

Elicitability and Identifiability of Systemic Risk Measures

Fissler, Tobias; Hlavinova, Jana; Rudloff, Birgit

Published in:
Finance and Stochastics

DOI:
[10.1007/s00780-020-00446-z](https://doi.org/10.1007/s00780-020-00446-z)

Published: 01/01/2021

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

Document License:
CC BY

[Link to publication](#)

Citation for published version (APA):
Fissler, T., Hlavinova, J., & Rudloff, B. (2021). Elicitability and Identifiability of Systemic Risk Measures. *Finance and Stochastics*, 25(1), 133 - 165. <https://doi.org/10.1007/s00780-020-00446-z>



Elicibility and identifiability of set-valued measures of systemic risk

Tobias Fissler¹ · Jana Hlavinová¹ · Birgit Rudloff¹

Received: 18 October 2019 / Accepted: 22 October 2020
© The Author(s) 2020

Abstract Identification and scoring functions are statistical tools to assess the calibration of risk measure estimates and to compare their performance with other estimates, e.g. in backtesting. A risk measure is called identifiable (elicitable) if it admits a strict identification function (strictly consistent scoring function). We consider measures of systemic risk introduced in Feinstein et al. (SIAM J. Financial Math. 8:672–708, 2017). Since these are set-valued, we work within the theoretical framework of Fissler et al. (preprint, available online at [arXiv:1910.07912v2](https://arxiv.org/abs/1910.07912v2), 2020) for forecast evaluation of set-valued functionals. We construct oriented selective identification functions, which induce a mixture representation of (strictly) consistent scoring functions. Their applicability is demonstrated with a comprehensive simulation study.

Keywords Consistent scoring functions · Diebold–Mariano tests · Forecast evaluation · M -estimation · Murphy diagrams

Mathematics Subject Classification (2010) 62F07 · 62F10 · 91B30 · 91G70

JEL Classification C52 · G32

T. Fissler received financial support via his Chapman Fellowship from Imperial College London for parts of this work. B. Rudloff acknowledges support from the OeNB anniversary fund, project number 17793.

✉ J. Hlavinová
jana.hlavinova@wu.ac.at

T. Fissler
tobias.fissler@wu.ac.at

B. Rudloff
birgit.rudloff@wu.ac.at

¹ Institute for Statistics and Mathematics, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020 Vienna, Austria

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 axiomatic approach to the quantitative assessment of risk was pioneered by Artzner et al. [6] and has since then been discussed from various angles in many further works; see Föllmer and Schied [33, Chap. 4] for a comprehensive overview.

The financial crisis 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 by Eisenberg and Noe [20]. The focus of this work, however, lies on modelling the financial system rather than measuring its systemic risk. Since then, financial mathematicians have developed a rich strand of literature, encompassing different approaches and emphasising various aspects of systemic risk. The model of [20] has been generalised in different ways, for instance by considering illiquidity (Rogers and Veraart [57]) or central clearing (Amini et al. [4]). 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 et al. [2], Adrian and Brunnermeier [3]). Recognising the drawbacks of treating the economy as a portfolio, Chen et al. [13] introduce an axiomatic approach to measuring systemic risk, further extended by Kromer et al. [48] and Hoffmann et al. [40]. The axiomatic approach of [13] 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, ρ is a scalar risk measure and $\Lambda: \mathbb{R}^d \rightarrow \mathbb{R}$ an increasing aggregation function. However, this approach 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 of capital allocations on the system.

As an alternative, so-called sensitive systemic risk measures have been introduced by Feinstein et al. [23]; see also Biagini et al. [11] and Armenti et al. [5] for related approaches. Here, one first adds capital to the d financial institutions and then applies an aggregation function, resulting in systemic risk measures of the form

$$R(Y) = \{k \in \mathbb{R}^d : \rho(\Lambda(Y + k)) \leq 0\}. \quad (1.1)$$

Thus one takes into account how the regulation changes the system itself as it aggregates the system $Y + k$ after regulation. In this paper, we mainly focus on this type of systemic risk measures; see Sect. 2.1 for more details. Above, $R(Y)$ 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 ρ after being aggregated via Λ . 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 crisis, 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 et al. [21], Emmer et al. [22] for detailed academic discussions and [8] for a regulatory perspective in banking. Besides differences in axiomatic properties such as *coherence* [6] and *convexity* [32] 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 [39] (see also Cont et al. [14] and Krätschmer et al. [47]) and *elicitability*.

The term elicibility is due to Osband [55, Chap. 2] and Lambert et al. [50]. Using the terminology of mathematical statistics, a law-invariant risk measure ρ mapping to $\mathbb{R}^* := (-\infty, \infty]$ is *elicitable* on some class \mathcal{M} of distributions if it admits an M -estimator in the sense of Huber and Ronchetti [42, Chap. 3], i.e., if there is a loss or scoring function $S: \mathbb{R}^* \times \mathbb{R} \rightarrow \mathbb{R}^*$ such that

$$\int S(\rho(F), y) dF(y) < \int S(x, y) dF(y) \quad (1.2)$$

for all $F \in \mathcal{M}$ and all $x \in \mathbb{R}^*$, $x \neq \rho(F)$. Any scoring function S satisfying (1.2) is called *strictly \mathcal{M} -consistent* for $\rho: \mathcal{M} \rightarrow \mathbb{R}^*$. Besides their usage in M -estimation and regression [46, 45, 53], the fact that strict consistency encourages truthful forecasting opens the way to meaningful forecast comparison, see Gneiting [35], which is closely related to comparative backtests in finance; see Fissler et al. [31] or Nolde and Ziegel [54]. In a nutshell, let $(Y_t)_{t=1, \dots, N}$ be the profits and losses at the time points $t = 1, \dots, N$ and let X_t be a one-step ahead point forecast for Y_t , based on the information \mathfrak{F}_{t-1} available to the forecaster at time point $t - 1$. That is, X_t is an \mathfrak{F}_{t-1} -measurable random variable. The aim of X_t is to correctly specify the conditional risk measure ρ of Y_t given \mathfrak{F}_{t-1} , that is, the minimal capital requirement one needs to add to Y_t to make $Y_t + X_t$ acceptable under ρ , given the information \mathfrak{F}_{t-1} . The simplest example is when ρ is the negative expectation. Then the ideal forecast can be written as $X_t^* = -\mathbb{E}[Y_t | \mathfrak{F}_{t-1}]$. For a general law-invariant risk measure ρ , the ideal forecast for Y_t can be expressed as $X_t^* = \rho(F_{Y_t | \mathfrak{F}_{t-1}})$, where $F_{Y_t | \mathfrak{F}_{t-1}}$ is the conditional distribution of Y_t given \mathfrak{F}_{t-1} . Note that X_t might be misspecified in the sense that $X_t \neq X_t^* = \rho(F_{Y_t | \mathfrak{F}_{t-1}})$, e.g. due to possible estimation errors when coming up with an estimate of $F_{Y_t | \mathfrak{F}_{t-1}}$, due to a misuse of the information \mathfrak{F}_{t-1} , or due to calculation errors when applying the risk measure to the conditional distribution. We refer to Sect. 5 for some specific examples. If there is another forecaster with his/her own information sets \mathfrak{G}_{t-1} for time $t - 1$ who issues alternative \mathfrak{G}_{t-1} -measurable point forecasts Z_t , $t = 1, \dots, N$, for Y_t , then the predictive performance of $(X_t)_{t=1, \dots, N}$ is deemed better than the performance of $(Z_t)_{t=1, \dots, N}$ with respect to the scoring function S if $\frac{1}{N} \sum_{t=1}^N S(X_t, Y_t) < \frac{1}{N} \sum_{t=1}^N S(Z_t, Y_t)$. **In the rest of this article**, we conveniently call a point forecast for Y_t aiming at correctly specifying $\rho(F_{Y_t | \mathfrak{F}_{t-1}})$ for some σ -algebra \mathfrak{F}_{t-1} a *forecast for ρ* .

