifiable (Proposition 4.6).

Notably, the main assumption behind Theorem 4.1 and the subsequent results relying on this identifiability is the identifiability of the underlying scalar risk measure $\rho$ in (3.1). A fortiori, it needs to admit an oriented identification function. Invoking Theorem 8 in Steinwart et al. (2014) this is equivalent to the elicitability of $\rho$ under mild regularity conditions. Propositions 4.8 and A.4 establish that under certain assumptions on the aggregation function $\Lambda$ also the converse holds. That is, the elicitability of $R$ implies the elicitability of $\rho$.

4.1. Identifiability results

**Theorem 4.1.** Let $\rho: \mathcal{M} \rightarrow \mathbb{R}$ be identifiable. Then the following assertions hold for $R_0: \mathcal{M}^d \rightarrow 2^{\mathbb{R}^d}$ defined at (3.3).

(i) $R_0$ is selectively identifiable. If $V_\rho: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is a strict $\mathcal{M}$-identification function for $\rho$, then

$$V_{R_0}: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}, \quad (k, y) \mapsto V_{R_0}(k, y) = V_\rho(0, \Lambda(y + k))$$

is a strict selective $\mathcal{M}^d$-identification function for $R_0$.

(ii) If $V_\rho: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is an oriented strict $\mathcal{M}$-identification function for $\rho$, then $V_{R_0}$ defined at (4.1) is oriented for $R_0$ in the sense that for all $F \in \mathcal{M}^d$ it holds that

$$\bar{V}_{R_0}(k, F) \begin{cases} < 0, & \text{if } k \notin R(F) \\ = 0, & \text{if } k \in R_0(F) \\ > 0, & \text{if } k \in R(F) \setminus R_0(F). \end{cases}$$

**Remark 4.2.** The orientation of $V_{R_0}$ can be considered as the multivariate counterpart of the orientation of $V_\rho$ with respect to the componentwise order on $\mathbb{R}^d$. Indeed, in both cases, a negative expected identification function corresponds to the case of predicting a capital requirement too small to make the system $Y \in \mathcal{Y}^d$ acceptable with respect to $R$ or the single firm $X \in \mathcal{Y}$ acceptable with respect to $\rho$.

Note that if $V_{R_0}: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a strict selective $\mathcal{M}^d$-identification function for $R_0$ which is oriented in the sense of (4.2) and which is such that the expected identification function $\bar{V}_{R_0}(\cdot, F)$ is continuous for any $F \in \mathcal{M}^d$, then the values of $R$ are closed sets.
Remark 4.3. Equation (4.1) states explicitly how to construct a strict selective \( \mathcal{M}^d \)-identification function \( V_{R_0} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \) for \( R_0 \), given a certain strict \( \mathcal{M} \)-identification function \( V_\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \) for \( \rho \). So \( V_{R_0} \) definitely depends on the choice of \( V_\rho \). Fissler (2017, Proposition 3.2.1) states that under some richness assumptions on the class \( \mathcal{M} \), any other strict identification function \( \tilde{V}_\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \) for \( \rho \) is of the form

\[
\tilde{V}_\rho(x, z) = g(x)V_\rho(x, z),
\]

where \( g : \mathbb{R} \to \mathbb{R} \) is a non-vanishing function. Moreover, if \( V_\rho \) is oriented, then \( \tilde{V}_\rho \) is oriented if and only if the function \( g \) in (4.3) is strictly positive; see also Steinwart et al. (2014, Theorem 8). Consequently, starting with such an identification function \( V_\rho \), the resulting (oriented) strict selective \( \mathcal{M}^d \)-identification function \( \tilde{V}_{R_0} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \) takes the form

\[
\tilde{V}_{R_0}(k, y) = \tilde{V}_\rho(0, \Lambda(y + k)) = g(0)V_\rho(0, \Lambda(y + k)).
\]

Hence, the only difference is that one ends up with a scaled version of \( V_{R_0} \) where the scaling factor \( g(0) \) is positive if both \( V_{R_0} \) and \( \tilde{V}_{R_0} \) are oriented.\(^9\)

In a similar spirit as Remark 4.3, one might also wonder whether the (oriented) strict selective identification functions constructed in Theorem 4.1 are the only (oriented) strict selective identification functions for \( R_0 \). This is definitely not the case since due to the linearity of the expectation, any function \( V'_{R_0} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \) with

\[
V'_{R_0}(k, y) = h(k)V_{R_0}(k, y) = h(k)V_\rho(0, \Lambda(y + k)),
\]

where \( h : \mathbb{R}^d \to \mathbb{R} \) is non-vanishing, is again a strict selective \( \mathcal{M}^d \)-identification function for \( R_0 \). Moreover, if \( V_{R_0} \) is oriented, then \( V'_{R_0} \) defined at (4.5) is oriented if and only if \( h > 0 \). In particular, the constant \( g(0) \) appearing in (4.4) can be incorporated into the function \( h \) such that we see that it does not matter which (oriented) strict exhaustive identification function \( \tilde{V}_\rho \) we choose to end up with the form at (4.5). The following theorem establishes that basically all selective \( \mathcal{M}^d \)-identification functions for \( R_0 \) are of the form at (4.5).

Proposition 4.4. Let \( A \subseteq \mathbb{R}^d \) and let \( V_{R_0}, V'_{R_0} : A \times \mathbb{R}^d \to \mathbb{R} \) be strict selective \( \mathcal{M}^d \)-identification functions for \( R_0 : \mathcal{M}^d \to 2^{\mathbb{R}^d} \). If for every \( x \in A \) there are \( F_1, F_2 \in \mathcal{M}^d \) such that \( \tilde{V}_{R_0}(x, F_1) > 0 \) and \( \tilde{V}_{R_0}(x, F_2) < 0 \) and \( \mathcal{M}^d \) is convex, then there is a non-vanishing function \( h : A \to \mathbb{R} \) such that

\[
\tilde{V}'_{R_0}(x, F) = h(x)\tilde{V}_{R_0}(x, F)
\]

for all \( x \in A \) and all \( F \in \mathcal{M}^d \).

Remark 4.5. (i) It is worth mentioning that the assumptions of Proposition 4.4 imply that \( R_0 \) is surjective on \( A \subseteq \mathbb{R}^d \). That is why we formulated the proposition in terms of a general action domain \( A \subseteq \mathbb{R}^d \) rather than \( \mathbb{R}^d \).

\(^9\)In the light of Lemma 5.3 (see also Remark 5.4), this provides also an argument that there is—up to scaling—only one translation invariant identification function for any translation equivariant functional; compare also to the discussion in Fissler and Ziegel (2019c).
(ii) If $\mathcal{M}^d$ is rich enough, and under additional regularity conditions on $V_{R_0}$, one can also establish a pointwise version of (4.6); see Fissler and Ziegel (2016, 2019b) for details.

For any vector $w \in \mathbb{R}^d$, we use the notation $w^\perp := \{x \in \mathbb{R}^d | w^\top x = 0\}$ for the orthogonal complement of the subspace spanned by $w$. With $\mathbb{R}^{w^\perp}$ we denote the space of all functions mapping from $w^\perp$ to $\mathbb{R}$.

**Proposition 4.6.** Let $\rho$ be a scalar risk measure, $R$ and $R_0$ as defined in (3.1) and (3.3), and suppose that Assumption (2) holds. Assume that $R_0$ is selectively identifiable with an oriented strict selective $\mathcal{M}^d$-identification function $V_{R_0}$. Let $w \in \mathbb{R}^{d^+}$ and define the map $V_{EAR_w}: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{w^\perp}$ via

$$V_{EAR_w}(k, y): w^\perp \rightarrow \mathbb{R}, \quad w^\perp \ni x \mapsto V_{EAR_w}(k, y)(x) = V_{R_0}(k + x, y).$$

for $(k, y) \in \mathbb{R}^d \times \mathbb{R}^d$. Then $V_{EAR_w}$ is a strict selective $\mathcal{M}^d$-identification function for $EAR_w$ in the sense that for any $F \in \mathcal{M}^d$ and any $k \in \mathbb{R}^d$

$$k \in EAR_w(F) \iff \bar{V}_{EAR_w}(k, F) \leq 0 \land \bar{V}_{EAR_w}(k, F)(0) = 0. \quad (4.7)$$

If the underlying risk measure $R$ is known to assume convex sets only (e.g. if $\rho$ is convex and $\Lambda$ concave), it is even sufficient to evaluate $\bar{V}_{EAR_w}(k, F)(x)$, or its empirical counterpart, for $x \in \mathbb{R}^d$ in a neighbourhood of 0, which can also nicely be seen in Figure 3 in Appendix B.2.

Since the selective identifiability in Proposition 4.6 deviates from the usual definition, it is worth investigating its implication in terms of level sets.

**Lemma 4.7.** Let $\mathcal{M}^d$ be convex. Under the conditions of Proposition 4.6, $EAR_w$ satisfies the selective CxLS property.

The proof of Lemma 4.7 is standard and therefore omitted. The selective CxLS property of $EAR_w$ suggests the open question as to whether $EAR_w$ also satisfies the selective CxLS* property or, alternatively, the exhaustive CxLS property. We conjecture that it satisfies the selective CxLS* property which would imply the possibility that $EAR_w$ might also be selectively elicitable. Note that for the situation when $EAR_w$ is a singleton and therefore can be considered as a selection of $R$, Lemma 4.7 and also this conjecture are not ruled out by the result of Proposition 2.16. Indeed, Theorem 4.9 below establishes that $R$ satisfies the exhaustive CxLS property, and the cash-invariance induces the proper-subset property which in turn implies that $R$ does not satisfy the selective CxLS* property.

We would like to compare the concept of identifiability introduced in Proposition 4.6 to the discussion about the backtestability of loss value at risk in Section 5 of Bignozzi, Burzoni, and Munari (2018). One can interpret their proposal as using a function-valued identification function, too. Then, their analogue of (4.7) is that the infimum of the function-valued identification function be 0 if using the correctly specified forecast.
Interestingly, this version of identifiability does not imply that the functional under consideration has convex level sets.

We end this section by noting that the identifiability of $\rho$ and the selective identifiability of $R_0$ are even equivalent if $\Lambda : \mathbb{R}^d \to \mathbb{R}$ possesses a measurable right inverse.

**Proposition 4.8.** Let $\rho : \mathcal{M} \to \mathbb{R}$ be a risk measure, $\Lambda : \mathbb{R}^d \to \mathbb{R}$ a surjective aggregation function, and $R_0 : \mathcal{M}^d \to 2^{\mathbb{R}^d}$ as defined in (3.3). Assume that there exists a measurable right inverse $\eta : \mathbb{R} \to \mathbb{R}^d$ such that $\Lambda \circ \eta = \text{id}_{\mathbb{R}}$, $Y$ is closed under translations, and that for any $X \in Y$, $\eta(X)$ belongs to $Y^d$. Then it holds that $\rho$ is (selectively) identifiable if and only if $R_0$ is selectively identifiable.

### 4.2. Elicitability results and mixture representation

In the seminal paper Ehm et al. (2016) it is shown that, subject to regularity conditions, any non-negative scoring function $S : \mathbb{R} \times \mathbb{R} \to [0, \infty]$ which is consistent for the $\alpha$-quantile (the $\tau$-expectile) can be written as a mixture or Choquet representation

$$S(x, y) = \int_{\mathbb{R}} S_{\theta}(x, y) \, dH(\theta), \quad x, y \in \mathbb{R},$$

where $H$ is a non-negative measure on $B(\mathbb{R})$ and $S_{\theta}, \theta \in \mathbb{R}$, are non-negative elementary scoring functions for the $\alpha$-quantile (the $\tau$-expectile). In particular, $S_{\theta}$ take the form

$$S_{\theta}(x, y) = (\mathbb{1}\{\theta < x\} - \mathbb{1}\{\theta < y\}) V(\theta, y)$$

where $V$ is an oriented identification function for the $\alpha$-quantile (the $\tau$-expectile). The score at (4.8) is strictly consistent if and only if the measure $H$ is strictly positive, that is, it puts positive mass on any open non-empty set. Ziegel (2016) and Dawid (2016) argued that this construction also works for more general one-dimensional functionals besides expectiles and quantiles which admit an oriented identification function; cf. Jordan et al. (2019). Steinwart et al. (2014) showed that, for one-dimensional functionals satisfying certain regularity conditions, the existence of such an oriented identification function is equivalent to the elicitability of the functional. While the orientation of the identification function immediately gives rise to the consistency of the elementary scores, and thus, of the mixtures at (4.8), an answer to the question as to whether all scoring functions for a certain functional are necessarily of the form at (4.8) can typically only be answered invoking Osband’s Principle (Fissler & Ziegel, 2016; Osband, 1985) hence assuming smoothness and regularity conditions.

Our construction of strictly consistent exhaustive scoring functions for the systemic risk measures $R$ also exploits the key result about the existence of oriented strict selective identification functions for $R_0$ and is similar in nature to the approach described above. For any $y \in \mathbb{R}^d$, we shall use the notation $R(y) := R(\delta_y)$.

**Theorem 4.9.** Let $V_{R_0} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be such that for all $F \in \mathcal{M}^d \cup \{\delta_y \mid y \in \mathbb{R}^d\}$

$$V_{R_0}(k, F) \in \begin{cases} (\infty, 0], & \text{if } k \notin R(F) \\ [0, \infty), & \text{if } k \in R(F). \end{cases}$$

26
(i) Under Assumption (1), for each \( k \in \mathbb{R}^d \), the map \( S_{R,k} : \tilde{\mathcal{P}}(\mathbb{R}^d;\mathbb{R}_+^d) \times \mathbb{R}^d \to [0,\infty) \),
\[
S_{R,k}(A,y) = (1_{R(y) \cap A}(k) - 1_{A \setminus R(y)}(k)) V_{R_0}(k,y) \tag{4.11}
\]
is a non-negative \( \mathcal{M}^d \)-consistent exhaustive scoring function for \( R : \mathcal{M}^d \to \tilde{\mathcal{P}}(\mathbb{R}^d;\mathbb{R}_+^d) \).

(ii) Under Assumption (1) and if \( \pi \) is a \( \sigma \)-finite non-negative measure on \( \mathcal{B}(\mathbb{R}^d) \), the map \( S_{R,\pi} : \tilde{\mathcal{P}}(\mathbb{R}^d;\mathbb{R}_+^d) \times \mathbb{R}^d \to [0,\infty] \),
\[
S_{R,\pi}(A,y) = \int_{\mathbb{R}^d} S_{R,k}(A,y) \pi(\mathrm{d}k) \tag{4.12}
\]
is a non-negative \( \mathcal{M}^d \)-consistent exhaustive scoring function for \( R : \mathcal{M}^d \to \tilde{\mathcal{P}}(\mathbb{R}^d;\mathbb{R}_+^d) \).

(iii) If Assumption (2) holds, if \( V_{R_0} \) is a strict selective \( \mathcal{M}^d \)-identification function for \( R_0 \) and if \( \pi \) is a \( \sigma \)-finite strictly positive measure on \( \mathcal{B}(\mathbb{R}^d) \), then the restriction of \( S_{R,\pi} \) defined at (4.12) to \( \mathcal{F}(\mathbb{R}^d;\mathbb{R}_+^d) \times \mathbb{R}^d \) is strictly \( \mathcal{M}^d_0 \)-consistent for \( R : \mathcal{M}^d_0 \to \mathcal{F}(\mathbb{R}^d;\mathbb{R}_+^d) \), where \( \mathcal{M}^d_0 \subseteq \mathcal{M}^d \) such that \( \hat{S}_{R,\pi}(R(F),F) < \infty \) for all \( F \in \mathcal{M}^d_0 \).

Note that the condition at (4.10) is some weak form of orientation. However, it does not imply that \( V_{R_0} \) is an identification function for \( R_0 \). In the one-dimensional setting, such a situation can occur in practice if the underlying risk measure is Value at Risk and the distributions are not continuous, implying that the corresponding quantile identification function will nowhere attain 0 in expectation.

Even though we defer the formal proof of Theorem 4.9 to Appendix B.2, we would still like to sketch and illustrate the idea, taking into account that Theorem 4.9 constitutes one of the main results of the paper. The key observation is the identity
\[
\hat{S}_{R,\pi}(A,F) - \hat{S}_{R,\pi}(R(F),F) = \int_{R(F) \setminus A} \hat{V}_{R_0}(k,F) \pi(\mathrm{d}k) - \int_{A \setminus R(F)} \hat{V}_{R_0}(k,F) \pi(\mathrm{d}k). \tag{4.13}
\]

Then, one uses the weak orientation of \( V_{R_0} \) given at (4.10) to conclude that the first integral on the right hand side of (4.13) is \( \geq 0 \) while the second integral is \( \leq 0 \). A graphic illustration of the situation is provided in Figure 1.

It is in order to make some comments about the scoring functions constructed in Theorem 4.9.