Ziegel [63] showed that expectiles are basically the only elicitable and coherent risk measures. In line with this, the prominent risk measure value at risk at level $\alpha \in (0, 1)$ (VaR_α), which corresponds to the negative of the lower α -quantile (see equation (4.1)), turns out to be elicitable, subject to mild conditions, but not coherent.

On the other hand, expected shortfall at level $\alpha \in (0, 1)$ (ES_α), a tail expectation, is coherent, but fails to be elicitable; see Gneiting [35] and Weber [62]. Interestingly, Fissler and Ziegel [27] showed that the pair $(\text{VaR}_\alpha, \text{ES}_\alpha)$ is elicitable despite ES 's failure to have a strictly consistent scoring function on its own; see also Acerbi and Szekely [1] for a slightly weaker result, and Fissler and Ziegel [30] for a similar result for range value at risk.

Closely related to the notion of elicibility 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 checks for calibration. A law-invariant risk measure $\rho: \mathcal{M} \rightarrow \mathbb{R}^*$ is *identifiable* on \mathcal{M} if it admits a Z -estimator, i.e., if there is some function $V: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\int V(x, y) dF(y) = 0 \quad \iff \quad x = \rho(F) \quad (1.3)$$

for all $F \in \mathcal{M}$ and all $x \in \mathbb{R}$. Any function V satisfying (1.3) is called a *strict \mathcal{M} -identification function* for ρ . Here we generalise the definition of identifiability in the literature to risk-measures possibly assuming the value ∞ . Clearly, if $\rho(F) = \infty$, (1.3) means that there is no real number x such that $\int V(x, y) dF(y) = 0$. Steinwart et al. [58] showed that under appropriate regularity conditions, the identifiability of a real-valued risk measure is equivalent to its elicibility. Coherently, VaR_α is identifiable under mild regularity conditions, using a simple coverage check, whereas ES_α fails to have a strict identification function. For a discussion of identifiability and calibration in the context of evaluating risk measures, we refer the reader to Davis [15] and Nolde and Ziegel [54].

1.3 Novel contributions and structure of the paper

The aim of this paper is to establish elicibility and identifiability results for systemic risk measures of the form (1.1) and for derived quantities thereof. This facilitates backtests of these risk measures and renders regression frameworks possible. Since these risk measures are set-valued, we use the terminological distinction between *selective* and *exhaustive* reports introduced in Fissler et al. [25] along with the corresponding notions of elicibility and identifiability. In a nutshell and translated to the setting of systemic risk measures of the form (1.1), a selective forecast specifies a single capital allocation that makes the system acceptable. On the other hand, exhaustive forecasts are more ambitious, aiming at reporting all adequate capital allocations simultaneously in the form of a set. Consequently, exhaustive scoring or identification functions take sets as their first argument, whereas their selective counterparts work with points as inputs. The corresponding definitions along with basic properties and assumptions on systemic risk measures defined in (1.1) and derived quantities such as *efficient cash-invariant allocation rules* (EARs), see Feinstein et al. [23], are gathered in Sect. 2.

Section 3 contains our main original contributions, most notably Theorem 3.1 asserting the existence of oriented selective identification functions for the functional $R_0(Y) = \{k \in \mathbb{R}^d: \rho(\Lambda(Y + k)) = 0\}$, and Theorem 3.10, which uses these identification functions to construct strictly consistent exhaustive scoring functions for R

from (1.1). Interestingly, these scoring functions arise as an integral construction of elementary scores, exploiting the orientation of the identification function. This can be considered a higher-dimensional analogue to the *mixture representation* of scoring functions for one-dimensional forecasts established in the seminal paper by Ehm et al. [19]. Similarly, this gives rise to the diagnostic tool of *Murphy diagrams* facilitating the assessment of forecast dominance; see Sect. 3.2.5. Thanks again to the orientation of the identification functions, we derive order-sensitivity results of these consistent scoring functions (Proposition 3.13). Concerning EARs as mentioned above, Proposition 3.6 establishes strict selective identification functions for EARs, interestingly mapping to a function space. On top of these main original contributions, we exploit the mutual exclusivity result of Fissler et al. [25, Theorem 3.7] to conclude that systemic risk measures of the form (1.1) are generally not selectively elicitable (Theorem 3.16).

The elicibility results on R rely on the identifiability of the underlying scalar risk measure ρ . This spells doom for the elicibility of systemic risk measures induced by ES as a scalar risk measure. Section 4 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 together with a VaR-related quantity leads to selective identifiability and exhaustive elicibility results (Proposition 4.1 and Theorem 4.3).

The practical applicability of our results is demonstrated in terms of a simulation study, which is the content of Sect. 5. 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, utilising a traffic-light approach suggested in Fissler et al. [31]. We close the paper with a brief discussion.

In an online preprint version [26], we gather results on positively homogeneous and translation-invariant scoring functions, additional simulation results as well as results concerning risk measures insensitive with respect to capital allocations.

2 Notation and terminology

2.1 Measures of systemic risk

Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be an atomless probability space, and for some integer $d \geq 1$, let $\mathcal{Y}^d \subseteq L^0(\Omega; \mathbb{R}^d)$ be a collection of d -dimensional random vectors which is closed under translation, meaning that $Y \in \mathcal{Y}^d$ and $k \in \mathbb{R}^d$ implies that $Y + k \in \mathcal{Y}^d$. From a risk management perspective, the random vector $Y = (Y_1, \dots, Y_d)^\top \in \mathcal{Y}^d$ 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 distributions of elements of \mathcal{Y}^d . Let $\Lambda: \mathbb{R}^d \rightarrow \mathbb{R}$ be a measurable *aggregation function*, meaning that it is non-constant and increasing with respect to the componentwise order. An aggregation function is typically, but not necessarily, assumed to be continuous or even concave.

Let $\mathcal{Y} \subseteq L^0(\Omega; \mathbb{R})$ be a collection of random variables which is closed under translation and contains $\{\Lambda(Y) : Y \in \mathcal{Y}^d\}$. Similarly to \mathcal{M}^d , let \mathcal{M} be the class of probability distributions of elements of \mathcal{Y} .

We consider a scalar monetary law-invariant risk measure $\rho : \mathcal{Y} \cup \mathbb{R} \rightarrow \mathbb{R}^*$; see Artzner et al. [6]. That is, for all $X, Z \in \mathcal{Y} \cup \mathbb{R}$ and $m \in \mathbb{R}$, it holds that $\rho(X + m) = \rho(X) - m$ (cash-invariance) and $\rho(X) \leq \rho(Z)$ if $X \geq Z$ \mathbb{P} -a.s. (monotonicity). Moreover, we assume that $\rho(0) \in \mathbb{R}$ so that cash-invariance induces a unique mapping $\rho : \mathbb{R} \rightarrow \mathbb{R}$. Exploiting law-invariance, we identify $\rho : \mathcal{Y} \cup \mathbb{R} \rightarrow \mathbb{R}^*$ with its induced risk functional $\hat{\rho} : \mathcal{M} \cup \{\delta_x : x \in \mathbb{R}\} \rightarrow \mathbb{R}^*$, where $\hat{\rho}(F) = \rho(X)$ for some $X \sim F$, and simply write ρ in both cases.