### 4.2.1. Comparison with one-dimensional case

The similarity of the mixture representation at (4.12) and (4.8) is obvious. With a closer look, one can also see the similarities on the level of the elementary scores given at (4.11) and (4.9). Indeed, (4.9) can be re-written as
\[
S_0(x,y) = (1_{[y,\infty]\setminus[x,\infty]}(\theta) - 1_{[x,\infty]\setminus[y,\infty]}(\theta)) V(\theta,y).
\]
The form of \( R(y) \) can be described explicitly in the following lemma, where we use the fact that \( \rho(0(0)) = \rho(0) - \rho(0) = 0 \).
Figure 1: A graphical illustration of Equation (4.13) for dimension $d = 2$. Suppose the red region corresponds to the correctly specified risk measure $R(F)$ and the blue region corresponds to some misspecified forecast $A$. The score difference $\tilde{S}_{R,\pi}(A,F) - \tilde{S}_{R,\pi}(R(F),F)$ is an integral of $\tilde{V}_{R_0}(\cdot,F)$ over $R(F) \setminus A$ (the red only region), plus an integral of $-\tilde{V}_{R_0}(\cdot,F)$ over $A \setminus R(F)$ (the blue only region).

Lemma 4.10. Let $R(Y) = \{k \in \mathbb{R}^d \mid \rho(\Lambda(Y + k)) \leq 0\}$, $Y \in \mathcal{Y}^d$, where the scalar risk measure $\rho$ is decreasing and cash-invariant, and the aggregation function $\Lambda: \mathbb{R}^d \to \mathbb{R}$ is increasing. Then, for each $y \in \mathbb{R}^d$ it holds that

$$R(y) = \Lambda^{-1}([\rho(0), \infty)) - y = \Lambda^{-1}(\{\rho(0)\}) + \mathbb{R}^d_+ - y.$$  

Accounting for the sign convention that the negative of a quantile or expectile are a scalar risk measure, one can see that the elementary scores at (4.11) essentially boil down to the ones at (4.9) for dimension $d = 1$.

4.2.2. Integrability

The non-negativity of the elementary scores at (4.11) guarantees that the integral at (4.12) always exists. However, as stated in part (iii) of Theorem 4.9, these scores are only strictly consistent if $\tilde{S}_{R,\pi}(R(F),F) < \infty$, which suggests the question as to when the integral at (4.12) is finite. A sufficient condition for the latter is that

$$\int_{\mathbb{R}^d} |V_{R_0}(k,y)| \pi(dk) < \infty.$$  

Therefore, a sufficient condition for the finiteness of $\tilde{S}_{R,\pi}(A,F)$ is that $V_{R_0}$ is $\pi \otimes F$-integrable.
4.2.3. Normalisation

By construction, the elementary scores at (4.11), and therefore the scores at (4.12), are non-negative. It is well known that if a scoring function \( S(x, y) \) is (strictly) \( \mathcal{M} \)-consistent for some functional \( T \) then for \( \lambda > 0 \) and some \( \mathcal{M} \)-integrable function \( a: \mathcal{O} \to \mathbb{R} \), the score \( S'(x, y) = \lambda S(x, y) + a(y) \), \( \lambda > 0 \) is also (strictly) \( \mathcal{M} \)-consistent for \( T \). Following Gneiting and Raftery (2007) we say that \( S \) and \( S' \) are equivalent. Therefore, if \( \mathcal{M} \) contains all point measures, \( S' \) is strictly \( \mathcal{M} \)-consistent for \( T \) and \( y \mapsto S'(T(\delta_y), y) \) is \( \mathcal{M} \)-integrable, the score \( S(x, y) = S'(x, y) - S'(T(\delta_y), y) \) is non-negative by construction. However, sometimes relaxing the normalisation condition that a score be non-negative can also help to relax integrability conditions on the scoring function. A standard example is the squared loss \( S(x, y) = (x - y)^2 \) which is non-negative, consistent for the mean relative to any class of distributions with a finite first moment, and strictly consistent for the mean relative to any class of distributions with a finite second moment. On the other hand, the equivalent score \( S'(x, y) = x^2 - 2xy \) maps to \( \mathbb{R} \), but is strictly consistent for the mean relative to any class of distributions with a finite first moment.

In that light, it might be interesting to consider scores \( S'_{R,k} \) which are equivalent to the elementary scores at (4.11). A natural choice might be \( S'_{R,k}(A, y) = -\mathbf{1}_A(k)V_{R_0}(k, y) \); cf. Dawid (2016). This leads to an alternative mixture representation akin to (4.12) of the form

\[
S'_{R,\pi}(A, y) = \int_{\mathbb{R}^d} S'_{R,k}(A, y) \pi(dk) = -\int_A V_{R_0}(k, y) \pi(dk).
\]

However, since the integrand \( S'_{R,k}(A, y) \) may attain both positive and negative values, one needs to impose that its negative part is \( \pi \)-integrable in order to guarantee the existence of the integral.

4.2.4. Characterisation of all consistent scoring functions

There is evidence that—under appropriate regularity conditions—all consistent scoring functions for the risk measure \( R \) are equivalent to a score of the form given at (4.12). That means, modulo equivalence, the choice of the consistent scoring function boils down to the choice of the measure \( \pi \).

Firstly, note that Proposition 4.4 implies that it does not matter what oriented strict \( \mathcal{M}^d \)-identification \( V_{R_0} \) we actually start with. Indeed, if \( V'_{R_0} \) were another such identification function, then \( V'_{R_0}(k, y) = h(k)V_{R_0}(k, y) \) for some positive function \( h \). But this solely amounts to a change of measure, since \( V_{R_0}(k, y)\pi(dk) = V'_{R_0}(k, y)\pi'(dk) \), where \( \pi' \) has the density \( 1/h \) with respect to \( \pi \). Secondly, the class of scoring functions of the form (4.12) is convex, which is a necessary condition (Gneiting, 2011a). Thirdly, as observed above, the mixture representation at (4.12) is the natural extension to the one-dimensional case. As remarked, for the one-dimensional case, one can typically establish this sort of necessary conditions only invoking Osband’s principle. Since Osband’s principle relies on a first-order-condition argument, it has only been established under smoothness conditions and for the finite dimensional case. We suspect that it is possible to generalise it to the infinite dimensional setting of predicting upper sets in
Possible approaches might work by borrowing ideas from the calculus of variations or by considering increments of scores rather than derivatives. However, the technical treatment of these approaches would be beyond the scope of the paper at hand such that we defer it to future research.

4.2.5. Order-sensitivity

It is known that—under weak assumptions on a scalar functional $T$—all strictly consistent scoring functions $S$ for $T$ are order-sensitive or accuracy-rewarding; see Nau (1985, Proposition 3), Lambert (2013, Proposition 2), Bellini and Bignozzi (2015, Proposition 3.4). In the scalar setting, this property means that $x_1 \leq x_2 \leq T(F)$ or $T(F) \leq x_2 \leq x_1$ implies that $\bar{S}(x_1, F) \geq \bar{S}(x_2, F)$. While one gets this useful property essentially ‘for free’ in the scalar case, asking for order-sensitivity in a multivariate setting is a lot more involved; see Fissler and Ziegel (2019c). One of the main questions in the multivariate setting is which order relation to use. In the present situation where our exhaustive action domain consists of closed upper subsets of $\mathbb{R}^d$, the canonical (partial) order relation is the subset relation. That means the canonical analogue of order-sensitivity in our setting is that for any distribution $F \in \mathcal{M}_d$ it holds that $A \subseteq B \subseteq R(F)$ or $A \supseteq B \supseteq R(F)$ implies that $\bar{S}_{R,\pi}(A, F) \geq \bar{S}_{R,\pi}(B, F)$. The following proposition establishes that this notion of order-sensitivity is fulfilled by all scoring functions introduced in Theorem 4.9(ii). The proof basically exploits the orientation of the underlying identification function $V_{R_0}$, which is a similar argument to the one given in Steinwart et al. (2014).

**Proposition 4.11.** Let the assumptions of Theorem 4.9(ii) prevail. Then, the scoring function $\bar{S}_{R,\pi}$ defined at (4.12) is $\mathcal{M}^d$-order-sensitive for $R$ in the sense that for all $A, B \in \hat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}^d_+)$ and for all $F \in \mathcal{M}_d$

$$\left( A \subseteq B \subseteq R(F) \text{ or } A \supseteq B \supseteq R(F) \right) \implies \bar{S}_{R,\pi}(A, F) \geq \bar{S}_{R,\pi}(B, F). \quad (4.14)$$

Under the assumptions of Theorem 4.9(iii), if $\bar{S}_{R,\pi}(B, F) < \infty$ and the inclusions $A \subseteq B$ or $A \supseteq B$ on the left hand side of (4.14) is strict, then the inequality on the right hand side is also strict.

4.2.6. Forecast dominance and Murphy diagrams

The notion of (strict) consistency implies that—in expectation—a correctly specified forecast will score at least as high as (strictly less than) any misspecified score. On the level of the prediction space setting (Gneiting & Ranjan, 2013; Strähl & Ziegel, 2017), Holzmann and Eulert (2014) showed that for two ideal forecasts, the one measurable with respect to a strictly larger information set is preferred under any strictly consistent scoring function; cf. Tsyplyakov (2014). Patton (2019) demonstrated that, in general, two misspecified forecasts rank differently under different (consistent) scoring functions. Therefore, the choice of the scoring function used in practice matters and
secondary quality criteria besides consistency, such as translation invariance or homo-
genity, may guide the decision what scoring function to use; see Section 5. For the rare situation when one forecast scores better than another one uniformly over all consistent scoring functions, Ehm et al. (2016) coined the term forecast dominance. We give the corresponding definition here for the situation of exhaustive forecasts for systemic risk measures $R$.

**Definition 4.12 (Dominance).** Let $Y \in \mathcal{Y}^d$ and $A, B$ two (stochastic) forecasts for some systemic risk measure $R$ of the form at (3.1), taking values in $\hat{\mathcal{P}}(\mathbb{R}_d^d; \mathbb{R}_d^d)$. Assume that Assumption (1) holds. Then $A$ dominates $B$ if $E[S_{R,\pi}(A, Y)] \leq E[S_{R,\pi}(B, Y)]$ for all consistent scoring functions $S_{R,\pi}$ of the form at (4.12) where $\pi$ is a $\sigma$-finite non-negative measure on $B(\mathbb{R}_d^d)$.

Note that the expectations are taken over the joint distribution of the forecasts and the observation.

Since the scores $S_{R,\pi}$ at (4.12) are parametrised by the class of non-negative $\sigma$-additive measures on $B(\mathbb{R}_d^d)$, it is not very handy to check forecast dominance in practice using the definition. To this end, the following corollary is helpful. The proof is straightforward and therefore omitted.

**Corollary 4.13.** Let $Y \in \mathcal{Y}^d$ and $A, B$ two (stochastic) forecasts for some systemic risk measure $R$ of the form at (3.1), taking values in $\hat{\mathcal{P}}(\mathbb{R}_d^d; \mathbb{R}_d^d)$. Assume that Assumption (1) holds. Then $A$ dominates $B$ if and only if $E[S_{R,k}(A, Y)] \leq E[S_{R,k}(B, Y)]$ for all elementary scores $S_{R,k}$ given at (4.11), where $k \in \mathbb{R}_d^d$.

Corollary 4.13 opens the way to an immediate multivariate analogue of Murphy diagrams considered in Ehm et al. (2016). That is, if $A$ is a $\hat{\mathcal{P}}(\mathbb{R}_d^d; \mathbb{R}_d^d)$-valued forecast of a systemic risk measure $R$ and $Y$ is the corresponding $\mathbb{R}_d^d$-valued observation of a financial system, we can consider the map

$$\mathbb{R}_d^d \ni k \mapsto s_A(k) = E[S_{R,k}(A, Y)]$$

(4.15)

as a diagnostic tool. For an empirical setting with forecasts $A_1, \ldots, A_N \in \hat{\mathcal{P}}(\mathbb{R}_d^d; \mathbb{R}_d^d)$ and observations $Y_1, \ldots, Y_N \in \mathbb{R}_d^d$, (4.15) takes the form

$$\mathbb{R}_d^d \ni k \mapsto \hat{s}_{N,A}(k) = \frac{1}{N} \sum_{t=1}^N S_{R,k}(A_t, Y_t).$$

(4.16)

We illustrate the usage of Murphy diagrams in simulation study presented in Subsection 7.2.

5. Homogeneous and translation invariant scoring functions

Recall that the systemic risk measure $R$ is cash-invariant, or translation equivariant, meaning that $R(Y + k) = R(Y) - k$ for all $Y \in \mathcal{Y}^d$ and for all $k \in \mathbb{R}_d^d$. Moreover, if
\( \rho \) and \( \Lambda \) are positively homogeneous, so is \( R \) in the sense that \( R(cY) = cR(Y) \) for all \( Y \in \mathbb{Y}^d \) and \( c > 0 \); see Lemma 5.2. Patton (2011), Nolde and Ziegel (2017) and Fissler and Ziegel (2019c) have argued that it is a reasonable requirement for a scoring function that ordering different sequences of forecasts in terms of their realised scores be invariant under transformations to which the functional of interest is equivariant. Therefore, we discuss translation invariance and positive homogeneity of scores (or score differences) for \( R \). We start by gathering some elementary definitions.

**Definition 5.1** (Homogeneity, translation invariance). (i) A function \( f : \mathbb{R}^d \to \mathbb{R} \) is called positively homogeneous of degree \( b \in \mathbb{R} \) if \( f(cx) = c^b f(x) \) for all \( c > 0 \) and for all \( x \in \mathbb{R}^d \).

(ii) A scalar risk measure \( \rho : \mathcal{Y} \to \mathbb{R} \) is called positively homogeneous if \( \rho(cX) = c\rho(X) \) for all \( c > 0 \).

(iii) An exhaustive scoring function \( S : \mathcal{F}(\mathbb{R}^d; \mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R} \) is said to have positively homogeneous score differences of degree \( b \in \mathbb{R} \) if \( S(cA, cy) - S(cB, cy) = c^b(S(A, y) - S(B, y)) \) for all \( A, B \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}^d) \), \( y \in \mathbb{R}^d \) and \( c > 0 \).

(iv) A selective identification function \( V : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \) is called translation invariant if \( V(k + l, y - l) = V(k, y) \) for all \( k, l, y \in \mathbb{R}^d \).

(v) An exhaustive scoring function \( S : \mathcal{F}(\mathbb{R}^d, \mathbb{R}^d) \to \mathbb{R} \) is said to have translation invariant score differences, if \( S(A + l, y - l) - S(B + l, y - l) = S(A, y) - S(B, y) \) for all \( A, B \in \mathcal{F}(\mathbb{R}^d, \mathbb{R}^d) \) and \( y \in \mathbb{R}^d \).

(vi) A measure \( \pi \) on \( \mathcal{B}(\mathbb{R}^d) \) is translation invariant if \( \pi(A) = \pi(A + l) \) for all \( A \in \mathcal{B}(\mathbb{R}^d) \) and for all \( l \in \mathbb{R}^d \).

(vii) A measure \( \pi \) on \( \mathcal{B}(\mathbb{R}^d) \) is positively homogeneous of degree \( b \in \mathbb{R} \) if \( \pi(cA) = c^b\pi(A) \) for all \( A \in \mathcal{B}(\mathbb{R}^d) \) and for all \( c > 0 \).

With these definitions in mind, we can now state the following results.

**Lemma 5.2.** If \( \rho \) is a positively homogeneous scalar risk measure and \( \Lambda \) is positively homogeneous of any degree \( b \in \mathbb{R} \), \( R \) and \( R^{ins} \) as defined in (3.1) and (3.2) are positively homogeneous, i.e. for all \( c > 0 \) and \( Y \in \mathbb{Y}^d \), \( R(cY) = cR(Y) \) and \( R^{ins}(cY) = cR^{ins}(Y) \).

**Lemma 5.3.** Assume that \( \rho : \mathcal{M} \to \mathbb{R} \) has a strict \( \mathcal{M} \)-identification function \( V_\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \). Then the following holds for \( V_{R_0} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \) defined at (4.1):

(i) \( V_{R_0} \) is translation invariant.

(ii) Assume that \( \mathcal{M} \) is convex and that for any \( x \in \mathbb{R}^d \) there are \( F_1, F_2 \in \mathcal{M}^d \) such that \( \tilde{V}_{R_0}(x, F_1) > 0 \) and \( V_{R_0}(x, F_2) < 0 \). Then for any translation invariant strict \( \mathcal{M}^d \)-identification function \( V'_{R_0} \) for \( R_0 \) there is some \( \lambda \neq 0 \) such that

\[
\tilde{V}'_{R_0}(x, F) = \lambda \tilde{V}_{R_0}(x, F)
\]

for all \( x \in \mathbb{R}^d \) and for all \( F \in \mathcal{M}^d \).

(iii) If \( V_\rho(0, \cdot) : \mathbb{R} \to \mathbb{R} \) is positively homogeneous of degree \( a \in \mathbb{R} \) and \( \Lambda \) is positively homogeneous of degree \( b \in \mathbb{R} \), then \( V_{R_0} \) is positively homogeneous of degree \( ab \).
Remark 5.4. Interestingly, part (i) of Lemma 5.3 implies that if $\rho: \mathcal{M} \to \mathbb{R}$ is identifiable and cash-invariant with a strict $\mathcal{M}$-identification function $V^\text{inv}_\rho: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, then $V^\text{inv}_\rho(x, y) := V^\rho(0, x + y)$ is translation invariant. This result can also be generalised to translation equivariant functionals, hence establishing the converse of Fissler and Ziegel (2019c, Proposition 4.7(i)).