We present the two most natural law-invariant set-valued measures of systemic risk that are based on ρ and Λ , namely

$$R : \mathcal{Y}^d \rightarrow 2^{\mathbb{R}^d}, \quad Y \mapsto R(Y) = \{k \in \mathbb{R}^d : \rho(\Lambda(Y + k)) \leq 0\}, \tag{2.1}$$

$$R^{\text{ins}} : \mathcal{Y}^d \rightarrow 2^{\mathbb{R}^d}, \quad Y \mapsto R^{\text{ins}}(Y) = \{k \in \mathbb{R}^d : \rho(\Lambda(Y) + \bar{k}) \leq 0\}. \tag{2.2}$$

In (2.2) and later, we use the shorthand $\bar{k} := \sum_{i=1}^d k_i$ for $k = (k_1, \dots, k_d)^\top \in \mathbb{R}^d$. Note the difference between R and R^{ins} . The risk measure R takes an 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 ρ . On the other hand, R^{ins} takes an ex post perspective on quantifying the risk of the system Y : It first considers the current aggregated system $\Lambda(Y)$ and then specifies the *total* capital requirement \bar{k} one needs to add to make the aggregated system acceptable, which amounts to specifying the *bail-out costs* of the aggregated system $\Lambda(Y)$ under ρ . In particular, the risk measure R^{ins} is *insensitive* to the capital allocation to each financial firm, disregarding possible transaction costs or other dependence structures between the financial firms and ignoring how the addition of capital changes the system itself. This justifies the mnemonic terminology. 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 ρ , R^{ins} takes the equivalent form $R^{\text{ins}}(Y) = \{k \in \mathbb{R}^d : \rho(\Lambda(Y)) \leq \bar{k}\}$. This means that R^{ins} is actually a bijection of the scalar risk measure $\rho \circ \Lambda : \mathcal{Y}^d \rightarrow \mathbb{R}^*$ considered in Chen et al. [13]. Therefore, one has to evaluate the risk measure ρ only once to determine R^{ins} . In contrast, such an appealing equivalent formulation is generally not available for R unless Λ is additive, or is even the sum in which case R and R^{ins} coincide. Consequently, in general, one is bound to evaluate ρ infinitely often to compute R ; see also the discussion in Feinstein et al. [23]. The main focus of this paper are elicibility and identifiability results for systemic risk measures of the form (2.1) and (2.2). However, since one can exploit the one-to-one relation between R^{ins} and $\rho \circ \Lambda$ and make use of the revelation principle, see Fissler [24, Sect. 2.3], Gneiting [35] and Osband [55, Sect. 2.1], to establish (exhaustive) elicibility and identifiability results, we do not present results about R^{ins} in this paper, but rather defer them to Fissler et al. [26, Supplementary Material].

For the sake of completeness, we recall the most important properties of R presented in Feinstein et al. [23]. Because ρ is cash-invariant and Λ is increas-

ing, the values of both R and R^{ins} defined in (2.1) and (2.2) are *upper sets*, i.e., $R(Y) = R(Y) + \mathbb{R}_+^d$ for any $Y \in \mathcal{Y}^d$, where \mathbb{R}_+^d denotes the collection of vectors in \mathbb{R}^d with only nonnegative elements and for any two sets $A, B \subseteq \mathbb{R}^d$, $A + B := \{a + b : a \in A, b \in B\}$ is the usual *Minkowski sum*. Recall that we have $A + \emptyset = \emptyset + A = \emptyset$. Following the notation of [23], we denote the collection of upper sets in \mathbb{R}^d with ordering cone \mathbb{R}_+^d by $\mathcal{P}(\mathbb{R}^d; \mathbb{R}_+^d) := \{B \subseteq \mathbb{R}^d : B = B + \mathbb{R}_+^d\}$. Both \mathbb{R}^d and \emptyset are elements of $\mathcal{P}(\mathbb{R}^d; \mathbb{R}_+^d)$. Moreover, R defined in (2.1) can attain these values even if the underlying scalar risk measure ρ maps to \mathbb{R} only, e.g. when Λ is bounded. While $R(Y) = \emptyset$ corresponds to the case that a scalar risk measure of the financial position Y is $+\infty$, meaning that the system Y 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 usually only discuss the case $R(Y) = \emptyset$, but remark that a treatment of the case $R(Y) = \mathbb{R}^d$ would also be possible for most results. Monotonicity and cash-invariance carry over to R in that $R(Y) \supseteq R(Z)$ for all $Y, Z \in \mathcal{Y}^d$ with $Y \geq Z$ \mathbb{P} -a.s. componentwise, and $R(Y + k) = R(Y) - k$ for all $k \in \mathbb{R}^d$. Monotonicity also carries over to R^{ins} ; note, however, that R^{ins} is in general not cash-invariant. We introduce further subclasses of $\mathcal{P}(\mathbb{R}^d; \mathbb{R}_+^d)$, where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel- σ -algebra on \mathbb{R}^d .

Definition 2.1 The class of *Borel-measurable upper subsets* of \mathbb{R}^d is denoted by $\widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d) := (\mathcal{P}(\mathbb{R}^d; \mathbb{R}_+^d) \cap \mathcal{B}(\mathbb{R}^d)) \setminus \{\mathbb{R}^d\}$. The class of *closed upper subsets* of \mathbb{R}^d is denoted by $\mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$. Note that $\mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d) \subseteq \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$.

For any set $A \subseteq \mathbb{R}^d$, we denote its topological boundary by ∂A . We introduce the law-invariant map

$$R_0 : \mathcal{Y}^d \rightarrow 2^{\mathbb{R}^d}, \quad Y \mapsto R_0(Y) = \{k \in \mathbb{R}^d : \rho(\Lambda(Y + k)) = 0\}. \tag{2.3}$$

Occasionally and when explicitly stated, we impose one of the following assumptions.

Assumption 2.2 For all $Y \in \mathcal{Y}^d$, $R(Y) \in \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$.

Assumption 2.3 For all $Y \in \mathcal{Y}^d$, $R(Y) \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$ and $R_0(Y) = \partial R(Y)$.

A sufficient condition for $R(Y) \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$ is that for any convergent sequence $(k_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}^d$ with limit k , we have

$$\rho(\Lambda(Y + k)) \leq \liminf_{n \rightarrow \infty} \rho(\Lambda(Y + k_n)). \tag{2.4}$$

The inequality in (2.4) holds e.g. if Λ is continuous and if $\rho(X) \leq \liminf_{n \rightarrow \infty} \rho(X_n)$ for all sequences (X_n) in \mathcal{Y} converging almost surely to $X \in \mathcal{Y}$. In particular, if $\mathcal{Y} = L^\infty(\Omega; \mathbb{R})$, ρ is convex, Λ continuous, and either (a) Λ is bounded (invoking the law-invariance of ρ and the results from Jouini et al. [44] and Svindland [60]) or

(b) Λ is uniformly continuous, then (2.4) holds. For instance in the network model considered in Sect. 5.1, Λ is bounded. Otherwise in the literature, Λ is often a concave function and thus not bounded (unless it is constant). Hence, for concave Λ , one could check whether it is uniformly continuous. This is clearly the case for instance for the most straightforward choice—the sum—as well as for instance for the aggregation function suggested by Amini et al. [4], also considered in Sect. 5.1. Moreover, note that we provide sufficient conditions only and not necessary ones, so that one may also check $R(Y) \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$ on a case-by-case basis.

If $\emptyset \neq R(Y) \in \mathcal{F}(\mathbb{R}^d; \mathbb{R})$, then $R_0(Y) \neq \emptyset$. Since Λ is increasing and ρ is cash-invariant, one obtains $R(Y) = R_0(Y) + \mathbb{R}_+^d$. Hence the values of R_0 determine R completely. Moreover, if Λ is strictly increasing, we have $R_0(Y) = \partial R(Y)$, meaning that $R_0(Y)$ contains the *efficient* capital allocations that make Y acceptable under R . Therefore, under Assumption 2.3, R and R_0 are connected via a one-to-one relation. Again invoking the revelation principle [35, Theorem 4], this means that exhaustive elicibility results for R (Theorem 3.10 (iii)) carry over to R_0 . In a nutshell, the *revelation principle* asserts that if there is a bijection, say g , such that $R_0 = g(R)$, then R_0 is (exhaustively) elicitable if and only if R is elicitable. Moreover, $S(A, y)$ is strictly consistent for R if and only if $S(g^{-1}(A), y)$ is strictly consistent for R_0 .

Finally, we recall the definition of an important scalarisation of the systemic risk measure R , called *efficient cash-invariant allocation rule* (EAR), as introduced in Feinstein et al. [23]. Roughly speaking, for $Y \in \mathcal{Y}^d$, $\text{EAR}(Y)$ specifies the capital allocations with minimal weighted cost among allocations in $R(Y)$. For simplicity, we confine our attention to the situation when $R(Y)$ is closed and to EARs with a *fixed* price or weight vector $w \in \mathbb{R}_{++}^d := \{x \in \mathbb{R}^d : x_1, \dots, x_d > 0\}$.

Definition 2.4 Suppose that $\emptyset \neq R(Y) \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$ for all $Y \in \mathcal{Y}^d$. An *efficient cash-invariant allocation rule* for a fixed price vector $w \in \mathbb{R}_{++}^d$ is given by

$$\text{EAR}_w(Y) := \arg \min_{k \in R(Y)} w^\top k. \tag{2.5}$$

For $Y \in \mathcal{Y}^d$, if there is a supporting hyperplane of $R(Y)$ orthogonal to w , then $\text{EAR}_w(Y)$ is the intersection of $\partial R(Y)$ and this hyperplane. Hence $\text{EAR}_w(Y)$ is not necessarily a singleton. If there is no supporting hyperplane of $R(Y)$ orthogonal to w , the function $R(Y) \ni k \mapsto w^\top k$ is unbounded from below and we set $\text{EAR}_w(Y) = \emptyset$.

Since ρ is law-invariant, so are the derived quantities R , R^{ins} , R_0 and EAR_w . Therefore, in analogy to our treatment of ρ , we identify R with the risk functional \hat{R} , where $\hat{R}(F) = R(Y)$ for $Y \sim F \in \mathcal{M}^d$, and simply write R for either; we use analogous conventions for R_0 and EAR_w .

2.2 Elicibility and identifiability of set-valued functionals

We have already mentioned the definitions of elicibility and identifiability for scalar risk measures $\rho : \mathcal{M} \rightarrow \mathbb{R}^*$ in (1.2) and (1.3), where we slightly extend the common definitions to account for ρ possibly attaining ∞ . All other risk measures considered here, R , R_0 and EAR, are set-valued, attaining subsets of \mathbb{R}^d . Hence we make use of

the theoretical framework on forecast evaluation of set-valued functionals introduced in Fissler et al. [25]. The main idea is to have a thorough distinction concerning the form of the forecasts between a *selective* notion where forecasts are single points, and an *exhaustive* mode where forecasts are set-valued. Moreover, corresponding notions of identifiability and elicibility are introduced and discussed in a very general setting, with the main result being that—subject to mild conditions—a set-valued functional is elicitable either in the selective, or the exhaustive sense, or not elicitable at all [25, Theorem 2.14]. We confine ourselves to introducing only the notions we discuss in this paper and do so directly in terms of R and R_0 ; the case of EARs is considered separately later. In the sequel, let $\mathcal{A} \subseteq 2^{\mathbb{R}^d}$. Moreover, for scoring functions $S: \mathcal{A} \times \mathbb{R}^d \rightarrow \mathbb{R}^*$ or identification functions $V: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, we use the shorthands $\bar{S}(A, F) := \int S(A, y) dF(y)$ and $\bar{V}(x, F) := \int V(x, y) dF(y)$ for $A \in \mathcal{A}$, $x \in \mathbb{R}^d$, and tacitly assume that these integrals exist for all $F \in \mathcal{M}^d$, where we say that the integral $\int g(y) dF(y)$ of a function $g: \mathbb{R} \rightarrow [-\infty, \infty]$ exists if g is measurable and $\int g(y)^+ dF(y) < \infty$ or $\int g(y)^- dF(y) < \infty$. In that case, we set

$$\int g(y) dF(y) := \int g(y)^+ dF(y) - \int g(y)^- dF(y) \in [-\infty, \infty].$$

Definition 2.5 A map $S: \mathcal{A} \times \mathbb{R}^d \rightarrow \mathbb{R}^*$ is an \mathcal{M}^d -consistent exhaustive scoring function for $R: \mathcal{M}^d \rightarrow \mathcal{A}$ if

$$\bar{S}(R(F), F) \leq \bar{S}(A, F), \quad \forall A \in \mathcal{A}, \forall F \in \mathcal{M}^d. \tag{2.6}$$

The map S is a *strictly* \mathcal{M}^d -consistent exhaustive scoring function for R if it is \mathcal{M}^d -consistent for R and if equality in (2.6) implies that $A = R(F)$. Finally, the risk measure $R: \mathcal{M}^d \rightarrow \mathcal{A}$ is *exhaustively elicitable* if there is a strictly \mathcal{M} -consistent exhaustive scoring function for R .

Note that the *strict* consistency of an exhaustive scoring function S for R implies that $\bar{S}(R(F), F) \in \mathbb{R}$ for all $F \in \mathcal{M}^d$.

Definition 2.6 A map $V: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a *selective* \mathcal{M}^d -identification function for $R_0: \mathcal{M}^d \rightarrow \mathcal{A}$ if $\bar{V}(x, F) = 0$ for all $x \in R_0(F)$ and all $F \in \mathcal{M}^d$. Moreover, V is a *strict selective* \mathcal{M}^d -identification function for R_0 if for all $x \in \mathbb{R}^d$ and all $F \in \mathcal{M}^d$, it holds that $\bar{V}(x, F) = 0$ if and only if $x \in R_0(F)$. Finally, the risk measure $R_0: \mathcal{M}^d \rightarrow \mathcal{A}$ is *selectively identifiable* if there is a strict selective \mathcal{M}^d -identification function for R_0 .

3 Main results

We present the main results of the paper in this section, with identifiability results in Sect. 3.1 and elicibility results in Sect. 3.2. Theorem 3.1 establishes the selective identifiability of R_0 . Notably, the main assumption behind Theorem 3.1 and the subsequent results relying on this identifiability is the identifiability of the underlying scalar risk measure ρ in (2.1). A fortiori, ρ needs to admit an oriented identification function. According to Steinwart et al. [58], a strict identification function

$V_\rho: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ for a scalar risk measure $\rho: \mathcal{M} \rightarrow \mathbb{R}^*$ is called *oriented* if for all $x \in \mathbb{R}$ and all $F \in \mathcal{M}$, it holds that $\bar{V}_\rho(x, F) < 0$ if and only if $x < \rho(F)$. Invoking [58, Theorem 8], the existence of an oriented identification function for ρ is equivalent to the elicibility of ρ under mild regularity conditions. Proposition 3.9 establishes that under certain assumptions on the aggregation function Λ , also the converse holds: the selective identifiability of R_0 implies the identifiability of ρ .

3.1 Identifiability results

Theorem 3.1 *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 in (2.3):*

(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 ρ , then*

$$V_{R_0}: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}, \quad (k, y) \mapsto V_{R_0}(k, y) = V_\rho(0, \Lambda(y + k)) \quad (3.1)$$

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

(ii) *If V_ρ is an oriented strict \mathcal{M} -identification function for ρ , then V_{R_0} defined in (3.1) is oriented for R_0 in the sense that for all $F \in \mathcal{M}^d$, it holds that*

$$\bar{V}_{R_0}(k, F) \begin{cases} < 0, & \text{if } k \notin R(F), \\ = 0, & \text{if } k \in R_0(F), \\ > 0, & \text{if } k \in R(F) \setminus R_0(F). \end{cases} \quad (3.2)$$

Proof (i) Let V_ρ be a strict \mathcal{M} -identification function for ρ . This means that for all $Y \sim F \in \mathcal{M}^d$ and all $k \in \mathbb{R}^d, x \in \mathbb{R}$, one has

$$\mathbb{E}_F[V_\rho(x, \Lambda(Y + k))] = 0 \iff x = \rho(\Lambda(Y + k)). \quad (3.3)$$

Setting $x = 0$ in (3.3) yields

$$\mathbb{E}_F[V_\rho(0, \Lambda(Y + k))] = 0 \iff 0 = \rho(\Lambda(Y + k)) \iff k \in R_0(Y),$$

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

(ii) Now assume that V_ρ is an oriented strict \mathcal{M} -identification function for ρ . This means that for all $Y \sim F \in \mathcal{M}^d$ and all $k \in \mathbb{R}^d, x \in \mathbb{R}$, one has

$$\mathbb{E}_F[V_\rho(x, \Lambda(Y + k))] \begin{cases} < 0, & \text{if } x < \rho(\Lambda(Y + k)), \\ = 0, & \text{if } x = \rho(\Lambda(Y + k)), \\ > 0, & \text{if } x > \rho(\Lambda(Y + k)). \end{cases} \quad (3.4)$$

Setting $x = 0$ in (3.4) yields the claim. □

Interestingly, if $V_{R_0}: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is an oriented strict selective \mathcal{M}^d -identification function for R_0 and $\bar{V}_{R_0}(\cdot, F), F \in \mathcal{M}^d$, is continuous, then $R(F)$ is closed.