Proposition 5.5. Let Assumption (2) hold and assume that $\rho$ is identifiable with an oriented strict $\mathcal{M}$-identification function $V^\rho$.  

(i) Let $L^d$ be the $d$-dimensional Lebesgue measure and $S_{R,k}$ be an elementary score of the form at (4.11) with identification function $V^R_0(k, y) = V^\rho(0, \Lambda(y + k))$. Then the scoring function 

$$S_{R,L^d}(A,y) = \int_{\mathbb{R}^d} S_{R,k}(A, y) L^d(dk), \quad A \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}^d_+), \quad y \in \mathbb{R}^d$$

(5.1)

is translation invariant and $\mathcal{M}^d$-consistent for $R$.

(ii) Any finite $\mathcal{M}^d$-consistent scoring function $S$ for $R$ of the form at (4.12) is translation invariant only if $S(A,y) = \gamma S_{R,L^d}(A,y)$ at (5.1) for some $\gamma \geq 0$.

Note that for all examples of $\rho$ and $\Lambda$ we are aware of, it holds that for the score at (5.1) $S_{R,L^d}(A,y)$ is finite if the symmetric difference $A \triangle R(y)$ is bounded and only if it has a finite Lebesgue measure. Hence, Proposition 5.5(ii) implies that the only finite translation invariant consistent score is the 0-score.

Proposition 5.6. Let Assumption (2) hold and suppose that $V^R_0$ is an oriented strict selective $\mathcal{M}^d$-identification function for $R_0$ which is positively homogeneous of degree $a \in \mathbb{R}$.

(i) Let $\pi$ be a non-negative $\sigma$-finite positively homogeneous measure of degree $b \in \mathbb{R}$ and $S_{R,k}$ be an elementary score of the form at (4.11) with identification function $V^R_0$. Then the scoring function 

$$S_{R,\pi}(A,y) = \int_{\mathbb{R}^d} S_{R,k}(A, y) \pi(dk), \quad A \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}^d_+), \quad y \in \mathbb{R}^d$$

(5.2)

is positively homogeneous of degree $a + b$.

(ii) Any finite $\mathcal{M}^d$-consistent scoring function $S$ for $R$ of the form at (4.12) is positively homogeneous of degree $a + b$ only if $S(A,y) = \gamma S_{R,\pi}(A,y)$ at (5.2) for some $\gamma \geq 0$ and for some non-negative $\sigma$-finite positively homogeneous measure $\pi$ of degree $b \in \mathbb{R}$.

Remark 5.7. For many measures $\pi$ and sets $A$ the score $S_{R,\pi}(A,y)$ defined at (4.12) might not be finite which diminishes the practical statistical applicability in the context of forecast comparison. More to the point, score differences involving $S_{R,\pi}(A,y)$ will not be finite or might even not be defined at all. To overcome this issue we suggest to work
with the following convention of score differences

\[ S_{R,\pi}(A, y) - S_{R,\pi}(B, y) := \int_{\mathbb{R}^d} S_{R,k}(A, y) - S_{R,k}(B, y) \pi(\text{d}k) \]

\[ = \int_{B \setminus A} V_{R_0}(k, y) \pi(\text{d}k) - \int_{A \setminus B} V_{R_0}(k, y) \pi(\text{d}k), \]

where \( S_{R,k} \) are the elementary scores defined at (4.11), which assume finite values only. Indeed, the integral at (5.3) might exist and might even be finite, even if \( S_{R,\pi}(A, y) \) or \( S_{R,\pi}(B, y) \) are \( \infty \). This can be particularly helpful when working with translation invariant or positively homogeneous scores.

6. Elicitability of systemic risk measures based on Expected Shortfall

The two most common scalar risk measures in quantitative risk management are Value at Risk (\( \text{VaR}_\alpha \)) and Expected Shortfall (\( \text{ES}_\alpha \)) at some level \( \alpha \in (0, 1) \). Both are law-invariant scalar risk measures such that we can define them directly as functionals on appropriate classes of distributions. For a probability distribution function \( F \) and \( \alpha \in (0, 1) \) we define

\[ \text{VaR}_\alpha(F) = -\inf\{x \in \mathbb{R} | \alpha \leq F(x)\}, \]

\[ \text{ES}_\alpha(F) = \frac{1}{\alpha} \int_0^\alpha \text{VaR}_\beta(F) \text{d}\beta \]

\[ = -\frac{1}{\alpha} \int_{[-\infty, -\text{VaR}_\alpha(F)]} x \text{d}F(x) - \frac{1}{\alpha} \text{VaR}_\alpha(F)(F(-\text{VaR}_\alpha(F)) - \alpha). \]

The last decade has seen quite a lively debate about which scalar risk measure is best to use in practice where the debate has mainly focused on the dichotomy of \( \text{VaR}_\alpha \) and \( \text{ES}_\alpha \); see Embrechts et al. (2014) and Emmer et al. (2015) for a comprehensive academic discussion and Bank for International Settlements (2014) for a regulatory perspective. \( \text{VaR}_\alpha \) is robust in the sense of Hampel (1971), but ignores losses beyond the level \( \alpha \). Moreover, Cont et al. (2010) showed that robustness and coherence are mutually exclusive implying that \( \text{VaR}_\alpha \) fails to be coherent. On the other hand, \( \text{ES}_\alpha \) is a coherent—thus non-robust—risk measure. As a tail expectation, it takes into account the losses beyond the level \( \alpha \) by definition. Another layer of the joust between the two risk measures is their backtestability (Acerbi & Szekely, 2014, 2017). While the identifiability of a risk measure is important, but not necessary for traditional backtesting, comparative backtesting relies on the elicitability of the risk measure at hand; see Fissler et al. (2016) and Nolde and Ziegel (2017).

As the negative of a selection of the \( \alpha \)-quantile, \( \text{VaR}_\alpha \) is elicitable on any class of distributions with a unique \( \alpha \)-quantile. In stark contrast, Gneiting (2011a) demonstrated that \( \text{ES}_\alpha \) does generally not satisfy the CxLS property which rules out its elicitability; cf. Weber (2006).
Recall that Theorem 4.1 and Theorem 4.9 establish identifiability and elicitation results for systemic risk measures based on a scalar risk measure \( \rho \) which is identifiable, and therefore—under weak regularity assumption—elicitable; see Steinwart et al. (2014). Moreover, Proposition 4.8 establishes that, under weak regularity conditions, the identifiability / elicibility of \( \rho \) is also necessary for the identifiability and elicibility of the systemic risk measure based on \( \rho \). Therefore, for some aggregation function \( \Lambda: \mathbb{R}^d \to \mathbb{R} \), the systemic risk measure \( R^{ES_\alpha}(Y) = \{ k \in \mathbb{R}^d | ES_\alpha(\Lambda(Y + k)) \leq 0 \} \), \( Y \in \mathcal{Y}^d \), generally fails to be elicitable. On the other hand, for scalar risk measures, Fissler and Ziegel (2016) established that the pair \( (VaR_\alpha, ES_\alpha) \) is elicitable under weak regularity conditions; cf. Acerbi and Szekely (2014). This might trigger the suspicion that the pair \( (VaR_\alpha, R^{ES_\alpha}) \) mapping to the product space \( \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d) \times \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d) \) is exhaustively elicitable. We conjecture, however, that \( (VaR_\alpha, R^{ES_\alpha}) \), in general, fails to have the exhaustive CxLS property for \( d \geq 2 \), ruling out its exhaustive elicibility.

Therefore, we need slightly more information than \( R^{VaR_\alpha}(Y) = \{ k \in \mathbb{R}^d | VaR_\alpha(\Lambda(Y + k)) \leq 0 \} \) in the other component to render the pair involving \( R^{ES_\alpha} \) elicitable. Note that \( R^{VaR_\alpha}(Y) \) only encodes information about the sign of \( VaR_\alpha(\Lambda(Y + k)) \) for each \( k \in \mathbb{R}^d \). Apart from \( k \) in the boundary of \( R^{VaR_\alpha}(Y) \) we know nothing about the actual size of \( VaR_\alpha(\Lambda(Y + k)) \). However, the positive result about the elicibility of the pair \( (VaR_\alpha, ES_\alpha) \) actually exploits the fact that for the scoring function \( S_\alpha(x, y) = -(1\{y \leq -x\} - \alpha)x/\alpha - 1\{y \leq -x\}y/\alpha, x, y \in \mathbb{R}, VaR_\alpha(F) \) is the minimiser of the expected score while \( ES_\alpha(F) \) is its minimum; see Frongillo and Kash (2015a). Therefore, we shall consider the function-valued functional \( T^{VaR_\alpha}: \mathcal{Y}^d \to \mathbb{R}^d \) where for each \( Y \in \mathcal{Y}^d \)

\[
T^{VaR_\alpha}(Y): \mathbb{R}^d \to \mathbb{R}, \quad \mathbb{R}^d \ni k \mapsto T^{VaR_\alpha}(Y)(k) = VaR_\alpha(\Lambda(Y + k)). \tag{6.2}
\]

### 6.1. Identifiability results

To simplify the exposition of the results, we shall make the following assumption about the class \( \mathcal{M} \).

**Assumption (3).** All distribution functions in \( \mathcal{M} \) are continuous and strictly increasing.

Note that this assumption imposes also implicit restrictions on the class \( \mathcal{M}^d \) since we assume that for any \( Y \) with distribution in \( \mathcal{M}^d \), the random variable \( \Lambda(Y + k) \) has a distribution in \( \mathcal{M} \) for any \( k \in \mathbb{R}^d \).

A strict \( \mathcal{M} \)-identification function \( V: \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2 \) for the pair \( (VaR_\alpha, ES_\alpha): \mathcal{M} \to \mathbb{R}^2 \) is given in terms of

\[
V(v, e, y) = \left( \alpha - 1\{y + v \leq 0\} + 1\{y + v \leq 0\}y/\alpha + \alpha 1\{y + v \leq 0\} - \alpha)v/\alpha \right),
\]

\((v, e) \in \mathbb{R}^2, y \in \mathbb{R} \), which can be verified by a straightforward calculation. This induces a (non-strict) selective \( \mathcal{M}^d \)-identification function \( U: \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^2 \)
for \((T^{VaR_α}, R^{ES_α}_0)\): \(\mathcal{M}^d \to \mathbb{R}^d \times 2^{\mathbb{R}^d}\). For \(v: \mathbb{R}^d \to \mathbb{R}\), \(k \in \mathbb{R}^d\) and \(y \in \mathbb{R}^d\) it is defined by

\[
U(v, k, y) = \begin{cases} 
\alpha - 1\{\Lambda(y + k) + v(k) \leq 0\} \\
1\{\Lambda(y + k) + v(k) \leq 0\} \Lambda(y + k)/\alpha + (1 - \alpha) v(k)/\alpha 
\end{cases} \quad (6.3)
\]

**Proposition 6.1.** For any \(F \in \mathcal{M}^d\), the component \(U_2\) of \(U\) defined at (6.3) is oriented in the sense that for any \(k \in \mathbb{R}^d\)

\[
\overline{U}_2(T^{VaR_α}(F), k, F) = \begin{cases} 
< 0, & \text{if } k \notin R^{ES_α}(F) \\
= 0, & \text{if } k \in R^{ES_α}_0(F) \\
> 0, & \text{if } k \in R^{ES_α}_0(F) \setminus R^{ES_α}(F).
\end{cases} \quad (6.4)
\]

Under Assumption (3) the map \(U\) is a selective \(\mathcal{M}^d\)-identification function for the functional \((T^{VaR_α}, R^{ES_α}_0)\): \(\mathcal{M}^d \to \mathbb{R}^d \times 2^{\mathbb{R}^d}\).

### 6.2. Elicitability results

We introduce the following regularity assumption on \(T^{VaR_α}\) defined at (6.2).

**Assumption (4).** The functional \(T^{VaR_α}: \mathcal{M}^d \to \mathbb{R}^d\) takes only values in \(\mathcal{C}(\mathbb{R}^d; \mathbb{R})\), the space of continuous functions from \(\mathbb{R}^d\) to \(\mathbb{R}\).

With a standard argument one can verify that Assumption (3) together with the continuity of \(\Lambda\) imply Assumption (4).

In order to present the following theorem more compactly, let us introduce \(S_{α,g}(x,y) = (1\{y ≤ x\} - α)(g(x) - g(y))\) for any increasing function \(g: \mathbb{R} \to \mathbb{R}\). Recall that \(S_{α,g}\) is a non-negative consistent selective scoring function for the \(α\)-quantile. Moreover, if \(g\) is strictly increasing, \(S_{α,g}\) is a strictly consistent selective scoring function for the \(α\)-quantile relative to any class \(\mathcal{M}\) of distributions such that \(g\) is \(\mathcal{M}\)-integrable; see Gneiting (2011b).

**Theorem 6.2.** (i) Under Assumption (1), for every \(k \in \mathbb{R}^d\) the function \(S_k: \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d; \mathbb{R}_+^d) \times \mathbb{R}^d \to [0, \infty)\),

\[
S_k(v, A, y) = -\Lambda_A(k) U_2(v, k, y) - 1_{R^{ES_α}_0(y, k)} \Lambda(y + k) \quad (6.5)
\]

is a non-negative \(\mathcal{M}^d\)-consistent exhaustive scoring function for \((T^{VaR_α}, R^{ES_α}_0): \mathcal{M}^d \to \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d; \mathbb{R}_+^d)\).

(ii) Under Assumption (1) and if \(π_1, π_2\) are σ-finite non-negative measures on \(\mathcal{B}(\mathbb{R}^d)\), the map \(S_{π_1, π_2}: \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d; \mathbb{R}_+^d) \times \mathbb{R}^d \to [0, \infty)\),

\[
S_{π_1, π_2}(v, A, y) = \int_{\mathbb{R}^d} S_{α,g_κ}(-v(k), \Lambda(y + k)) π_1(\text{dk}) + \int_{\mathbb{R}^d} S_k(v, A, y) π_2(\text{dk}), \quad (6.6)
\]

where for each \(k \in \mathbb{R}^d\) the function \(g_k: \mathbb{R} \to \mathbb{R}\) is non-decreasing and \(S_k\) is given at (6.5), is a non-negative \(\mathcal{M}^d\)-consistent exhaustive scoring function for \((T^{VaR_α}, R^{ES_α}_0): \mathcal{M}^d \to \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d; \mathbb{R}_+^d)\).
(iii) If Assumptions (2), (3), and (4) hold, if $g_k$ is strictly increasing for all $k \in \mathbb{R}^d$ and if $\pi_1, \pi_2$ are strictly positive, then the restriction of $S_{\pi_1, \pi_2}$ defined at (6.6) to $\mathcal{C}(\mathbb{R}^d; \mathbb{R}) \times \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d) \times \mathbb{R}^d$ is a non-negative strictly $M_{d}^0$-consistent exhaustive scoring function for $(T^{VaR}_\alpha, R^{ES}_\alpha)$: $M_{d}^0 \rightarrow \mathcal{C}(\mathbb{R}^d; \mathbb{R}) \times \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$, where $M_{d}^0 \subseteq \mathcal{M}^d$ is such that $\bar{S}_{\pi_1, \pi_2}(T^{VaR}_\alpha(F), R^{ES}(F), F) < \infty$ for all $F \in M_{d}^0$.

Theorem 6.2(ii) suggests that there is again the possibility to consider Murphy diagrams to assess the quality of forecasts for $(T^{VaR}_\alpha, R^{ES}_\alpha)$ simultaneously over all scoring functions given at (6.6). However, a direct implementation would amount to defining them on the 2d-dimensional Euclidean space. If one further decomposes the functions $g_k$ in the spirit of Ehm et al. (2016), one would even end up with a map defined on $\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d$. However, arguing along the lines of Ziegel, Krüger, Jordan, and Fasciati (2019), the measure $\pi_1$ only accounts for forecast accuracy in the Value at Risk component. Therefore, if interest focuses on the Expected Shortfall component, it makes sense to set $\pi_1 = 0$ to facilitate the analysis. This implies that one can consider the Murphy diagram

$$
\mathbb{R}^d \ni k \mapsto \mathbf{E}[S_k(v, A, Y)]
$$

with the elementary scores $S_k$ given at (6.5). The empirical formulation in the spirit of (4.16) is straightforward.

7. Examples and simulations

7.1. Consistency of exhaustive scoring function for $R$

In this subsection, we shall demonstrate the discrimination ability of the consistent exhaustive scoring functions constructed in Theorem 4.9 via a simulation study. We shall do so in the context of the prediction space setting introduced in Gneiting and Ranjan (2013). That means we explicitly model the information sets of each forecaster. For the sake of simplicity and following Gneiting et al. (2007) and Fissler and Ziegel (2019a) we choose to consider prediction-observation-sequences that are independent and identically distributed over time. Despite this simplification, there is still a variety of parameters to consider in the simulation study:

(i) the dimension of the financial system $d$,
(ii) the (unconditional) distribution of $Y_t$,
(iii) the aggregation function $\Lambda$;
(iv) the scalar risk measure $\rho$;
(v) the competing forecasts $A_t$ and $B_t$, along with their joint distributions with $Y_t$;
(vi) the measure $\pi$ (and thus the scoring function $S_{R, \pi}$);
(vii) the time horizon $N$.