Remark 3.2 If V_{R_0} is oriented in the sense of (3.4), the full risk measure R can also be selectively ‘identified’ by checking the sign of the expected identification function $\bar{V}_{R_0}(k, F)$. Even though we are unaware of a result that excludes its exhaustive identifiability in the sense of Fissler et al. [25] per se, we do not see a way to come up with a (statistically feasible) exhaustive identification function for R .

Remark 3.3 Equation (3.1) explicitly constructs a strict selective \mathcal{M}^d -identification function $V_{R_0} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ for R_0 , given a certain strict \mathcal{M} -identification function $V_\rho : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ for ρ . Hence, such a V_{R_0} definitely depends on the choice of V_ρ . Fissler [24, Proposition 3.2.1] states that under some richness assumptions on the class \mathcal{M} , any other strict identification function $\tilde{V}_\rho : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ for ρ is of the form $\tilde{V}_\rho(x, z) = g(x)V_\rho(x, z)$, where $g : \mathbb{R} \rightarrow \mathbb{R}$ is non-vanishing. If V_ρ is oriented, then \tilde{V}_ρ is oriented if and only if $g > 0$. Consequently, starting with such an identification function \tilde{V}_ρ , the resulting (oriented) strict selective \mathcal{M}^d -identification function $\tilde{V}_{R_0} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ 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 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.

In a similar spirit as Remark 3.3, one might also wonder whether the (oriented) strict selective identification functions constructed in Theorem 3.1 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 $V'_{R_0} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ with

$$V'_{R_0}(k, y) = h(k)V_{R_0}(k, y) = h(k)V_\rho(0, \Lambda(y + k)), \tag{3.5}$$

where $h : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 in (3.5) is oriented if and only if $h > 0$. In particular, the constant $g(0)$ appearing in Remark 3.3 can be incorporated into the function h so that we see that it does not matter which (oriented) strict exhaustive identification function \tilde{V}_ρ we choose to end up with the form in (3.5). The next result establishes that basically all selective \mathcal{M}^d -identification functions for R_0 are of the form (3.5).

Proposition 3.4 *Let $A \subseteq \mathbb{R}^d$ and let $V_{R_0}, V'_{R_0} : A \times \mathbb{R}^d \rightarrow \mathbb{R}$ be strict selective \mathcal{M}^d -identification functions for $R_0 : \mathcal{M}^d \rightarrow 2^{\mathbb{R}^d}$. If for every $x \in A$, there are $F_1, F_2 \in \mathcal{M}^d$ such that $\bar{V}_{R_0}(x, F_1) > 0$ and $\bar{V}_{R_0}(x, F_2) < 0$ and if \mathcal{M}^d is convex, then there is a non-vanishing function $h : A \rightarrow \mathbb{R}$ such that*

$$\bar{V}'_{R_0}(x, F) = h(x)\bar{V}_{R_0}(x, F) \tag{3.6}$$

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

Proof The proof follows along the lines of the proof of Fissler and Ziegel [27, Theorem 3.2]; see Osband [55, Theorem 2.1]. The dimensionality of x does not play any

role in the proof. As our identification functions map to \mathbb{R} , we use $k = 1$ in the proof of [27, Theorem 3.2]. 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 [27, Assumption (V1)]. If we replace $\nabla \bar{S}(x, F)$ in the proof of [27, Theorem 3.2] by $\bar{V}'(x, F)$, we obtain that there is a function $h: A \rightarrow \mathbb{R}$ 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 \mathbb{B}_G in the proof of [27, Theorem 3.2] will be a 2×3 matrix of rank 1 for any $x \in A$, $h(x)$ has to be nonzero for all $x \in A$. \square

Remark 3.5 (i) Note that the assumptions of Proposition 3.4 imply that for all $x \in A$, there is some $F \in \mathcal{M}^d$ such that $x \in R_0(F)$. That is why we formulated the result in terms of a general action domain $A \subseteq \mathbb{R}^d$.

(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 (3.6); see [27, 28] for details.

Finally, we turn our attention to EARs as introduced in Definition 2.4. For any price or weight 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 as usual the space of all functions from w^\perp to \mathbb{R} .

Proposition 3.6 *Suppose that Assumption 2.3 holds and that R_0 has a selective oriented strict \mathcal{M}^d -identification function V_{R_0} in the sense of (3.2). Let $w \in \mathbb{R}_{++}^d$ and define the map $V_{\text{EAR}_w} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{w^\perp}$ via*

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

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 $F \in \mathcal{M}^d, k \in \mathbb{R}^d$, we have

$$k \in \text{EAR}_w(F) \iff \bar{V}_{\text{EAR}_w}(k, F) \leq 0 \text{ and } \bar{V}_{\text{EAR}_w}(k, F)(0) = 0. \quad (3.7)$$

Proof Under Assumption 2.3, (3.7) is equivalent to

$$k \in \text{EAR}_w(F) \iff (k + w^\perp) \cap \text{int}(R(F)) = \emptyset \text{ and } k \in R_0(F).$$

For the ‘ \Rightarrow ’ part, assume that $k \in \text{EAR}_w(F)$. Then clearly $k \in R_0(F)$. Suppose $k + h \in \text{int}(R(F))$ for some $h \in w^\perp$. Then there is $\epsilon > 0$ such that $k + h - \epsilon \mathbf{1} \in R(F)$ and we get $w^\top(k + h - \epsilon \mathbf{1}) = w^\top k - \epsilon w^\top \mathbf{1} < w^\top k$, which is a contradiction.

For the ‘ \Leftarrow ’ part, assume that $k \in R_0(F)$ and $(k + w^\perp) \cap \text{int}(R(F)) = \emptyset$, and let $h \in R(F)$ be such that $w^\top h < w^\top k$, that is, $w^\top(k - h) > 0$. Then there are $\alpha \in \mathbb{R}$ and $\ell \in w^\perp$ with $\mathbf{1} = \alpha(k - h) + \ell$. By multiplying with w^\top from the left, one gets that $\alpha = \frac{w^\top \mathbf{1}}{w^\top(k - h)} > 0$. The fact that $R(F)$ is an upper set implies that $k + \frac{1}{\alpha} \ell = h + \frac{1}{\alpha} \mathbf{1} \in R(F) + \mathbb{R}_{++}^d \subseteq \text{int} R(F)$, which contradicts the assumption that $(k + w^\perp) \cap \text{int}(R(F)) = \emptyset$. Therefore $w^\top k \leq w^\top h$ for all $h \in R(F)$, which means that $k \in \text{EAR}_w(F)$. \square

If the underlying risk measure R is known to assume convex sets only (e.g. if ρ is convex and Λ concave, see Feinstein et al. [23]), it is even sufficient to evaluate

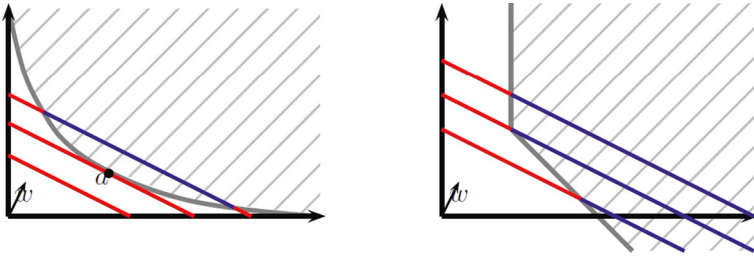


Fig. 1 An illustration of Proposition 3.6 for dimension $d = 2$. In both panels, $R(Y)$ (depicted in grey) is convex. Straight lines correspond to affine translations of w^\perp . For k in the red area, $\bar{V}_{R_0}(k + x, F) \leq 0$, and if k is in the blue area, $\bar{V}_{R_0}(k + x, F) > 0$. In the left panel, $\text{EAR}_w(Y)$ exists and is a singleton consisting of the point a . In the right panel, the function $k \mapsto w^\top k, k \in R(Y)$, is unbounded from below so that $\text{EAR}_w(Y) = \emptyset$

$\bar{V}_{\text{EAR}_w}(k, F)(x)$, or its empirical counterpart, for $x \in w^\perp$ in a neighbourhood of 0, which can also be seen nicely in Fig. 1.