We confine ourselves to the following choices of these parameters.
We work with two different combinations of $Y_t$ and $\Lambda$. In both cases, we work with a system with $d = 5$ participants.

(a) The vector $Y_t$ models the gains and losses of the participants in the system. At any time point $t$, $Y_t = \mu_t + \epsilon_t$ where $\mu_t$ follows a 5-dimensional normal distribution with mean 0, correlations 0.5 and variances 1, and $\epsilon_t$ follows a 5-dimensional standard normal distribution. Thus, conditionally on $\mu_t$, $Y_t$ has distribution $\mathcal{N}_5(\mu_t, I_5)$, whereas unconditionally, $Y_t \sim \mathcal{N}_5(0, \Sigma)$ with $(\Sigma)_{ij} = 0.5$ for $i, j = 1, \ldots, 5$, $i \neq j$ and 2 otherwise. The aggregation function $\Lambda_1$ is of the form $\Lambda_1(Y_t) = (1 - \beta)\sum_{i=1}^{d} Y_{it}^+ - \beta \sum_{i=1}^{d} Y_{it}^-$, as suggested in Amini et al. (2015), and we set $\beta = 0.75$. This way, both gains and losses influence the value of the aggregation function, however, the losses have a higher weight. Here and in what follows, $x^+$ and $x^-$ denote the positive and negative parts of $x$, such that $x^+ = \max(0, x)$ and $x^- = -\min(0, x)$.

(b) We consider an extended model of Eisenberg and Noe (2001); see Feinstein et al. (2017): The participants have liabilities towards each other, $L_{ij,t}$ represents the nominal liability of participant $i$ towards participant $j$ at time point $t$, $i, j = 1, \ldots, 5$. Moreover, each participant $i$ owes an amount $L_{is,t}$ to society at time point $t$. To simplify the simulations and shorten the computing time, we assume that the liabilities matrix is deterministic and constant in time, so that we can write $L_{is}$ instead of $L_{is,t}$. Moreover, we denote by $\bar{L}_s$ the sum of all payments promised to society, i.e., $\bar{L}_s = \sum_{i=1}^{d} L_{is}$. The vector $Y_t$ represents the endowments of the participants at time point $t$. As suggested in Eisenberg and Noe (2001), if some of the endowments are negative, we introduce a so called sink node and interpret the negative endowments as liabilities towards this node. The value of the aggregation function $\Lambda_2$ corresponds to the sum of all payments society obtains in the clearing process as described in Eisenberg and Noe (2001). To simulate the endowments of the participants $Y_t$, we assume that $Y_{it} = (\mu_{it} + \epsilon_{it})^2$ for $i = 1, \ldots, 5$ with $\mu_{it}$ and $\epsilon_{it}$ specified in (a). We construct the system in the following way:

- The probability of a participant owing to another participant is 0.8. If there is a liability from $i$ to $j$, its nominal value is 2.
- In addition, each participant owes 2 to the society.

(iv) In setting (a), we consider the scalar risk measures $\text{VaR}_\alpha$, $\alpha \in (0, 1)$, defined at (6.1), and its expectile-based version defined as $\text{EVaR}_\tau(X) = -e_\tau(X)$, $\tau \in (0, 1)$, where $e_\tau$ satisfies the equation $\tau \mathbb{E}[(X - e_\tau)^+] = (1 - \tau) \mathbb{E}[(X - e_\tau)^-]$ (Newey & Powell, 1987). For the interpretation of expectile-based risk measures in finance we refer to Bellini and Di Bernardino (2015) and to Ehm et al. (2016) for a novel economic angle on expectiles. In case (b), however, the aggregation function $\Lambda_2$ takes nonnegative values only and therefore any financial system would be deemed acceptable when working with $\rho = \text{VaR}_\alpha$ or $\rho = \text{EVaR}_\tau$. Following Feinstein et al. (2017) we overcome this issue by considering the shifted risk measure $\tilde{\rho}(X) = \rho(X) + 0.9\bar{L}_s$, where $\rho = \text{VaR}_\alpha$ or $\rho = \text{EVaR}_\tau$, thus considering the system.
acceptable if \( VaR_{\alpha} \) or \( EVaR_{\tau} \) of the amount that society obtains from the nodes is at most \(-0.9L_s\). Using the standard identification functions for \( VaR_{\alpha} \) and \( EVaR_{\tau} \) (Gneiting, 2011a), the selective identification functions for \( R_0 \) are the following:

- for \( \rho(X) = VaR_{\alpha}(X) + a \):
  \[
  V_{R_0}(k, y) = \alpha - 1\{\Lambda(k + y) - a \leq 0\};
  \tag{7.1}
  \]

- for \( \rho(X) = EVaR_{\tau}(X) + a \):
  \[
  V_{R_0}(k, y) = \tau(\Lambda(k + y) - a)^+ - (1 - \tau)(\Lambda(k + y) - a)^-.
  \tag{7.2}
  \]

(v) We consider two ideal forecasters with different information sets: Anne has access to \( \mu_t \) and uses the correct conditional distribution of \( Y_t \). That is, she issues \( A_t = R(N_5(\mu_t, I_5)) = R(N_5(0_5, I_5)) - \mu_t \) in case (a) and \( A_t = R(N_5(\mu_t, I_5)^2) \)\(^{10}\) in case (b) for each \( t = 1, \ldots, N \). Bob is uninformed and issues the climatological forecast. That is, he uses the correct unconditional distribution of \( Y_t \) for his forecasts. Therefore, he constantly predicts \( B_t = R(N(0_t, \Sigma)) \) in case (a) and \( B_t = R(N(0_t, \Sigma)^2) \) in case (b).

(vi) We choose \( \pi \) to be a 5-dimensional Gaussian measure with mean \( m \in \mathbb{R}^5 \) and covariance \( I_5 \). To enhance the discrimination ability of the score \( S_{R,\pi} \), we aim at choosing \( m \) close to the boundary of \( R(Y_t) \). Here we work with \( m = 2 \cdot 1 \) as this value appears to be fairly close to the (deterministic) forecasts of Bob in all four cases. This choice of \( \pi \) turns out to be beneficial with respect to the integrability considerations and renders our scores finite. Indeed, since \( V_{R_0} \) for \( \rho = VaR_{\alpha} + a \) is bounded, it is \( \pi \otimes F \)-integrable for any finite measure \( \pi \). In the case of \( \rho = EVaR_{\tau} + a \), more considerations are necessary. From the construction of \( A_2 \) it is clear that it is a bounded function, in particular, the values lie in the interval \( [0, \sum_{i=1}^{d} L_{i8}] \). This in turn implies that the identification function \( V_{R_0} \) is bounded. Therefore \( V_{R_0} \) is \( \pi \otimes F \)-integrable for any finite measure \( \pi \). Finally, since \( A_1 \) only grows linearly and both \( \pi \) and \( Y_t \) are Gaussian, the integrability is also guaranteed in this case.

(vii) We work with sample sizes \( N = 250 \), being a good proxy for the number of working (and trading) days in a year.

To compare Anne’s with Bob’s forecast performance, we employ the classical Diebold- Mariano test (Diebold & Mariano, 1995) based on the scoring functions \( S_{R,\pi} \) of the form at (4.12) arising from our choice of \( \pi \) and identification functions introduced in (7.1) and (7.2). We repeat the experiment 1 000 times for setting (a) and 100 times for setting (b).\(^{11}\) We approximate \( \pi \) with a Monte Carlo draw of size 100 000. The computations are performed with the statistics software \( \mathbb{R} \), and in particular its \texttt{Rcpp} package to also integrate parts of \texttt{C++} code to enhance the computational speed.

\(^{10}\)We use the notation \( N_d(m, \Sigma)^2 \) for the distribution of a random variable \( Y = X^2 \) where \( X \sim N_d(m, \Sigma) \).

\(^{11}\)Due to the presence of clearing, the computation time tends to be quite lengthy in setting (b).
That is, $\text{VaR}$ only consider follows a 2-dimensional standard normal distribution. As the scalar risk measure $\rho \in \mathbb{R}^2$, we define it and issues sign-reversed forecasts $\mu_f^{\top} Y_t$. Moreover, both in case (a) and (b), the number of instances when Anne’s forecasts are 250. This might be due to the fact that $\Lambda_B$ is considerably higher for model (a) as opposed to (b) where we yield rejection rates ranging from 0.74 to 0.90. This might be due to the fact that $\Lambda_1$ is unbounded whereas $\Lambda_2$ only takes values between 0 and $\bar{L}_s$, which might translate into a smaller influence of the predictive distributions upon which the forecasts are based. Moreover, both in case (a) and (b), the number of instances when Anne’s forecasts are preferred over Bob’s ones is higher for $\rho = EVaR_\alpha + a$ than for $\rho = VaR_\alpha + a$.

### 7.2. Murphy diagrams

In this subsection, we illustrate the use of Murphy diagrams, following Corollary 4.13. To allow for graphical illustrations, we reduce the dimension to $d = 2$, translating case (a) of subsection 7.1 to $d = 2$. In particular, we have $Y_t = \mu_t + \epsilon_t$ where $\mu_t$ follows a 2-dimensional normal distribution with mean 0, variances 1 and correlations 0.5, and $\epsilon_t$ follows a 2-dimensional standard normal distribution. As the scalar risk measure $\rho$ we only consider $VaR_{0.05}$ and we use the aggregation function $\Lambda_1 : \mathbb{R}^2 \to \mathbb{R}$, i.e., $\Lambda(x) = 0.25(x_1^+ + x_2^+) - 0.75(x_1^- + x_2^-)$. Besides focused Anne and climatological Bob introduced above both using their respective information sets ideally, we also consider Celia. Just like Anne, Celia has access to $\mu_\xi$ resulting in the same information set. However, she misinterprets it and issues sign-reversed forecasts $C_t$ assuming that $Y_t \sim \mathcal{N}_2(-\mu_t, I_2)$. That is, $C_t = R(\mathcal{N}_2(-\mu_t, I_2)) = R(\mathcal{N}_2(0, I_2)) + \mu_t$. Again, we consider a time horizon of $N = 250$.

In the left panel of Figure 2 we illustrate the differences of empirical Murphy diagrams $[-5, 5]^2 \ni k \mapsto \hat{s}_{250, f_1}(k) - \hat{s}_{250, f_2}(k) = \frac{1}{250} \sum_{t=1}^{250} S_{R,k}(f_{1t}, Y_t) - S_{R,k}(f_{2t}, Y_t)$ where $f_{1t}$, $f_{2t}$, $f_{3t}$, and $f_{4t}$.

| $H_0$ | $VaR_{0.01}$ | $VaR_{0.05}$ | $EVaR_{0.01}$ | $EVaR_{0.05}$ |
|------|--------------|--------------|----------------|--------------|
| $\Lambda_1$ | $A \succeq B$ | 0.995 | 0.940 | 1.000 | 1.000 |
| $A \preceq B$ | 0.000 | 0.000 | 0.000 | 0.000 |
| $\Lambda_2$ | $A \succeq B$ | 0.740 | 0.870 | 0.790 | 0.900 |
| $A \preceq B$ | 0.000 | 0.000 | 0.000 | 0.000 |

Table 1: Ratios of rejections of the null hypotheses at significance level 0.05.
Figure 2: Left panel: Differences of empirical Murphy diagrams $\hat{s}_{250, f_1}(k) - \hat{s}_{250, f_2}(k)$ from (4.16) versus $k \in \mathbb{R}^2$. Right panel: Three-zone traffic light illustration of pointwise comparative backtests following Fissler et al. (2016). The green area corresponds to the region where the null $H_0^+: f_1 \preceq f_2$ is rejected, the red one is where $H_0^-: f_1 \succeq f_2$ is rejected, at level 0.05, respectively. Yellow means that neither $H_0^+$ nor $H_0^-$ are rejected. In the grey region, the two Murphy diagrams identically coincide.
\(f_{2t}\) stand for one of the three considered forecasts, \(A_t\), \(B_t\) or \(C_t\). In each pairwise comparison, we choose \(f_{1t}\) to be inferior to \(f_{2t}\) such that we expect a non-negative difference of the corresponding Murphy diagrams. Indeed, only in the comparison between Bob and Anne, there are some \(k\) where \(\hat{s}_{250,f_1}(k) - \hat{s}_{250,f_2}(k) < 0\). For the remaining regions and situations, the Murphy diagrams behave consistently with our expectations.

For all three pairwise comparisons, one can nicely recognise the region where the respective two forecasts differ, resulting in a positive score difference depicted in blue. This bluish region seems to correspond to a blurred version of the boundary of the considered risk measure. Interestingly, while these regions illustrating positive score differences are similar in shape and location for the two pairs involving Celia, this region seems to be slightly translated to the upper right corner in the comparison between Anne and Bob. Quite intuitively, the magnitude of the score difference with a maximum of approximately 0.05 is smaller in the joust between the two ideal forecasts issued by Bob and Anne in comparison to the situations involving the sign-reversed Celia where the maximal difference between the Murphy diagrams is larger than 0.15. We have performed this experiment several times and observed that the stylised facts are qualitatively stable. For transparency reasons, we have depicted the first experiment performed, but we report some more experiments in Appendix C.

In the right panel of Figure 2 we depict the results of pointwise comparative backtests using the traffic-light illustration suggested in Fissler et al. (2016), which is akin to the three-zone approach of the Bank for International Settlements (2013, pp. 103–108). That is, we perform a Diebold-Mariano test using the elementary score \(S_{R,k}\) for each \(k\) in a grid of \([-5,5]^2\). This means, we would like to see whether the superiority of the forecasts \(f_{1t}\) is recognised at a significance level of 0.05 deploying the two possible one-sided null hypotheses \(H_0^+: f_1 \preceq f_2\) and \(H_0^-: f_1 \succeq f_2\), using the notation introduced in the previous subsection. If for a certain \(k\) the null \(H_0^+\) is rejected deeming \(f_2\) significantly superior to \(f_1\), we colour the corresponding \(k\) in green. Similarly, if the null \(H_0^-\) is rejected, considering \(f_1\) to be superior to \(f_2\), we illustrate \(k\) in red. For all \(k\) in the yellow region, none of the two nulls is rejected, meaning that the procedure is indecisive at the significance level 0.05. Finally, the grey area corresponds to those points where the score difference is constantly zero for all \(t = 1, \ldots, N\). Due to the vanishing variance, a Diebold-Mariano test is apparently not possible there. But clearly, this still means that the two forecasts are just equally good in that region.

The specific results nicely correspond to the situations obtained in the left panel of Figure 2. For all three pairwise comparisons and for \(k\) close to the four corners of the area \([-5,5]^2\), the score differences identically vanish, resulting in a grey colouration. Again, in all three cases, there is a “continuous” behaviour in that the grey region adjoins a yellow stripe before turning into a fairly broad green stripe. For the comparisons involving Celia, clearly using an inferior predictive distribution to both Anne’s and Bob’s, it is reassuring that a substantial region is coloured in green. In this region, the procedure is decisive, deeming Celia significantly inferior to Anne and to Bob. Moreover, for this particular simulation, there is no red region. The situation comparing the two ideal forecasters Anne and Bob is somewhat more involved. While most of the previous observations also apply to that situation, there is a small red stripe close to the upper right
corner. For $k$ in that region and for this particular simulation, this means that Bob’s forecasts outperform Anne’s ones. While this observation is somewhat unexpected, it reflects the finite sample nature of the simulation, rendering such outcomes possible. Having a look at some more experiments, the results of which are again reported in Appendix C, shows that this red region is not stable over different simulations (which would clearly violate the sensitivity of consistent scoring functions with respect to increasing information sets established in Holzmann and Eulert (2014)), but it moves and occasionally also vanishes (on the region $[-5,5]^2$ considered). Interestingly, in all events with a red region present, this red region was still roughly located in a similar area.

8. Discussion

As mentioned in the introduction, the aims and novel contributions of this paper are two-fold. On the one hand, we introduce a principled framework for the assessment of forecasts for set-valued quantities: an exhaustive notion where forecasts specify the entire set of interest versus a selective notion where forecasters are content with issuing a single point in the set of interest. We unveil the structural connection between these two alternative notions, notably their mutual exclusivity (Corollary 2.13). The other main contribution consists of establishing selective identifiability results in Theorem 4.1 and exhaustive elicitation results in Theorem 4.9 for systemic risk measures sensitive with respect to capital allocations. Notably, the construction of consistent exhaustive scoring functions relies on a mixture representation of easily computable elementary scores, which opens the way to the diagnostic tool of Murphy diagrams.

While the structural insights and the framework for evaluating forecasts for set-valued quantities might help to enhance the academic and applied avenue of research in that area (see Subsection 2.5), we would like to briefly outline specific fields where the identifiability and elicitation results of this paper can be of advantage.