In Fissler et al. [25, Sect. 3.1], different versions of the convex level sets (CxLS) property are introduced and their necessity for identifiability and elicibility for set-valued functionals is discussed. Our next result establishes the so-called *selective CxLS* property* of EAR_w . That is, for all $F_0, F_1 \in \mathcal{M}^d$ and all $\lambda \in (0, 1)$ such that $F_\lambda := (1 - \lambda)F_0 + \lambda F_1 \in \mathcal{M}^d$, it holds that

$$\text{EAR}_w(F_0) \cap \text{EAR}_w(F_1) \neq \emptyset \implies \text{EAR}_w(F_0) \cap \text{EAR}_w(F_1) = \text{EAR}_w(F_\lambda). \tag{3.8}$$

Proposition 3.7 *Suppose that Assumption 2.3 holds and R_0 has a selective oriented strict \mathcal{M}^d -identification function V_{R_0} in the sense of (3.2). Then $\text{EAR}_w, w \in \mathbb{R}_{++}^d$, satisfies the selective CxLS* property (3.8).*

Proof Assume that $k \in \text{EAR}_w(F_0) \cap \text{EAR}_w(F_1)$. Then for $\lambda \in (0, 1)$ with $F_\lambda \in \mathcal{M}^d$, we obtain

$$\bar{V}_{\text{EAR}_w}(k, F_\lambda)(x) = (1 - \lambda)\bar{V}_{R_0}(k + x, F_0) + \lambda\bar{V}_{R_0}(k + x, F_1) \leq 0$$

for all $x \in w^\perp$ and $\bar{V}_{R_0}(k, F_\lambda) = 0$, hence $\text{EAR}_w(F_0) \cap \text{EAR}_w(F_1) \subseteq \text{EAR}_w(F_\lambda)$. Now let $\ell \in \text{EAR}_w(F_\lambda)$, but $\ell \notin \text{EAR}_w(F_0) \cap \text{EAR}_w(F_1)$. We obtain that

$$(1 - \lambda)\bar{V}_{R_0}(\ell, F_0) = -\lambda\bar{V}_{R_0}(\ell, F_1)$$

and since $\ell \notin \text{EAR}_w(F_0) \cap \text{EAR}_w(F_1)$, both $\bar{V}_{R_0}(\ell, F_0)$ and $\bar{V}_{R_0}(\ell, F_1)$ must be nonzero and of opposite signs. Assume without loss of generality that $\bar{V}_{R_0}(\ell, F_0) < 0$ and $\bar{V}_{R_0}(\ell, F_0) > 0$. For any $k \in \text{EAR}_w(F_0) \cap \text{EAR}_w(F_1)$, we have $w^\top \ell = w^\top k$ so that $\ell \in k + w^\perp$. This, however, leads to a contradiction since $k \in \text{EAR}_w(F_1)$ implies $\bar{V}_{R_0}(x, F_1) \leq 0$ for all $x \in k + w^\perp$. \square

If \mathcal{M} is such that EAR_w is ‘truly set-valued’ on \mathcal{M} , and in particular if it satisfies the *proper subset property* of Fissler et al. [25, Definition 3.4], i.e., there exist

$F, G \in \mathcal{M}$ such that $\emptyset \neq \text{EAR}_w(G) \subsetneq \text{EAR}_w(F)$ and \mathcal{M} is convex, then [25, Theorem 3.5] asserts that EAR_w is not exhaustively elicitable on \mathcal{M} under the conditions of Proposition 3.7.

Remark 3.8 We should like to compare the concept of identifiability introduced in Proposition 3.6 to the discussion about the backtestability of *loss value at risk* in Bignozzi et al. [12, Sect. 5]. One can interpret the backtesting procedure suggested in [12] as using a function-valued identification function as well. From that angle, the analogue of (3.7) in the context of [12] would be that the infimum of the function-valued identification function is 0 when using the correctly specified forecast. Interestingly, this version of identifiability does not imply that the functional under consideration satisfies one of the CxLS properties of [25, Sect. 3.1].

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

Proposition 3.9 *Let $\Lambda: \mathbb{R}^d \rightarrow \mathbb{R}$ be surjective with a measurable right inverse $\eta: \mathbb{R} \rightarrow \mathbb{R}^d$, i.e., η satisfies $\Lambda \circ \eta = \text{id}_{\mathbb{R}}$. Assume that $\eta(X)$ belongs to \mathcal{Y}^d for any $X \in \mathcal{Y}$. Then ρ is identifiable if and only if R_0 is selectively identifiable.*

Proof The ‘only if’ part is a special case of Theorem 3.1. For the ‘if’ part, assume that $V_{R_0}: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a strict selective \mathcal{M}^d -identification function for R_0 . For any $Y \in \mathcal{Y}^d$, it holds that

$$\mathbb{E}[V_{R_0}(0, Y)] = 0 \iff 0 \in R_0(Y) \iff \rho(\Lambda(Y)) = 0.$$

Then for any $s \in \mathbb{R}$ and $X \in \mathcal{Y}$, we obtain that

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

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

3.2 Elicitability results and mixture representation

In the seminal paper Ehm et al. [19], it is shown that subject to regularity conditions, any nonnegative scoring function $S: \mathbb{R} \times \mathbb{R} \rightarrow [0, \infty]$ consistent for the α -quantile (the τ -expectile) can be written as a *mixture* or *Choquet representation*

$$S(x, y) = \int_{\mathbb{R}} S_\theta(x, y) dH(\theta), \quad x, y \in \mathbb{R}, \tag{3.9}$$

where H is a nonnegative measure on $\mathcal{B}(\mathbb{R})$ and $S_\theta, \theta \in \mathbb{R}$, are nonnegative elementary scoring functions for the α -quantile (the τ -expectile). In particular, S_θ takes the form

$$S_\theta(x, y) = (\mathbb{1}_{\{\theta < x\}} - \mathbb{1}_{\{\theta < y\}})V(\theta, y), \tag{3.10}$$

where V is an *oriented* identification function for the α -quantile (the τ -expectile). The score in (3.9) is strictly consistent if and only if the measure H is strictly positive, that is, puts positive mass on any open nonempty set. Ziegel [64] and Dawid [16] argued that besides expectiles and quantiles, this construction also works for more general one-dimensional functionals which admit an oriented identification function; see Jordan et al. [43]. Steinwart et al. [58] showed that for one-dimensional functionals satisfying certain regularity conditions, the existence of an oriented identification function is equivalent to the elicibility 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 in (3.9), an answer to the question as to whether all nonnegative scoring functions for a certain functional are *necessarily* of the form in (3.9) can typically only be answered by invoking Osband’s principle (see Fissler and Ziegel [27] and Osband [55, Theorem 2.1]), 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 in Theorem 3.1 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 use the notation $R(y) := R(\delta_y)$.

Theorem 3.10 *Let $V_{R_0} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be measurable in its first argument and such that for all $F \in \mathcal{M}^d \cup \{\delta_y : y \in \mathbb{R}^d\}$, we have*

$$\bar{V}_{R_0}(k, F) \in \begin{cases} (-\infty, 0], & \text{if } k \notin R(F), \\ [0, \infty), & \text{if } k \in R(F). \end{cases} \tag{3.11}$$

(i) *Under Assumption 2.2, the map $S_{R,k} : \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d) \times \mathbb{R}^d \rightarrow [0, \infty)$,*

$$S_{R,k}(A, y) = (\mathbb{1}_{R(y) \setminus A}(k) - \mathbb{1}_{A \setminus R(y)}(k)) V_{R_0}(k, y) \tag{3.12}$$

is for each $k \in \mathbb{R}^d$ a nonnegative \mathcal{M}^d -consistent exhaustive scoring function for $R : \mathcal{M}^d \rightarrow \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$.

(ii) *Under Assumption 2.2 and if π is a σ -finite nonnegative measure on $\mathcal{B}(\mathbb{R}^d)$, the nonnegative map $S_{R,\pi} : \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d) \times \mathbb{R}^d \rightarrow [0, \infty]$,*

$$S_{R,\pi}(A, y) = \int_{\mathbb{R}^d} S_{R,k}(A, y) \pi(dk) \tag{3.13}$$

is an \mathcal{M}^d -consistent exhaustive scoring function for $R : \mathcal{M}^d \rightarrow \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$.

(iii) *If Assumption 2.3 holds, if for all $F \in \mathcal{M}^d$ it holds that $\bar{V}_{R_0}(k, F) < 0$ for $k \notin R(F)$ and $\bar{V}_{R_0}(k, F) > 0$ for $k \in \text{int}(R(F))$, and if π is a σ -finite strictly positive measure on $\mathcal{B}(\mathbb{R}^d)$, then the restriction of $S_{R,\pi}$ defined in (3.13) to $\mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d) \times \mathbb{R}^d$ is strictly \mathcal{M}_0^d -consistent for the map $R : \mathcal{M}_0^d \rightarrow \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$, where $\mathcal{M}_0^d \subseteq \mathcal{M}^d$ is such that $\bar{S}_{R,\pi}(R(F), F) < \infty$ for all $F \in \mathcal{M}_0^d$.*

For the proof of Theorem 3.10, we need the following lemma.

Lemma 3.11 *For any $A_1, A_2 \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$, the symmetric difference $A_1 \Delta A_2$ defined as $(A_1 \setminus A_2) \cup (A_2 \setminus A_1)$ is empty if and only if its interior $\text{int}(A_1 \Delta A_2)$ is empty.*

Proof If $A_1 \Delta A_2 = \emptyset$, it is clear that $\text{int}(A_1 \Delta A_2) = \emptyset$. Assume that there is an $x \in A_1 \Delta A_2$. Without loss of generality, we can assume that $x \in A_1 \setminus A_2$. As A_2 is closed, one can find $r > 0$ such that $B_r(x) \cap A_2 = \emptyset$. Then there is $\epsilon > 0$ such that $x + \epsilon \mathbf{1} \in B_r(x)$. Because A_1 is an upper set, we also get that $x + \epsilon \mathbf{1} \in \text{int}(A_1)$ which shows that $\text{int}(A_1 \setminus A_2) \neq \emptyset$. \square

Proof of Theorem 3.10 (i) Let $F \in \mathcal{M}^d \cup \{\delta_y : y \in \mathbb{R}^d\}$, $A \in \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$ and $k \in \mathbb{R}^d$. A direct calculation yields that

$$\bar{S}_{R,k}(A, F) - \bar{S}_{R,k}(R(F), F) = (\mathbb{1}_{R(F) \setminus A}(k) - \mathbb{1}_{A \setminus R(F)}(k)) \bar{V}_{R_0}(k, F) \geq 0,$$

where the last inequality is a direct consequence of (3.11). The nonnegativity of $S_{R,k}$ follows by choosing $F = \delta_y$, $y \in \mathbb{R}^d$, which yields $S_{R,k}(A, y) \geq S_{R,k}(R(y), y) = 0$.

Claim (ii) is a direct consequence of the nonnegativity and consistency of $S_{R,k}$.

(iii) Let $F \in \mathcal{M}^d$ and $A^* := R(F)$, $A \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$ with $A \neq A^*$. Assume that $\bar{S}_{R,\pi}(A, F), \bar{S}_{R,\pi}(A^*, F) < \infty$ (otherwise, there is nothing to show). Using Fubini's theorem, we obtain that $\bar{S}_{R,\pi}(A, F) - \bar{S}_{R,\pi}(A^*, F)$ equals

$$\int_{A^* \setminus A} \bar{V}_{R_0}(k, F) \pi(dk) - \int_{A \setminus A^*} \bar{V}_{R_0}(k, F) \pi(dk).$$

Then Lemma 3.11 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 π assigns positive mass to any nonempty open set in $\mathcal{B}(\mathbb{R}^d)$ implies $\int_{A \setminus A^*} \bar{V}_{R_0}(k, F) \pi(dk) < 0$, 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 $\partial A^* = R_0(F)$ is a closed set. That means that $\text{int}(A^* \setminus A) \setminus \partial A^*$ is open and nonempty. Moreover, the functional $\bar{V}_{R_0}(\cdot, F)$ is strictly positive on $\text{int}(A^* \setminus A) \setminus \partial A^*$. Hence we obtain

$$\int_{A^* \setminus A} \bar{V}_{R_0}(k, F) \pi(dk) \geq \int_{\text{int}(A^* \setminus A) \setminus \partial A^*} \bar{V}_{R_0}(k, F) \pi(dk) > 0,$$

which implies that $\bar{S}_{R,\pi}(A, F) - \bar{S}_{R,\pi}(A^*, F) > 0$. A graphic illustration of the situation is provided in Figure 2. \square

Note that condition (3.11) or the similar condition on \bar{V}_{R_0} in part (iii) of Theorem 3.10 does not imply that V_{R_0} is an identification function for R_0 . This relaxation is particularly beneficial if ρ is value at risk and \mathcal{M} contains discontinuous distributions, so that value at risk possibly fails to be identifiable on \mathcal{M} .

3.2.1 Comparison with the one-dimensional case

The similarity of the mixture representations in (3.13) and (3.9) is obvious. With a closer look, one can also see the similarities on the level of the elementary scores

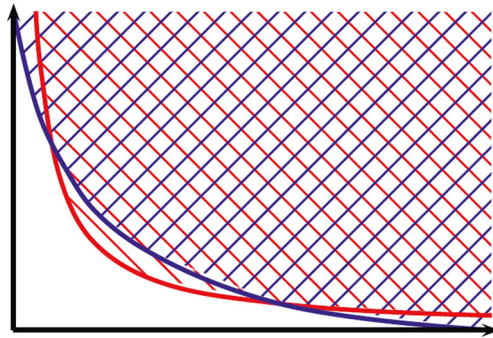


Fig. 2 A graphical illustration of the proof of Theorem 3.10 for dimension $d = 2$. The red region corresponds to the correctly specified risk measure $R(F)$ and the blue region to some misspecified forecast A . The score difference $\bar{S}_{R,\pi}(A, F) - \bar{S}_{R,\pi}(R(F), F)$ is then an integral of $\bar{V}_{R_0}(\cdot, F)$ over $R(F) \setminus A$ (the red-only region) plus an integral of $-\bar{V}_{R_0}(\cdot, F)$ over $A \setminus R(F)$ (the blue-only region)

given in (3.12) and (3.10). Indeed, (3.10) can be rewritten as

$$S_\theta(x, y) = (\mathbb{1}_{[y, \infty) \setminus [x, \infty)}(\theta) - \mathbb{1}_{[x, \infty) \setminus [y, \infty)}(\theta))V(\theta, y).$$

The form of $R(y)$ is described explicitly in Lemma 3.12, where we use the fact that $\rho(\rho(0)) = \rho(0) - \rho(0) = 0$.

Lemma 3.12 *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 an expectile is a scalar risk measure, one can see that the elementary scores in (3.12) essentially boil down to the ones in (3.10) for dimension $d = 1$.

3.2.2 Integrability

The nonnegativity of the elementary scores in (3.12) guarantees that the integral in (3.13) always exists. However, as stated in part (iii) of Theorem 3.10, these scores are only strictly consistent if $\bar{S}_{R,\pi}(R(F), F) < \infty$, which suggests the question as to when the integral in (3.13) is finite. A sufficient condition for this integrability is that $\int_{\mathbb{R}^d} |V_{R_0}(k, y)| \pi(dk) < \infty$. Therefore, a sufficient condition for the finiteness of $\bar{S}_{R,\pi}(A, F)$ is that V_{R_0} is $(\pi \otimes F)$ -integrable.

3.2.3 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 in (3.13). This means that modulo equivalence, the choice of the consistent scoring function boils down to the choice of the measure π .