**Backtesting.** A strictly consistent exhaustive scoring function $S_R$ for a systemic risk measure $R$ can be used for comparative backtests of competing exhaustive forecasts, that is, set-valued forecasts, as described in Fissler et al. (2016) and Nolde and Ziegel (2017); see also Subsection 7. More precisely, having competing forecasts $A_1,\ldots,A_N \in \mathcal{P}(R^d;R^d_+), B_1,\ldots,B_N \in \mathcal{P}(R^d;R^d_+)$, and verifying observations of the gains and losses of the financial system $Y_1,\ldots,Y_N \in R^d$, one can consider a properly normalised version of the test statistic

$$\frac{1}{N} \sum_{t=1}^N S_R(A_t,Y_t) - S_R(B_t,Y_t)$$

to assess which forecast sequence is superior under $S_R$.

On the other hand, the fact that one can construct oriented selective identification functions for risk measures might open the way to one-sided traditional backtests (Nolde & Ziegel, 2017, Subsection 2.21). That is, if one has a sequence of vector-valued predictions for capital requirements $k_1,\ldots,k_N \in \mathbb{R}^d$ along with verifying observations $Y_1,\ldots,Y_N \in \mathbb{R}^d$, one might wonder if the forecasted capital requirements are adequate.
to eliminate the risk of the financial system under $R$. That means, we would like to judge if $k_t \in R(Y_t)$ for all $t = 1, \ldots, N$ with a certain level of certainty $\alpha$. If $V_{R_0}$ is an oriented strict selective identification function for $R_0$, this amounts to testing the one-sided null hypothesis

$$H_0: E[V_{R_0}(k_t, Y_t)] \leq 0 \text{ for all } t = 1, \ldots, N.$$ 

Under suitable mixing conditions, one can construct an (asymptotic) level $\alpha$ test for this null hypothesis by considering a rescaled version of the test statistic

$$\frac{1}{N} \sum_{t=1}^{N} V_{R_0}(k_t, Y_t).$$

Note that from a regulatory perspective, testing this one-sided null hypothesis is more sensible than testing the two-sided null

$$H'_0: E[V_{R_0}(k_t, Y_t)] = 0 \text{ for all } t = 1, \ldots, N.$$ 

Indeed, this corresponds to assessing whether $\rho(\Lambda(Y_t + k_t)) = 0$. However, overestimating the financial requirements to make the system acceptable is even more prudent from a regulatory angle.

**M-Estimation.** If a systemic risk measure $R$ is exhaustively elicitable, one can make inference for it in the form of $M$-estimation (Huber & Ronchetti, 2009). That is, if one has a sample $Y_1, \ldots, Y_N \in \mathbb{R}^d$ of stationary observations fulfilling sufficient mixing conditions, one might estimate the set-valued risk measure $R(Y) \in F(\mathbb{R}^d; \mathbb{R}^d_+)$, where $Y$ has the same distribution as the observations, via

$$\hat{R}(Y) = \arg \min_{A \in F(\mathbb{R}^d; \mathbb{R}^d_+)} \frac{1}{N} \sum_{t=1}^{N} S_R(A, Y_t),$$

where $S_R: F(\mathbb{R}^d; \mathbb{R}^d_+ \times \mathbb{R}^d) \rightarrow \mathbb{R}$ is a strictly consistent exhaustive scoring function for $R$. Under suitable conditions, the $M$-estimator $\hat{R}(Y)$ at (8.1) is consistent for $R(Y)$. However, computationally, the optimisation problem at (8.1) might be rather expensive, if feasible at all. The reason is that one needs to optimise over the collection of all closed upper sets of $\mathbb{R}^d$.

**Regression.** A closely connected concept to the notion of $M$-estimation is regression where it is possible to bypass the complication to optimise over a collection of sets. Consider a time series $(X_t, Y_t)_{t \in \mathbb{N}}$. Sticking to the usual denomination, let $Y_t$ denote the response variable, taking values in $\mathbb{R}^d$, and let $X_t$ be a $p$-dimensional vector of regressors. The regressors might consist of quantities which seem relevant to the systemic risk of the financial system. Examples might consist of macroeconomic quantities such as GDP, unemployment, inflation, net-investments etc. Let $\Theta \subseteq \mathbb{R}^q$ be a parameter space and let $M: \mathbb{R}^p \times \Theta \rightarrow F(\mathbb{R}^d; \mathbb{R}^d_+)$ be a parametric model taking values in the collection of
closed upper subsets of $\mathbb{R}^d$. Suppose the model is correctly specified in that there exists a unique parameter $\theta_0 \in \Theta$ such that

$$R(F_{Y_t|X_t}) = M(X_t, \theta_0) \quad \mathbb{P}\text{-a.s. for all } t \in \mathbb{N}. \tag{8.2}$$

Here, $R: \mathcal{M}^d \to \mathcal{F}(\mathbb{R}^d; \mathbb{R}^d_+)$ is a law-invariant risk measure of the form at (3.1) satisfying the conditions of Theorem 4.9(iii). Further, suppose that the (regular version of the) conditional distribution $F_{Y_t|X_t}$ of $Y_t$ given $X_t$ is an element of $\mathcal{M}^d_0$ almost surely, where we use the notation of Theorem 4.9. Note that the time series does not need to be strongly stationary, but only the conditional distribution $F_{Y_t|X_t}$ needs to satisfy the ‘semi-parametric stationarity condition’ specified via (8.2). Let $S_{R,\pi}$ be a strictly $\mathcal{M}^d_0$-consistent exhaustive scoring function for $R$. Then, under certain mixing and integrability assumptions specified in White (2001, Corollary 3.48) one yields the following Law of Large Numbers

$$\frac{1}{N} \sum_{t=1}^{N} \left\{ S_{R,\pi} (M(X_t, \theta), Y_t) - \mathbb{E} [S_{R,\pi} (M(X_t, \theta), Y_t)] \right\} \to 0 \quad \mathbb{P}\text{-a.s. as } N \to \infty$$

for all $\theta \in \Theta$. It is essentially a uniform version (in the parameter $\theta$) of this Law of Large Numbers result which yields the consistency for the empirical estimator

$$\hat{\theta}_N := \arg \min_{\theta \in \Theta} \frac{1}{N} \sum_{t=1}^{N} S_{R,\pi} (M(X_t, \theta), Y_t) \tag{8.3}$$

for $\theta_0$; see van der Vaart (1998), Huber and Ronchetti (2009) or Nolde and Ziegel (2017) for details. The advantage of this regression approach in comparison to $M$-estimation is that the optimisation at (8.3) needs to be performed over a subset $\Theta$ of $\mathbb{R}^q$ only (which is often assumed to be compact). This makes the result computationally a lot more feasible than the optimisation procedure over a collection of upper sets.\(^{12}\)

Besides the usual practical challenge of constructing reasonable parametric models $M$ to model the systemic risk of a financial system $Y_t$ given regressors $X_t$, we see some interesting theoretical problems related to this regression framework. While, under correct model specification given at (8.2), any strictly consistent scoring function $S_{R,\pi}$ induces a consistent estimator $\hat{\theta}_N$ at (8.3), the estimator will generally depend on the choice of $S_{R,\pi}$ (or $\pi$) in finite samples. Moreover, the efficiency of the estimator $\hat{\theta}_N$, expressed in terms of the asymptotic variance of $\sqrt{N}(\hat{\theta}_N - \theta_0)$, will depend on the choice of $S_{R,\pi}$, suggesting an interesting optimality criterion for $S_{R,\pi}$.

A very modern and interesting approach circumventing this issue is to perform regression simultaneously with respect to the class of all consistent scoring functions (or a reasonably large subclass), which is explored in the recent paper Jordan et al. (2019). To perform this efficiently, the mixture representation of scoring functions in terms of elementary scores might prove beneficial. We defer this interesting problem to future research.

\(^{12}\)In other words, $M$-estimation can be considered as a special instance of regression where the regressor $X_t$ is constant and where $\Theta$ corresponds to $\mathcal{F}(\mathbb{R}^d; \mathbb{R}^d_+)$. 

45
Acknowledgement

We would like to express our sincere gratitude to Tilmann Gneiting and Johanna Ziegel for insightful discussions about the topic, to Rafael Frongillo and Alexander Jordan who coined the terminology of exhaustive versus selective elicitability in a joint discussion, to Dario Azzimonti, Peter Barančok, Zied Ben Bouallègue, Seamus Bradley, Jonas Brehmer, Timo Dimitriadis, David Ginsbourger and Ilya Molchanov for valuable discussions and helpful references, to Yuan Li for a careful proofreading of an earlier version of this paper, and to Lukáš Šablica for helpful programming advice on the simulation part of this project.

Tobias Fissler gratefully acknowledges financial support from Imperial College London via his Chapman Fellowship. He also enjoyed the hospitality of the Institute for Statistics and Mathematics at Vienna University of Economics and Business during several research visits when main parts of the project have jointly been developed.

Appendix

A. Systemic risk measures insensitive to capital allocations

In this appendix, we state some elicitability and identifiability results for systemic risk measures $R^{ins}$ defined at (3.2) which are insensitive with respect to capital allocation. They are similar to some results of Section 4, however often less technically involved. To this end, we first introduce the following law-invariant risk measures connected to $R^{ins}$:

$$r: Y^d \rightarrow \mathbb{R}, \quad Y \mapsto r(Y) = \rho(\Lambda(Y)), \quad \text{(A.1)}$$

$$R^{ins}_0: Y^d \rightarrow 2^{\mathbb{R}^d}, \quad Y \mapsto R^{ins}_0(Y) = \{ \tilde{k} \in \mathbb{R}^d | \rho(\Lambda(Y) + \tilde{k}) = 0 \}, \quad \text{(A.2)}$$

where we recall the shorthand $\tilde{k} = \sum_{i=1}^d k_i$ for some vector $k = (k_1, \ldots, k_d) \in \mathbb{R}^d$. Due to the cash-invariance of $\rho$, $R^{ins}_0$ is a bijection of $r$. Note that $R^{ins}$ is always a closed half-space above the hyperplane with normal $(1, \ldots, 1)^\top \in \mathbb{R}^d$ and $R^{ins}_0$ corresponds to its topological boundary. This implies a one-to-one relationship between $R^{ins}_0$ and $R^{ins}$. Thanks to these facts, one can make use of an extension of the so called revelation principle which originates from Osband’s (1985) seminal thesis.

**Proposition A.1** (Revelation principle). Let $T: \mathcal{M}^d \rightarrow \mathbb{R}$ be an identifiable and elicitable functional, $g: \mathbb{R}^d \rightarrow \mathbb{R}$ some map, and $h: A \rightarrow \mathbb{R}, A \subseteq 2^{\mathbb{R}^d}$ a bijection with inverse $h^{-1}$. Then the following assertions hold true:

(i) $T: \mathcal{M}^d \rightarrow \mathbb{R}$ is identifiable if and only if $T_{h^{-1}} = h^{-1} \circ T: \mathcal{M}^d \rightarrow A$ is exhaustively identifiable. The function $V: \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a strict $\mathcal{M}^d$-identification function for $T$ if and only if

$$V_{h^{-1}}: A \times \mathbb{R}^d \rightarrow \mathbb{R}, \quad (B, y) \mapsto V(h^{-1}(B), y)$$

is a strict $\mathcal{M}^d$-identification function for $T_{h^{-1}} = h^{-1} \circ T: \mathcal{M}^d \rightarrow A$. 

46
(ii) $T : \mathcal{M}^d \to \mathbb{R}$ is elicitable if and only if $T_{h^{-1}} = h^{-1} \circ T : \mathcal{M}^d \to \mathbb{A}$ is exhaustively elicitable. The function $S : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is a strictly $\mathcal{M}^d$-consistent scoring function for $T$ if and only if

$$S_{h^{-1}} : \mathbb{A} \times \mathbb{R}^d \to \mathbb{R}, \quad (B, y) \mapsto S(h^{-1}(B), y)$$

is a strictly $\mathcal{M}^d$-consistent exhaustive scoring function for $T_{h^{-1}} = h^{-1} \circ T : \mathcal{M}^d \to \mathbb{A}$.

(iii) If $S : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is a (strictly) consistent scoring function for $T$, then

$$S_g^{-1} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}, \quad (x, y) \mapsto S(g(x), y)$$

is a (strictly) $\mathcal{M}^d$-consistent selective scoring function for $T_g^{-1} = g^{-1} \circ T : \mathcal{M}^d \to 2^{\mathbb{R}^d}$.

(iv) If $V : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is a (strict) $\mathcal{M}^d$-identification function for $T$, then

$$V_g^{-1} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}, \quad (x, y) \mapsto V(g(x), y)$$

is a (strict) selective $\mathcal{M}^d$-identification function for $T_g^{-1}$.

(v) If $V : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is an oriented strict identification function for $T$ and $g$ is strictly increasing with respect to the componentwise order, then $V_g^{-1} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ is an oriented strict selective $\mathcal{M}^d$-identification function for $T_g^{-1}$ in the following sense:

$$\bar{V}_{g^{-1}}(x, Y) \begin{cases} < 0, & \text{if } x \in (g^{-1}(\{T(F)\}) - \mathbb{R}_+^d) \setminus g^{-1}(\{T(F)\}) \\ = 0, & \text{if } x \in g^{-1}(\{T(F)\}) \\ > 0, & \text{if } x \in (g^{-1}(\{T(F)\}) + \mathbb{R}_+^d) \setminus g^{-1}(\{T(F)\}). \end{cases}$$

Proof. (i)–(ii) These statements are a special case of Lemma 2.3.2 in Fissler (2017).

(iii) This is a special case of Lemma 2.3.3 in Fissler (2017).

(iv) Let $F \in \mathcal{M}^d$. If $T_{g^{-1}}(F) = g^{-1}(\{T(F)\}) = \emptyset$, there is nothing to show. Assume that $x \in T_{g^{-1}}(F)$. Then we have $g(x) = T(F)$ and thus $\bar{V}_{g^{-1}}(x, F) = \tilde{V}(g(x), F) = 0$. Moreover, if $x' \notin T_{g^{-1}}(F)$, we have $g(x') \neq T(F)$ and thus if $V$ is a strict identification function for $T$, we have $\bar{V}_{g^{-1}}(x', F) = \tilde{V}(g(x'), F) \neq 0$.

(v) From the previous part we already have $\bar{V}_{g^{-1}}(x, F) = 0$ for $x \in g^{-1}(\{T(F)\})$. Now assume $x \in (g^{-1}(\{T(F)\}) - \mathbb{R}_+^d) \setminus g^{-1}(\{T(F)\})$. Then there is some $x' \in g^{-1}(\{T(F)\})$ such that $x \leq x'$ componentwise and $x \neq x'$ and thus $g(x) < g(x') = T(F)$. Therefore, using the orientation of $V$, we get $\bar{V}_{g^{-1}}(x, F) < 0$. The last part follows by similar considerations.

\[ \square \]

**Lemma A.2.** Let $\rho : \mathcal{M} \to \mathbb{R}$ be identifiable and elicitable with a strict $\mathcal{M}$-identification function $V_{\rho} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and a strictly $\mathcal{M}$-consistent scoring function $S_{\rho} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Then the following assertions hold for $r : \mathcal{M}^d \to \mathbb{R}$ defined at (A.1):

47
(i) $r$ is identifiable and
\[ V_r : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}, \quad (x, y) \mapsto V_\rho(x, \Lambda(y)) \] (A.3)
is a strict $\mathcal{M}^d$-identification function for $r$.

(ii) If $V_\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is an oriented strict $\mathcal{M}$-identification function for $\rho$, then $V_r$ defined at (A.3) is oriented for $r$.

(iii) $r$ is elicitable and
\[ S_r : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}, \quad (x, y) \mapsto S_\rho(x, \Lambda(y)) \]
is a strictly $\mathcal{M}^d$-consistent scoring function for $r$.

\textbf{Proof.} Obvious. \hfill \Box

\textbf{Corollary A.3.} Let $\rho : \mathcal{Y} \to \mathbb{R}$ be identifiable and elicitable with a strict $\mathcal{M}$-identification function $V_\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and a strictly $\mathcal{M}$-consistent scoring function $S_\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$.

Then the following assertions hold for $R_{\text{ins}}$ defined at (3.2) and $R_{\text{ins}}^0$ defined at (A.2):

(i) $R_{\text{ins}}$ is exhaustively identifiable and exhaustively elicitable. Define $\bar{k}_{\text{min}}(B) = \min \{\bar{k} | k \in B\}$ for $B \in \mathcal{A}$ where $\mathcal{A}$ is the collection of all closed half-spaces above the hyperplane with normal $(1, \ldots, 1)\top \in \mathbb{R}^d$. Then
\[ V_{R_{\text{ins}}} : \mathcal{A} \times \mathbb{R}^d \to \mathbb{R}, \quad (B, y) \mapsto V_\rho(\bar{k}_{\text{min}}(B), y), \]
\[ S_{R_{\text{ins}}} : \mathcal{A} \times \mathbb{R}^d \to \mathbb{R}, \quad (B, y) \mapsto S_\rho(\bar{k}_{\text{min}}(B), y) \]
are a strict exhaustive $\mathcal{M}^d$-identification function and a strictly $\mathcal{M}^d$-consistent exhaustive scoring function for $R_{\text{ins}}$, respectively.

(ii) $R_{\text{ins}}^0$ is selectively identifiable and
\[ V_{R_{\text{ins}}^0} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \quad (k, y) \mapsto V_{R_{\text{ins}}^0}(k, y) = V_\rho(\bar{k}, \Lambda(y)) \]
is a strict selective $\mathcal{M}^d$-identification function for $R_{\text{ins}}^0$. Moreover, if $V_\rho$ is oriented, $V_{R_{\text{ins}}^0}$ is oriented in the sense defined at (4.2).

(iii) $R_{\text{ins}}^0$ is selectively elicitable and
\[ S_{R_{\text{ins}}^0} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \quad (k, y) \mapsto S_{R_{\text{ins}}^0}(k, y) = S_\rho(\bar{k}, \Lambda(y)) \]
is a strictly $\mathcal{M}^d$-consistent selective scoring function for $R_{\text{ins}}^0$.

\textbf{Proof.} This is a direct consequence of Proposition A.1 and Lemma A.2. \hfill \Box

\textbf{Proposition A.4.} Let $\rho : \mathcal{Y} \to \mathbb{R}$ be a risk measure, $\Lambda : \mathbb{R}^d \to \mathbb{R}$ an aggregation function and $r : \mathcal{Y}^d \to \mathbb{R}$ as in (A.1). Assume that there exists a measurable right inverse $\eta : \Lambda(\mathbb{R}^d) \to \mathbb{R}^d$ such that $\Lambda \circ \eta = \text{id}_\mathbb{R}$, and for any $X \in \mathcal{Y}$, $\eta(X)$ belongs to $\mathcal{Y}^d$. Then $\rho$ is identifiable if and only if $r$ is identifiable.

\textbf{Proof.} The proof is analogous to the proof of Proposition 4.8. \hfill \Box

48
B. Proofs of the main part

B.1. Proofs of Section 2

Proof of Proposition 2.7. Let \( S_M \) (\( S_{M_1} \)) be the class of strictly \( M \)-consistent \((M_1\)-consistent) scoring functions for \( T \). It holds that

\[
S_M \subseteq S_{M_1} = S'_{M_1} = S'_M.
\]

However, any \( S' \in S'_{M_1} \) fails to be strictly \( M \)-consistent for \( T \). Hence, \( S_M = \emptyset \). \( \square \)

Proof of Proposition 2.10. Let \( F_0, F_1 \in M \), \( \lambda \in (0, 1) \) and define \( F_\lambda = (1 - \lambda)F_0 + \lambda F_1 \). Let \( S: A_{sel} \times O \to \mathbb{R}^* \) be a strictly \( M \)-consistent selective scoring function for \( T \). If \( T(F_0) \cap T(F_1) = \emptyset \), there is nothing to show. So let \( t \in T(F_0) \cap T(F_1) \neq \emptyset \). Moreover, let \( x \in A_{sel} \). 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 \( M \)-consistency of \( S \). This implies that

\[
\bar{S}(x, F_\lambda) - \bar{S}(t, F_\lambda) = (1 - \lambda)(\bar{S}(x, F_0) - \bar{S}(t, F_0)) + \lambda(\bar{S}(x, F_1) - \bar{S}(t, F_1)) \tag{B.1}
\]

The identity in (B.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 \( M \)-consistency of \( S \), the assertion follows. \( \square \)

Proof of Theorem 2.12. Assume \( S \) is a strictly \( M \)-consistent exhaustive scoring function for \( T \). Let \( F, G \in M \) such that \( \emptyset \neq T(G) \subset T(F) \). Then

\[
\bar{S}(T(F), F) - \bar{S}(T(G), F) < 0 < \bar{S}(T(F), G) - \bar{S}(T(G), G).
\]

For any \( \lambda \in (0, 1) \) the selective CxLS* property implies that \( T((1-\lambda)F + \lambda G) = T(G) \). Then there is a sufficiently small \( \lambda_0 \in (0, 1) \) such that

\[
\bar{S}(T(F), (1 - \lambda_0)F + \lambda_0 G) - \bar{S}(T(G), (1 - \lambda_0)F + \lambda_0 G)
= (1 - \lambda_0)(\bar{S}(T(F), F) - \bar{S}(T(G), F)) + \lambda_0(\bar{S}(T(F), G) - \bar{S}(T(G), G)) < 0,
\]

which violates the strict \( M \)-consistency. \( \square \)

Proof of Corollary 2.13. Implication (i) is an immediate consequence of Proposition 2.10 and Theorem 2.12. Implication (ii) is merely the contraposition of (i). \( \square \)
Proof of Proposition 2.16. Let $T_{\text{sel}} \colon \mathcal{F} \to \mathbb{A}$ be a selection of $T$ and suppose $S \colon \mathbb{A} \times \mathcal{O} \to \mathbb{R}^*$ is a strictly $\mathcal{M}$-consistent scoring function for $T_{\text{sel}}$. Let $F, G, H \in \mathcal{M}$ satisfy (2.7). If $t_0$ and $t_1$ are topologically distinguishable, the selective CxLS* property implies that for any $\lambda \in (0, 1)$ we have that $t_1 = T_{\text{sel}}((1 - \lambda)F + \lambda G)$ and $t_2 = T_{\text{sel}}((1 - \lambda)F + \lambda H)$. Then, for $t_0 = T_{\text{sel}}(F)$ we have that $t_0 \neq t_1$ or $t_0 \neq t_2$. Without loss of generality, assume $t_0 \neq t_1$. Then the map $\gamma : [0, 1] \to \mathbb{A}, \lambda \mapsto T_{\text{sel}}((1 - \lambda)F + \lambda G)$ is not continuous,\footnote{Let $U \subset \mathbb{A}$ 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 $[0, 1]$.} which shows that $T_{\text{sel}}$ is not mixture-continuous. Moreover, $\gamma$ is neither injective nor constant, such that Lemma B.1 in Fissler and Ziegel (2019c) implies that $T_{\text{sel}}$ is not identifiable. Finally, the strict $\mathcal{M}$-consistency of $S$ for $T_{\text{sel}}$ implies that for all $\lambda \in (0, 1)$

$$S(t_0, F) - S(t_1, F) < 0 < S(t_0, (1 - \lambda)F + \lambda G) - S(t_1, (1 - \lambda)F + \lambda G).$$

This contradicts the elementary fact that the map $[0, 1] \ni \lambda \mapsto S(t_0, (1 - \lambda)F + \lambda G) - S(t_1, (1 - \lambda)F + \lambda G)$ is continuous (and even affine). \hfill \qed

B.2. Proofs of Section 4

Proof of Theorem 4.1. (i) Let $V_\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a strict $\mathcal{M}$-identification function for $\rho$. That means for all $Y \in \mathcal{Y}^d$ with distribution $F \in \mathcal{M}^d$, for all $k \in \mathbb{R}^d$ and for all $x \in \mathbb{R}$, one has that

$$E_F[V_\rho(x, \Lambda(Y + k))] = 0 \iff x = \rho(\Lambda(Y + k)). \quad (B.2)$$

Setting $x = 0$ in (B.2) yields

$$E_F[V_\rho(0, \Lambda(Y + k))] = 0 \iff 0 = \rho(\Lambda(Y + k)) \iff k \in R_0(Y),$$

which holds in particular for $R_0(Y) = \emptyset$. Therefore $V_{R_0}(k, y) = V_\rho(0, \Lambda(y + k))$ is a strict selective $\mathcal{M}$-identification function for $R_0$.

(ii) Now assume that $V_\rho : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is an oriented strict $\mathcal{M}$-identification function for $\rho$. That means for all $Y \in \mathcal{Y}^d$ with distribution $F \in \mathcal{M}^d$, for all $k \in \mathbb{R}^d$ and for all $x \in \mathbb{R}$, one has that

$$E_F[V_\rho(x, \Lambda(Y + k))] \begin{cases} < 0, & \text{if } x < \rho(\Lambda(Y + k)) \\ = 0, & \text{if } x = \rho(\Lambda(Y + k)) \\ > 0, & \text{if } x > \rho(\Lambda(Y + k)). \end{cases} \quad (B.3)$$

Setting $x = 0$ in (B.3) yields the claim. \hfill \qed

Proof of Proposition 4.4. The proof follows along the lines of the proof of Theorem 3.2 in Fissler and Ziegel (2016); cf. Osband (1985). The dimensionality of $x$ does not play
Figure 3: A graphical illustration of the proof of Proposition 4.6 for dimension \(d = 2\). Suppose the blue region corresponds to the correctly specified risk measure \(R(F)\). In the left picture, \(EAR_w(F)\) is a singleton, containing only point \(A\). Point \(B\) corresponds to case (i), whereas points \(C\) and \(D\) are examples of case (ii) for points that are not in \(EAR_w(F)\). In the right picture, \(EAR_w(F) = \emptyset\).

For any \(k \in \mathbb{R}^d\) there is some \(x \in w^\perp\) such that \(\bar{V}_{R_0}(k + x, F) > 0\).

As our identification functions map to \(\mathbb{R}\), we use \(k = 1\) in the proof of Theorem 3.2 of Fissler and Ziegel (2016). The assumption on the existence of \(F_1, F_2 \in \mathcal{M}^d\) such that the signs of \(\bar{V}_{R_0}\) are different plus the convexity of \(\mathcal{M}^d\) are equivalent to Assumption (V1) in Fissler and Ziegel (2016). If we replace \(\nabla \bar{S}(x, F)\) by \(\bar{V}'(x, F)\), we obtain that there is a function \(h: A \to \mathbb{R}\) such that \(\bar{V}'(x, F) = h(x) \bar{V}(x, F)\) for all \(x \in A\) and all \(F \in \mathcal{M}^d\). Since the matrix \(B_G\) in the proof will be a \(2 \times 3\) matrix of rank 1 for any \(x \in A\), \(h(x)\) has to be nonzero for all \(x \in A\).

\[\bar{V}'(x, F) = h(x) \bar{V}(x, F)\]

for all \(x \in A\) and all \(F \in \mathcal{M}^d\). Since the matrix \(B_G\) in the proof will be a \(2 \times 3\) matrix of rank 1 for any \(x \in A\), \(h(x)\) has to be nonzero for all \(x \in A\).

\[\bar{V}'(x, F) = h(x) \bar{V}(x, F)\]

Proof of Proposition 4.6. Let \(F \in \mathcal{M}^d\) and \(EAR_w(F) \neq \emptyset\). Note that for any \(k \in \mathbb{R}^d\), \(\bar{V}_{EAR_w}(k, F)\) evaluates \(\bar{V}_{R_0}(\cdot, F): \mathbb{R}^d \to \mathbb{R}\) on the hyperplane orthogonal to \(w\) containing \(k\). Since \(\bar{V}_{EAR_w}(F) \subseteq R_0(F)\), \(k \in \bar{V}_{EAR_w}(F)\) implies that \(\bar{V}_{R_0}(k, F) = 0\). The orientation of \(\bar{V}_{R_0}\), and the facts that \(\bar{V}_{EAR_w}(F)\) is the intersection of \(R(F)\) with the supporting hyperplane for \(R(F)\) orthogonal to \(w\) and that \(R(F)\) is an upper set imply that \(\bar{V}_{R_0}(k + x, F) = \bar{V}_{EAR_w}(k, F)(x) \leq 0\) for all \(x \in w^\perp\). If \(k \notin \bar{V}_{EAR_w}(F)\), there are two possibilities:

(i) The orthogonal hyperplane containing \(k\) has an empty intersection with \(R(F)\) such that \(\bar{V}_{R_0}(k + x, F) < 0\) for all \(x \in w^\perp\).

(ii) The orthogonal hyperplane containing \(k\) has a non-empty intersection with \(R(F) \setminus R_0(Y)\) such that there is some \(x \in w^\perp\) with \(\bar{V}_{R_0}(k + x, F) > 0\).

Now let \(EAR_w(F) = \emptyset\). If \(R(F) = \emptyset\), (4.7) holds trivially. If \(R(F)\) is non-empty, \(EAR_w(F)\) is only empty if there is no supporting hyperplane for \(R(F)\) orthogonal to \(w\). Then for any \(k \in \mathbb{R}^d\), there are \(x_1, x_2 \in w^\perp\) such that \(\bar{V}_{EAR_w}(k, F)(x_1) > 0\) and \(\bar{V}_{EAR_w}(k, F)(x_2) < 0\), as depicted in the right part of Figure 3.
Proof of Proposition 4.8. The ‘only if’ part is a special case of Theorem 4.1. For the ‘if’ part, assume $V_{R_0}: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a strict selective $\mathcal{M}$-identification function for $R_0$. For any $Y \in \mathcal{Y}$ it holds that $E[V_{R_0}(0,Y)] = 0 \Leftrightarrow 0 \in R_0(Y) \Leftrightarrow \rho(A(Y)) = 0$. Then we obtain that for any $s \in \mathbb{R}$ and any $X \in \mathcal{Y}$

$$\rho(X) = s \iff \rho(X + s) = 0 \iff E[V_{R_0}(0, \eta(X+s))] = 0.$$ 

Thus $\rho$ is identifiable with a strict selective $\mathcal{M}$-identification function $V_\rho: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, $V_\rho(s,x) = V_{R_0}(0, \eta(x+s))$. \hfill $\square$

For the proof of Theorem 4.9, we need the following lemma.

**Lemma B.1.** Let $A_1, A_2 \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}^d_+)$. Then, the symmetric difference $A_1 \triangle A_2 = (A_1 \setminus A_2) \cup (A_2 \setminus A_1)$ is empty if and only if its interior, int$(A_1 \triangle A_2)$, is empty.

**Proof of Lemma B.1.** If $A_1 \triangle A_2 = \emptyset$ it is clear that int$(A_1 \triangle A_2) = \emptyset$.

Assume that there is an $x \in A_1 \triangle A_2$. Without loss of generality, we can assume that $x \in A_1 \setminus A_2$. If $x \in \text{int}(A_1 \setminus A_2)$, we are done. Hence, let $x \in (A_1 \setminus A_2) \setminus \text{int}(A_1 \setminus A_2)$ which implies that $x \in \partial(A_1 \setminus A_2)$, where $\partial(A_1 \setminus A_2)$ denotes the boundary of $A_1 \setminus A_2$. It holds that $\partial(A_1 \setminus A_2) = \partial(A_1) \setminus \partial(A_2)$, which follows since $\partial A_2 \subseteq A_2$ and $x \in A_1 \setminus A_2$. Due to the definition of the boundary, this means that for all $\varepsilon > 0$ it holds that $B_\varepsilon(x) \setminus A_1 \neq \emptyset$, where $B_\varepsilon(x)$ is the open ball with centre $x$ and radius $\varepsilon$. Assume that for all $\varepsilon > 0$ we have $B_\varepsilon(x) \cap A_2 \neq \emptyset$, then $x \in A_2 = A_2$, which is a contradiction to the assumption that $x \in \partial(A_1 \setminus A_2)$. That means there exists an $\varepsilon_0 > 0$ such that $B_{\varepsilon_0}(x) \cap A_2 = \emptyset$. Moreover, since $A_1$ is an upper set, $x + \mathbb{R}^d_+$ is a non-empty open subset of $A_1$. Furthermore, we see that $B_{\varepsilon_0}(x) \cap (x + \mathbb{R}^d_+)$ is a non-empty open subset of $A_1$ which is disjoint from $A_2$. This means that int$(A_1 \setminus A_2) \neq \emptyset$. \hfill $\square$

**Proof of Theorem 4.9.** (i) Let $k \in \mathbb{R}^d$, $A \in \mathcal{P}(\mathbb{R}^d; \mathbb{R}^d_+)$ and $F \in \mathcal{M}$. A direct calculation yields that

$$S_{R,k}(A,F) - S_{R,k}(R(F),F) = (1_{R(F) \setminus A}(k) - 1_{A \setminus R(F)}(k)) V_{R_0}(k,F) \geq 0, \quad (B.4)$$ 

where the last inequality is a direct consequence of the weak form of orientation given at (4.10). The non-negativity of $S_{R,k}$ follows from the $\mathcal{M}$-consistency, exploiting that $\delta_y \in \mathcal{M}$ for all $y \in \mathbb{R}^d$ and $S_{R,k}(A,y) \geq S_{R,k}(R(y),y) = 0$.

(ii) This is a direct consequence of the non-negativity and consistency of the scores $S_{R,k}$.

(iii) Let $F \in \mathcal{M}$, and $A^* := R(F), A \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}^d_+)$ with $A \neq A^*$. Assume that $S_{R,\pi}(A,F), S_{R,\pi}(A^*,F) < \infty$ (otherwise, there is nothing to show). Using Fubini’s Theorem, we obtain

$$S_{R,\pi}(A,F) - S_{R,\pi}(A^*,F) = \int_{A \setminus A^*} V_{R_0}(k,F) \pi(dk) - \int_{A \setminus A^*} V_{R_0}(k,F) \pi(dk).$$ 

52
Then Lemma B.1 yields that \( \text{int}(A \setminus A^*) \neq \emptyset \) or \( \text{int}(A^* \setminus A) \neq \emptyset \). If \( \text{int}(A \setminus A^*) \neq \emptyset \), the fact that \( \bar{V}(\cdot, F) \) is strictly negative on \( (A^*)^c \) and the assumption that \( \pi \) assigns positive mass to any non-empty open set in \( \mathcal{B}(\mathbb{R}^d) \) implies that
\[
\int_{A \setminus A^*} \bar{V}_{R_0}(k, F) \pi(\text{d}k) < 0,
\]
which implies that \( \bar{S}_{R,\pi}(A, F) - \bar{S}_{R,\pi}(A^*, F) > 0 \).
Assume \( \text{int}(A^* \setminus A) \neq \emptyset \). The boundary \( \partial A^* = R_0(F) = \{ k \in \mathbb{R}^d \mid \bar{V}_{R_0}(k, F) = 0 \} \) is a closed set. That means that \( \text{int}(A^* \setminus A) \setminus \partial A^* \) is open and non-empty. Moreover \( \bar{V}_{R_0}(\cdot, F) \) is strictly positive on \( \text{int}(A^* \setminus A) \setminus \partial A^* \). Hence,
\[
\int_{A^* \setminus A} \bar{V}_{R_0}(k, F) \pi(\text{d}k) \geq \int_{\text{int}(A^* \setminus A) \setminus \partial A^*} \bar{V}_{R_0}(k, F) \pi(\text{d}k) > 0,
\]
which implies that \( \bar{S}_{R,\pi}(A, F) - \bar{S}_{R,\pi}(A^*, F) > 0 \).

**Proof of Proposition 4.11.** For the first part of the Proposition, it is sufficient to show order-sensitivity for the elementary scores \( S_{R,k} \) given at (4.11). Let \( A \subseteq B \subseteq R(F) \). Then, for any \( k \in \mathbb{R}^d \),
\[
\bar{S}_{R,k}(A, F) - \bar{S}_{R,k}(B, F) = \mathbbm{1}_{B \setminus A}(k) \bar{V}_{R_0}(k, F) \geq 0,
\]
due to the orientation given at (4.10). On the other hand, for \( A \supseteq B \supseteq R(F) \) we obtain that
\[
\bar{S}_{R,k}(A, F) - \bar{S}_{R,k}(B, F) = -\mathbbm{1}_{A \setminus B}(k) \bar{V}_{R_0}(k, F) \geq 0,
\]
where the inequality follows again by (4.10).

The second part of the proposition follows along the lines of the proof of Theorem 4.9(iii). \( \square \)

**B.3. Proofs of Section 5**

**Proof of Lemma 5.2.** Assume that \( \rho \) is a positively homogeneous scalar risk measure and \( \Lambda \) is positively homogeneous of degree \( b \in \mathbb{R} \). Let \( c > 0 \) and \( Y \in \mathbb{R}^d \).
\[
R(cY) = \left\{ k \in \mathbb{R}^d \mid \rho(\Lambda(cY + k)) \leq 0 \right\} = \left\{ k \in \mathbb{R}^d \mid \rho\left(c^b \Lambda(Y + k/c)\right) \leq 0 \right\} = \left\{ k \in \mathbb{R}^d \mid c^b \rho(\Lambda(X + k/c)) \leq 0 \right\} = c \left\{ k \in \mathbb{R}^d \mid \rho(\Lambda(X + k)) \leq 0 \right\} = c R(Y).
\]

*Mutatis mutandis*, the proof works also for \( R^{ins} \). \( \square \)

**Proof of Lemma 5.3.** (i) Let \( y, k, l \in \mathbb{R}^d \). Then
\[
V_{R_0}(k + l, y - l) = V_{\rho}(0, \Lambda(k + l + y - l)) = V_{\rho}(0, \Lambda(k + y)) = V_{R_0}(k, y).
\]

53
(ii) Let $V'_{R_0}$ be another translation invariant strict $\mathcal{M}^d$-identification function for $R_0$. Using Proposition 4.4 there is a non-vanishing function $h: \mathbb{R}^d \to \mathbb{R}$ such that $V'_{R_0}(x, F) = h(x)V_{R_0}(x, F)$ for all $x \in \mathbb{R}^d$ and for all $F \in \mathcal{M}^d$. One can show that $h$ is constant along the lines of the proof of Proposition 4.7(ii) in Fissler and Ziegel (2019c).

(iii) Let $c > 0$, $k, y \in \mathbb{R}^d$. Then
\[
V_{R_0}(ck, cy) = V_{\rho}(0, \Lambda(ck + cy)) = V_{\rho}(0, c^b\Lambda(k + y)) = c^{ab}V_{\rho}(0, \Lambda(k + y)) = c^{ab}V_{R_0}(k, y).
\]

\[\blacksquare\]

Proof of Proposition 5.5. First observe that $V_{R_0}(k, y) = V_{\rho}(0, \Lambda(y + k))$ is a selective translation invariant strict $\mathcal{M}^d$-identification function for $R_0$, invoking Lemma 5.3(i) and Theorem 4.1. Then a direct computation yields that
\[
S_{R, k}(A + l, y - l) = S_{R, k - l}(A, l)
\]
for all $A \subseteq \mathbb{R}^d$, $y, l \in \mathbb{R}^d$. Therefore, (B.5) and the translation invariance of the Lebesgue measure show part (i).

For part (ii), assume that $S$ is of the form at (4.12) and is translation invariant where, invoking the discussion above, we may assume without loss of generality that the elementary scores $S$ are based on the translation invariant identification function $V_{R_0}$. For $l \in \mathbb{R}^d$ define the measure $\pi_l(A) = \pi(A + l)$ for $A \in \mathcal{B}(\mathbb{R}^d)$. Note that since $S$ is assumed to be finite this implies that the score differences are well-defined and are translation invariant as well. For $A \subseteq B$ this implies that for any $l \in \mathbb{R}$
\[
\int_{B \setminus A} V_{R_0}(z, y)\pi_l(dz) = \int_{B \setminus A} V_{R_0}(z, y)\pi(dz).
\]

Any set of the form $I = [a_1, b_1) \times \cdots \times [a_d, b_d)$, $a_i, b_i \in \mathbb{R}$, $a_i \leq b_i$, can be represented as $B \setminus A$ for some $A, B \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}^d_{+})$ with $A \subseteq B$. The system of these sets $I$, however, is a generator of $\mathcal{B}(\mathbb{R}^d)$, and we conclude that for any $l \in \mathbb{R}^d$
\[
\nu_{y, l}(D) := \int_D V_{R_0}(z, y)\pi_l(dz) = \int_D V_{R_0}(z, y)\pi(dz) =: \nu_y(D).
\]

for all $D \in \mathcal{B}(\mathbb{R}^d)$. For each $D \in \mathcal{B}(\mathbb{R}^d)$ we obtain the decomposition (depending on $y$)
\[
D = D^D_y \cup D^D_y \cup D^0_y,\quad \text{where } D^D_y = D \cap R(y)^c, \quad D^D_y = D \cap R_0(y) \text{ and } D^+ y = D \cap R(y) \setminus R_0(y).
\]

Hence, (B.6) and the strictness of the identification function $V_{R_0}$ imply that for $E = D^+ y$ or $E = D^D_y$
\[
\pi_l(E) = \int_E \frac{1}{V_{R_0}(z, y)}\nu_{y, l}(dz) = \int_E \frac{1}{V_{R_0}(z, y)}\nu_y(dz) = \pi(E).
\]

The translation equivariance of $R$ and Assumption (2) imply that (B.7) holds for all $E \in \mathcal{B}(\mathbb{R}^d)$. That means that $\pi = \gamma L^d$ for some $\gamma \geq 0$. \[\blacksquare\]
Proof of Proposition 5.6. Assume that $V_{R_0}$ is an oriented strict selective $\mathcal{M}^d$-identification function for $R_0$ which is positively homogeneous of degree $a \in \mathbb{R}$. A direct computation yields that

$$S_{R,k}(cA,cy) = c^a S_{R,k}(A,y)$$  \hspace{1cm} (B.8)

for all $A \in 2^\mathbb{R}^d$, $y \in \mathbb{R}^d$ and $c > 0$. If $\pi$ is positively homogeneous of degree $b \in \mathbb{R}$, (B.8) implies that $S_{R,\pi}$ in (5.2) is positively homogeneous of degree $a + b$.

For part (ii), assume that $S$ is of the form at (4.12) and is positively homogeneous of degree $a + b$. For $c \in \mathbb{R}$ define the measure $\pi_c(A) = \pi(cA)$ for $A \in \mathcal{B}(\mathbb{R}^d)$. Note that since $S$ is assumed to be finite, the positive homogeneity of $S$ implies that the score differences are well-defined and are also positively homogeneous of degree $a + b$. For $A \subseteq B$ a direct computation shows that for any $c > 0$

$$\int_{B \setminus A} V_{R_0}(z,y) \pi_c(dz) = c^b \int_{B \setminus A} V_{R_0}(z,y) \pi(dz).$$

With the same arguments as in the proof of Proposition 5.5, we conclude that

$$\nu_y,c(D) := \int_D V_{R_0}(z,y) \pi_c(dz) = c^b \int_D V_{R_0}(z,y) \pi(dz) =: c^b \nu_y(D)$$

for all $D \in \mathcal{B}(\mathbb{R}^d)$ and

$$\pi_c(E) = \frac{1}{V_{R_0}(z,y)} \nu_y,c(dz) = \frac{1}{V_{R_0}(z,y)} c^b \nu_y(dz) = c^b \pi(E)$$  \hspace{1cm} (B.9)

for $E = D \cap R(y)^c$ or $E = D \cap R(y) \setminus R_0(y)$. Finally, the translation equivariance of $R$ and Assumption (2) imply that (B.9) holds for all $E \in \mathcal{B}(\mathbb{R}^d)$. That means $\pi$ is positively homogeneous of degree $b$. \hfill \Box

B.4. Proofs of Section 6

Proof of Proposition 6.1. Let $F \in \mathcal{M}^d$ and $k \in \mathbb{R}^d$. Then

$$\hat{U}_2(T^{VaR_\alpha}(F),k,F) = \frac{1}{\alpha} \mathbb{E}_F[\Lambda(Y + k) \mathbb{I}\{\Lambda(Y + k) \leq -T^{VaR_\alpha}(F)(k)\}]$$

$$+ \frac{1}{\alpha} VaR_\alpha(\Lambda(Y + k))(F_{\Lambda(Y + k)}(-VaR_\alpha(\Lambda(Y + k))) - \alpha)$$

$$- ES_\alpha(\Lambda(Y + k))$$

$$\begin{cases} < 0, & \text{if } k \notin R^{ES_\alpha}(F) \\ = 0, & \text{if } k \in R_0^{ES_\alpha}(F) \\ > 0, & \text{if } k \in R^{ES_\alpha}(F) \setminus R_0^{ES_\alpha}(F), \end{cases}$$

where $F_{\Lambda(Y + k)}$ is the distribution function of $\Lambda(Y + k)$. Under Assumption (3) it holds that $\hat{U}_1(T^{VaR_\alpha}(F),k,F) = 0$. Therefore, one ends up with the second assertion. \hfill \Box

55
Proof of Theorem 6.2. (i) Let $F \in \mathcal{M}^d$, $v \in \mathbb{R}^d$, $A \in \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}^d)$ and $v^* = TV^{aR_0}(F)$, $A^* = R^{ES_0}(F)$ and $k \in \mathbb{R}^d$. If $\tilde{S}_k(v, A, F) = \infty$ there is nothing to show. So we assume that $\tilde{S}_k(v, A, F)$ is finite. Consider

$$
\tilde{S}_k(v, A, F) - \tilde{S}_k(v^*, A, F) = 1_A(k)[-\tilde{U}_2(v, k, F) + U_2(v^*, k, F)]
$$

$$
= 1_A(k)E_F[\tilde{S}_{\alpha, id}(-v(k), \Lambda(Y + k)) - \tilde{S}_{\alpha, id}(-v^*(k), \Lambda(Y + k))] \geq 0,
$$

since $S_{\alpha, id}$ is consistent for the $\alpha$-quantile. If $\tilde{S}_k(v^*, A, F) = \infty$ we are done. Otherwise, consider

$$
\tilde{S}_k(v^*, A, F) - \tilde{S}_k(v^*, A^*, F) = (1_{A^* \setminus A}(k) - 1_{A \setminus A^*}(k))\tilde{U}_2(v, k, F) \geq 0,
$$

where the inequality follows from (6.4). The non-negativity follows from the consistency and the fact that $S_k(T^{VaR_0}(\delta_y), R^{ES_0}(\delta_y), y) = 0$.

(ii) Due to part (i), the score $S_{0, \pi_2}$ is $\mathcal{M}^d$-consistent for $(T^{VaR_0}, R^{ES_0}) : \mathcal{M}^d \rightarrow \mathbb{R}^d \times \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}^d)$. Since $S_{\alpha, g_\pi}$ is a consistent selective scoring function for the $\alpha$-quantile, the assertion follows invoking Fubini’s Theorem.

(iii) Let $F \in \mathcal{M}_0^d$, $v \in \mathcal{C}(\mathbb{R}^d; \mathbb{R})$, $A \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}^d)$ and $v^* = TV^{aR_0}(F)$, $A^* = R^{ES_0}(F)$. If $v \neq v^*$ then $K = \{k \in \mathbb{R}^d | v(k) \neq v^*(k)\} \neq \emptyset$ is open. If $\tilde{S}_{\pi_1, \pi_2}(v, A, F) = \infty$ there is nothing to show. Otherwise

$$
E_F[S_{\pi_1, \pi_2}(v, A, Y) - S_{\pi_1, \pi_2}(v^*, A, Y)]
$$

\[
\geq \int_K E_F[S_{\alpha, g_\pi}(-v(k), \Lambda(Y + k)) - S_{\alpha, g_\pi}(-v^*(k), \Lambda(Y + k))] \pi_1(dk)
\]

\[
+ \frac{1}{\alpha} \int_{A \cap K} E_F[S_{\alpha, id}(-v(k), \Lambda(Y + k)) - S_{\alpha, id}(-v^*(k), \Lambda(Y + k))] \pi_2(dk) > 0,
\]

where the first integral is strictly positive and the second one is non-negative (and strictly positive if and only if $\pi_2(A \cap K) > 0$).

If $A \neq A^*$, then $E_F[S_{\pi_1, \pi_2}(v^*, A, Y) - S_{\pi_1, \pi_2}(v^*, A^*, Y)] > 0$, which follows with similar arguments as in the proof of Theorem 4.9(iii).

\[\square\]

C. Further simulation results

In Figures 4 and 5, depict 4 more experiments as described in Subsection 7.2.
Figure 4: See description of Figure 2.
Figure 5: See description of Figure 2.
References

Acerbi, C., & Szekely, B. (2014). Backtesting Expected Shortfall. *Risk Magazine*. Retrieved from https://www.msci.com/documents/10199/22aa9922-f874-4060-b57a-0f0e627a489b

Acerbi, C., & Szekely, B. (2017). General properties of backtestable statistics. *Preprint*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2905109

Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2016). Measuring Systemic Risk. *Rev. Financial Stud.*, 30(1), 2–47. Retrieved from https://doi.org/10.1093/rfs/hhw088

Adrian, T., & Brunnermeier, M. K. (2016). CoVaR. *Am. Econ. Rev.*, 106(7), 1705–41. Retrieved from http://www.aeaweb.org/articles?id=10.1257/aer.20120555

Amini, H., Filipovic, D., & Minca, A. (2015). Systemic Risk and Central Clearing Counterparty Design. *Swiss Finance Institute Research Paper No. 13–34*. Retrieved from http://dx.doi.org/10.2139/ssrn.2275376

Armenti, Y., Crépey, S., Drapeau, S., & Papapantoleon, A. (2018). Multivariate Shortfall Risk Allocation and Systemic Risk. *SIAM J. Financial Math.*, 9(1), 90–126. Retrieved from https://doi.org/10.1137/16M1087357

Aronow, P. M., & Lee, D. K. K. (2018). A Curious Result on Breaking Ties among Sample Medians. *Preprint*. Retrieved from https://arxiv.org/abs/1807.03462

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Math. Finance*, 9, 203–228. Retrieved from https://doi.org/10.1111/1467-9965.00068

Artzner, P., Delbaen, F., & Koch-Medina, P. (2009). Risk Measures and Efficient Use of Capital. *Astin Bull.*, 39, 101–116. Retrieved from https://doi.org/10.2143/AST.39.1.2038058

Askanazi, R., Diebold, F. X., Schorheide, F., & Shin, M. (2018). On the Comparison of Interval Forecasts. *J. Time Series Anal.*, 39(6), 953–956. Retrieved from https://dx.doi.org/10.1111/jtsa.12426

Azzimonti, D., Bect, J., Chevalier, C., & Ginsbourger, D. (2016). Quantifying Uncertainties on Excursion Sets Under a Gaussian Random Field Prior. *SIAM/ASA J. Uncertain. Quantif.*, 4(1), 850–874. Retrieved from https://doi.org/10.1137/141000749

Azzimonti, D., Ginsbourger, D., Chevalier, C., Bect, J., & Richet, Y. (2018). Adaptive Design of Experiments for Conservative Estimation of Excursion Sets. *Preprint*. Retrieved from https://arxiv.org/abs/1611.07256

Bank for International Settlements. (2013). *Consultative Document: Fundamental review of the trading book: A revised market risk framework*. Retrieved from https://www.bis.org/bcbs/publ/d305.htm

Bank for International Settlements. (2014). *Consultative Document: Fundamental review of the trading book: Outstanding issues*. Retrieved from https://www.bis.org/publ/bcbs265.pdf

Bellini, F., & Bignozzi, V. (2015). On elicitable risk measures. *Quant. Finance, 15*(5), 725–733. Retrieved from https://doi.org/10.1080/14697688.2014.946955

Bellini, F., & Di Bernardino, E. (2015). Risk Management with Expectiles. *Eur. J. Finance*. Retrieved from http://dx.doi.org/10.2139/ssrn.2475106

Biagini, F., Fouque, J.-P., Frittelli, M., & Meyer-Brandis, T. (2019). A unified approach to systemic risk measures via acceptance sets. *Math. Finance*, 29(1), 329–367. Retrieved from https://doi.org/10.1111/mafi.12170
Bignozzi, W., Burzoni, M., & Munari, C. (2018). Risk Measures Based on Benchmark Loss Distributions. *Swiss Finance Institute Research Paper No. 18–48*. Retrieved from https://ssrn.com/abstract=3088423

Bolin, D., & Lindgren, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 77(1), 85–106. Retrieved from https://doi.org/10.1111/rssb.12055

Brehmer, J., & Strokorb, K. (2019). Why scoring functions cannot assess tail properties. *Preprint*. Retrieved from https://arxiv.org/abs/1905.04233

Brier, G. W. (1950). Verification of Forecasts Expressed in Terms of Probability. *Mon. Weather Rev.*, 78(1), 1–3. Retrieved from https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Chen, C., Iyengar, G., & Moallemi, C. C. (2013). An Axiomatic Approach to Systemic Risk. *Manag. Sci.*, 59(6), 1373–1388. Retrieved from https://doi.org/10.1287/mnsc.1120.1631

Chevalier, C., Ginsbourger, D., Bect, J., & Molchanov, I. (2013). Estimating and quantifying uncertainties on level sets using the Vorob’ev expectation and deviation with Gaussian process models. In D. Uciński, A. Atkinson, & C. Patan (Eds.), *mODa 10 – Advances in Model-Oriented Design and Analysis* (Vol. 1). Springer International Publishing. Retrieved from https://hal.archives-ouvertes.fr/hal-00731783/document

Cont, R., Deguest, R., & Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quant. Finance*, 10, 593–606. Retrieved from https://doi.org/10.1080/14697681003685597

Daouia, A., & Paindaveine, D. (2019). From Halfspace M-depth to Multiple-output Expectile Regression. *Preprint*. Retrieved from https://arxiv.org/abs/1905.12718

Davis, M. H. A. (2016). Verification of internal risk measure estimates. *Stat. Risk Model.*, 33(3–4), 67–93. Retrieved from https://doi.org/10.1515/strm-2015-0007

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(3), 505–562.

Dearborn, K., & Frongillo, R. (2019). On the Indirect Elicitability of the Mode and Modal Interval. *Ann. Inst. Statist. Math.*. Retrieved from https://doi.org/10.1007/s10657-019-00719-1

Delbaen, F., Bellini, F., Bignozzi, V., & Ziegel, J. F. (2016). Risk measures with the CxLS property. *Finance Stoch.*, 20(2), 433–453. Retrieved from https://doi.org/10.1007/s00780-015-0279-6

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *J. Bus. Econom. Statist.,* 13, 253–263. Retrieved from https://doi.org/10.1080/07350015.1995.10524599

Dorninger, M., Gilleland, E., Casati, B., Mittermaier, M. P., Ebert, E. E., Brown, B. G., & Wilson, L. J. (2018). The Setup of the MesoVICT Project. *Bull. Am. Meteorol. Soc.*, 99(9), 1887–1906. Retrieved from https://doi.org/10.1175/BAMS-D-17-0164.1

Ehm, W., Gneiting, T., Jordan, A., & Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(3), 505–562. Retrieved from https://doi.org/10.1111/rssb.12154
Eisenberg, L., & Noe, T. (2001). Systemic Risk in Financial Networks. *Manag. Sci.*, 47(2), 236–249. Retrieved from https://doi.org/10.1287/mnsc.47.2.236.9835

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., & Beleraj, A. (2014). An Academic Response to Basel 3.5. *Risks*, 2(1), 25–48. Retrieved from https://doi.org/10.3390/risks2010025

Emmer, S., Kratz, M., & Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard risk measures. *Journal of Risk (The)*, 8, 31–60. Retrieved from https://doi.org/10.1137/16M1066087

Engelberg, J., Manski, C. F., & Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *J. Bus. Econ. Stat.*, 27, 30–41. Retrieved from https://www.jstor.org/stable/27639017

Feinstein, Z., Rudloff, B., & Weber, S. (2017). Measures of systemic risk. *SIAM J. Financial Math.*, 8, 672–708. Retrieved from https://doi.org/10.1137/16M1066087

Fissler, T. (2017). *On Higher Order Elicitability and Some Limit Theorems on the Poisson and Wiener Space* (Doctoral dissertation, University of Bern). Retrieved from http://biblio.unibe.ch/download/eldiss/17fissler_t.pdf

Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband’s principle. *Ann. Statist.*, 44(4), 1680–1707. Retrieved from https://doi.org/10.1214/16-AOS1439

Fissler, T., & Ziegel, J. F. (2019a). Elicitability of Range Value at Risk. *Preprint*. Retrieved from https://arxiv.org/abs/1902.04489v2

Fissler, T., & Ziegel, J. F. (2019b). Erratum: Higher Order Elicitability and Osband’s Principle. *Preprint*. Retrieved from https://arxiv.org/abs/1901.08826v1

Fissler, T., & Ziegel, J. F. (2019c). Order-sensitivity and equivariance of scoring functions. *Electron. J. Stat.*, 13(1), 1166–1211. Retrieved from https://doi.org/10.1214/19-EJS1552

Fissler, T., Ziegel, J. F., & Gneiting, T. (2016). Expected shortfall is jointly elicitable with value-at-risk: implications for backtesting. *Risk Magazine*, 58–61.

Föllmer, H., & Schied, A. (2002). Convex measures of risk and trading constraints. *Finance Stoch.*, 6(4), 429–447. Retrieved from https://doi.org/10.1007/s007800200072

Föllmer, H., & Schied, A. (2004). *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter.

Föllmer, H., & Weber, S. (2015). The Axiomatic Approach to Risk Measures for Capital Determination. *Annual Review of Financial Economics*, 7(1), 301–337. Retrieved from https://doi.org/10.1146/annurev-finance-111914-040203

Frongillo, R., & Kash, I. (2015a). On Elicitation Complexity. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems 28* (pp. 3258–3266). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/5832-on-elicitation-complexity.pdf

Frongillo, R., & Kash, I. (2015b). Vector-Valued Property Elicitation. *JMLR Workshop Conf. Proc.*, 40, 1–18. Retrieved from http://proceedings.mlr.press/v40/Frongillo15.html

Frongillo, R., & Kash, I. A. (2018). Elicitation Complexity of Statistical Properties. *Preprint*. Retrieved from https://arxiv.org/abs/1506.07212v2

Gao, T., Chen, T., Chenthamarakshan, V., & Witbrock, M. (2019). A Sequential Set Generation Method for Predicting Set-Valued Outputs. *Preprint*. Retrieved from https://arxiv.org/abs/1903.05153
Gneiting, T. (2011a). Making and Evaluating Point Forecasts. *J. Amer. Statist. Assoc.*, 106, 746–762. Retrieved from https://www.jstor.org/stable/41416407

Gneiting, T. (2011b). Quantiles as optimal point forecasts. *Int. J. Forecast.*, 27(2), 197–207. Retrieved from https://doi.org/10.1016/j.ijforecast.2009.12.015

Gneiting, T. (2017). When is the mode functional the bayes classifier? *Stat*, 6(1), 204–206. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/ston.12117

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 69, 243–268. Retrieved from https://doi.org/10.1111/j.1467-9868.2007.00587.x

Gneiting, T., & Raftery, A. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Amer. Statist. Assoc.*, 102, 359–378. Retrieved from https://doi.org/10.1198/016214506000001437

Gneiting, T., & Ranjan, R. (2013). Combining predictive distributions. *Electron. J. Statist.*, 1747–1782. Retrieved from https://doi.org/10.1214/13-EJS823

Good, I. J. (1952). Rational decisions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 14(1), 107–114. Retrieved from http://www.jstor.org/stable/2984087

Hamel, A. H., & Kostner, D. (2018). Cone distribution functions and quantiles for multivariate random variables. *J. Multivariate Anal.*, 167, 97–113. Retrieved from https://doi.org/10.1016/j.jmva.2018.04.004

Hampel, F. R. (1971, 12). A General Qualitative Definition of Robustness. *Ann. Math. Statist.*, 42(6), 1887–1896. Retrieved from https://doi.org/10.1214/aoms/1177693054

Heinrich, C. (2014). The mode functional is not elicitable. *Biometrika*, 101(1), 245–251. Retrieved from https://doi.org/10.1093/biomet/ast048

Hoffmann, H., Meyer-Brandis, T., & Svindland, G. (2016). Risk-consistent conditional systemic risk measures. *Stochastic Process. Appl.*, 126(7), 2014–2037. Retrieved from https://doi.org/10.1016/j.spa.2016.01.002

Holzmann, H., & Eulert, M. (2014). The role of the information set for forecasting – with applications to risk management. *Ann. Appl. Stat.*, 8, 79–83. Retrieved from https://doi.org/10.1214/13-AOAS709

Huber, P. J., & Ronchetti, E. M. (2009). *Robust statistics* (Second ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Jordan, A., Mühlemann, A., & Ziegel, J. F. (2019). Optimal solutions to the isotonic regression problem. *Preprint*. Retrieved from https://arxiv.org/abs/1904.04761

Jouni, E., Schachermayer, W., & Touzi, N. (2006). Law invariant risk measures have the Fatou property. In S. Kusuoka & A. Yamazaki (Eds.), *Adv. math. econ.* (Vol. 9, pp. 49–72). Tokyo: Springer. Retrieved from https://doi.org/10.1007/4-431-34342-3_4

Koenker, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.

Koenker, R., & Basset, G. (1978). Regression Quantiles. *Econometrica*, 46(1), 33–50. Retrieved from https://doi.org/10.2307/1913643

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *J. Econométrics*, 128(1), 137–164. Retrieved from http://dx.doi.org/10.1016/j.jeconom.2004.08.010

Krätschmer, V., Schied, A., & Zähle, H. (2012). Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *J. Multivariate Anal.*, 103, 35–47. Retrieved from https://doi.org/10.1016/j.jmva.2011.06.005
Kratzschmer, V., Schied, A., & Zähle, H. (2014). Comparative and qualitative robustness for law-invariant risk measures. *Finance Stoch.*, 18, 271–295. Retrieved from https://doi.org/10.1007/s00780-013-0225-4

Kromer, E., Overbeck, L., & Zilch, K. (2016, Oct 01). Systemic risk measures on general measurable spaces. *Math. Methods Oper. Res.*, 84(2), 323–357. Retrieved from https://doi.org/10.1007/s00186-016-0545-1

Lambert, N. (2013). Elicitation and Evaluation of Statistical Functionals. *Preprint*. Retrieved from http://web.stanford.edu/~nlambert/papers/elicitation.pdf

Lambert, N., Pennock, D. M., & Shoham, Y. (2008). Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce* (pp. 129–138). Chicago, Il, USA: ACM. Retrieved from https://dl.acm.org/citation.cfm?id=1386813

Matheson, J. E., & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Manag. Sci.*, 22(10), 1087–1096. Retrieved from http://www.jstor.org/stable/2629907

Mayo-Wilson, C., & Wheeler, G. (2016). Scoring Imprecise Credences: A Mildly Immodest Proposal. *Philos. Phenomenol. Res.*, 93(1), 55–78. Retrieved from https://doi.org/10.1111/phpr.12256

Mizera, I. (2010). Qualitative robustness and weak continuity: the extreme unction? In J. Antoch, M. Hušková, & P. Sen (Eds.), *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková* (Vol. Volume 7, pp. 169–181). Beachwood, Ohio, USA: Institute of Mathematical Statistics. Retrieved from https://doi.org/10.1214/10-IMSCOLL717

Molchanov, I. (2017). *Theory of random sets* (2nd ed.). London: Springer-Verlag London.

Mösching, A., & Dümbgen, L. (2019). Monotone least squares and isotonic quantiles. *Preprint*. Retrieved from https://arxiv.org/abs/1901.02398

Murphy, A. H., & Daan, H. (1985). Forecast Evaluation. In A. H. Murphy & R. W. Katz (Eds.), *Probability, Statistics and Decision Making in the Atmospheric Sciences* (pp. 379–437). Westview Press, Boulder, Colorado.

Nau, R. F. (1985). Should Scoring Rules Be ‘Effective’? *Manag. Sci.*, 31(5), 527–535. Retrieved from https://doi.org/10.1287/mnsc.31.5.527

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. Engle & D. McFadden (Eds.), *Handbook of Econometrics* (Vol. 4, pp. 2111–2245). Elsevier. Retrieved from https://doi.org/10.1016/S1573-4412(05)80005-4

Newey, W. K., & Powell, J. L. (1987). Asymmetric Least Squares Estimation and Testing. *Econometrica*, 55, 819–847. Retrieved from https://www.jstor.org/stable/1911031

Nolde, N., & Ziegel, J. F. (2017, 12). Elicitability and backtesting: Perspectives for banking regulation. *Ann. Appl. Stat.*, 11(4), 1833–1874. Retrieved from https://doi.org/10.1214/17-AOAS1041

Osband, K. H. (1985). *Providing Incentives for Better Cost Forecasting* (Unpublished doctoral dissertation). University of California, Berkeley.

Patton, A. J. (2011). Data-based ranking of realised volatility estimators. *J. Econometrics*, 161(2), 284–303. Retrieved from https://doi.org/10.1016/j.jeconom.2010.12.010

Patton, A. J. (2019). Comparing possibly misspecified forecasts. *J. Bus. Econom. Statist.*, 0(0), 1–23. Retrieved from https://doi.org/10.1080/07350015.2019.1585256
Rogers, L. C. G., & Veraart, L. A. M. (2013). Failure and Rescue in an Interbank Network. *Manag. Sci.*, 59(4), 882–898. Retrieved from https://doi.org/10.1287/mnsc.1120.1569

Savage, L. J. (1971). Elicitation of Personal Probabilities and Expectations. *J. Amer. Statist. Assoc.*, 66, 783–801. Retrieved from https://www.jstor.org/stable/2284229

Schlag, K. H., & van der Weele, J. J. (2015). A method to elicit beliefs as most likely intervals. *Judgm. Decis. Mak.*, 10(5). Retrieved from https://hdl.handle.net/11245/1.497336

Seidenfeld, T., Schervish, M. J., & Kadane, J. B. (2012). Forecasting with imprecise probabilities. *Internat. J. Approx. Reason.*, 53(8), 1248–1261. Retrieved from https://doi.org/10.1016/j.ijar.2012.06.018

Steinwart, I., Pasin, C., Williamson, R., & Zhang, S. (2014). Elicitation and Identification of Properties. *JMLR Workshop Conf. Proc.*, 35, 1–45. Retrieved from http://proceedings.mlr.press/v35/steinwart14.html

Strähl, C., & Ziegel, J. F. (2017). Cross-calibration of probabilistic forecasts. *Electron. J. Statist.*, 11(1), 608–639. Retrieved from https://doi.org/10.1214/17-EJS1244

Tsyplakov, A. (2014). *Theoretical guidelines for a partially informed forecast examiner* (Tech. Rep.). Novosibirsk: 60–380. Tsyplakov, A. (2014) Theoretical guidelines for a partially informed forecast examiner. Working Paper. Department of Economics, Novosibirsk State University. Retrieved from https://mpra.ub.uni-muenchen.de/67333/1/MPRA_paper_67333.pdf

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press.

Wang, R., & Wei, Y. (2018). Risk functionals with convex level sets. *Preprint*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3292661

Weber, S. (2006). Distribution-Invariant Risk Measures, Information, and Dynamic Consistency. *Math. Finance*, 16, 419–441. Retrieved from https://doi.org/10.1111/j.1467-9965.2006.00277.x

White, H. (2001). *Asymptotic Theory for Econometricians*. San Diego: Academic Press.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R., & Smola, A. (2017). Deep sets. *Preprint*. Retrieved from https://arxiv.org/abs/1703.06114

Ziegel, J. F. (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(3), 505–562.

Ziegel, J. F., Krüger, F., Jordan, A., & Fasciati, F. (2019, 02). Robust Forecast Evaluation of Expected Shortfall. *J. Financ. Economet.*. Retrieved from https://dx.doi.org/10.1093/jjfinec/nby035