First, note that Proposition 3.4 implies that it does not matter which 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 strictly positive function h . But this only amounts to a change of measure as $V_{R_0}(k, y)\pi(dk) = V'_{R_0}(k, y)\pi'(dk)$, where π' has the density $1/h$ with respect to π . Secondly, the class of scoring functions of the form (3.13) is convex, which is a necessary condition; see Gneiting [35, Theorem 2]. Thirdly, as observed above, the mixture representation in (3.13) is the natural extension of the one-dimensional case. To answer the question whether indeed all scoring functions for R are equivalent to a score of the form given in (3.13), one would need to generalise Osband’s principle from the finite-dimensional case in Fissler and Ziegel [27] to the infinite-dimensional setting of reporting upper sets in $\mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$.

3.2.4 Order-sensitivity

Under weak assumptions on ρ , all strictly consistent scoring functions S for ρ are *order-sensitive* or *accuracy-rewarding*; see Nau [52, Proposition 3], Lambert [49, Proposition 1], Bellini and Bignozzi [9, Proposition 3.4]. In the scalar setting, this property means that $x_1 \leq x_2 \leq \rho(F)$ or $\rho(F) \leq x_2 \leq x_1$ implies $\bar{S}(x_1, F) \geq \bar{S}(x_2, F)$. That is, if two forecasts are on the same side of the true value of the risk measure, the closer the forecast is to the risk measure the better, evaluated in terms of the expected score. 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 [29]. 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. This means that 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 $\bar{S}_{R,\pi}(A, F) \geq \bar{S}_{R,\pi}(B, F)$.

Proposition 3.13 *Let the assumptions of Theorem 3.10 (ii) prevail. Then the scoring function $S_{R,\pi}$ defined in (3.13) is \mathcal{M}^d -order-sensitive for R in the sense that for all $A, B \in \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$ and all $F \in \mathcal{M}^d$, we have*

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

Under the assumptions of Theorem 3.10 (iii) and if $\bar{S}_{R,\pi}(B, F) < \infty$, it holds that

$$(A \subsetneq B \subseteq R(F) \text{ or } A \supsetneq B \supseteq R(F)) \implies \bar{S}_{R,\pi}(A, F) > \bar{S}_{R,\pi}(B, F).$$

Proof For the first part, it is enough to show order-sensitivity for the elementary scores $S_{R,k}$ in (3.12). Let $A \subseteq B \subseteq R(F)$. Then for any $k \in \mathbb{R}^d$, (3.11) yields $\bar{S}_{R,k}(A, F) - \bar{S}_{R,k}(B, F) = \mathbb{1}_{B \setminus A}(k)\bar{V}_{R_0}(k, F) \geq 0$. On the other hand, if we have $A \supseteq B \supseteq R(F)$, then $\bar{S}_{R,k}(A, F) - \bar{S}_{R,k}(B, F) = -\mathbb{1}_{A \setminus B}(k)\bar{V}_{R_0}(k, F) \geq 0$ again by (3.11). The second part follows along the lines of the proof of Theorem 3.10 (iii). \square

3.2.5 Forecast dominance and Murphy diagrams

The notion of (strict) consistency implies that—in expectation—a correctly specified forecast will score at most as high as (strictly less than) any misspecified forecast. On the level of the prediction space setting (see Gneiting and Ranjan [38], Strähl and Ziegel [59]), which is also sketched in Sect. 1.2 after (1.2), Holzmann and Euler [41] 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; see Tsyplakov [61]. Patton [56] 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 homogeneity, may guide the decision on what scoring function to use; see Fissler et al. [26, Sect. 4] for results in the current context. For the rare situation when one forecast scores better than another uniformly over all consistent scoring functions, Ehm et al. [19] coined the term *forecast dominance*. We give the corresponding definition here for the situation of exhaustive forecasts for systemic risk measures R .

Definition 3.14 Let $Y \in \mathcal{Y}^d$ and let A, B be two (stochastic) forecasts for some systemic risk measure R of the form in (2.1) which take values in $\widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$. Then A *dominates* B if $\mathbb{E}[S_{R,\pi}(A, Y)] \leq \mathbb{E}[S_{R,\pi}(B, Y)]$ for all consistent scoring functions $S_{R,\pi}$ of the form (3.13), where π is a σ -finite nonnegative measure on $\mathcal{B}(\mathbb{R}^d)$.

Note that the expectations are taken over the joint distribution of the forecasts and the observation, implicitly assuming joint measurability of $S_{R,\pi}$ in both arguments.

Since the scores $S_{R,\pi}$ in (3.13) are parametrised by the class of nonnegative σ -additive measures on $\mathcal{B}(\mathbb{R}^d)$, it is not very handy to check forecast dominance in practice by using the definition. To this end, the following corollary is helpful. The proof is straightforward and therefore omitted.

Corollary 3.15 Let $Y \in \mathcal{Y}^d$ and let A, B be two (stochastic) forecasts for some systemic risk measure R of the form (2.1) which take values in $\widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$. Then A *dominates* B if and only if $\mathbb{E}[S_{R,k}(A, Y)] \leq \mathbb{E}[S_{R,k}(B, Y)]$ for all elementary scores $S_{R,k}$ given in (3.12), where $k \in \mathbb{R}^d$.

Corollary 3.15 opens the way to an immediate multivariate analogue of *Murphy diagrams* considered in Ehm et al. [19]. That is, if A is a $\widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$ -valued forecast of a systemic risk measure R and Y is the corresponding \mathbb{R}^d -valued observation of a financial system, we can consider the map

$$\mathbb{R}^d \ni k \mapsto s_A(k) = \mathbb{E}[S_{R,k}(A, Y)] \tag{3.14}$$

(also referred to as a *Murphy diagram*) as a diagnostic tool. For an empirical setting with observations $Y_1, \dots, Y_N \in \mathbb{R}^d$ and forecasts $A_1, \dots, A_N \in \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$, (3.14) takes the form

$$\mathbb{R}^d \ni k \mapsto \hat{s}_{N,A}(k) = \frac{1}{N} \sum_{t=1}^N S_{R,k}(A_t, Y_t). \tag{3.15}$$

We illustrate the practical usage of Murphy diagrams in Sect. 5.2.

3.2.6 Selective elicibility

Having established the exhaustive elicibility and selective identifiability of systemic risk measures $R: \mathcal{M}^d \rightarrow \mathcal{A}$ defined in (2.1), one may wonder about their selective elicibility. As defined in Fissler et al. [25], a selective scoring function is a map $s: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^*$. It is *strictly \mathcal{M}^d -consistent* for R if $\bar{s}(x, F) \leq \bar{s}(k, F)$ for all $F \in \mathcal{M}^d, k \in \mathbb{R}^d$ and $x \in R(F)$, where equality implies that $k \in R(F)$. R is called *selectively elicitable* on \mathcal{M}^d if there exists a strictly \mathcal{M}^d -consistent selective scoring function for it.

The *mutual exclusivity result* of [25, Theorem 3.7] asserts that under mild conditions, a set-valued functional cannot be both selectively and exhaustively elicitable. One of the conditions for this mutual exclusivity result is that R satisfies the *proper subset property* in the sense that there are $F, G \in \mathcal{M}^d$ such that $\emptyset \neq R(G) \subsetneq R(F)$. In the given context, R satisfies the proper subset property as can be seen by invoking the monotonicity or the cash-invariance of R . Another condition is that for all $\varepsilon \in (0, 1)$, there exists some $\lambda_0 \in (0, \varepsilon)$ such that $(1 - \lambda_0)F + \lambda_0G \in \mathcal{M}^d$. Finally, one needs to impose a certain finiteness assumption. The technical details are summarised in the following theorem.

Theorem 3.16 *Under the conditions of Theorem 3.10 (iii) and additionally assuming the convexity of \mathcal{M}_0^d introduced there, it holds that if there is a strictly \mathcal{M}_0^d -consistent exhaustive scoring function $S_{R,\pi}: \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d) \times \mathbb{R}^d \rightarrow \mathbb{R}^*$ for R such that the expected score $\bar{S}_{R,\pi}(A, F)$ is finite for all $F \in \mathcal{M}_0^d$ and $A \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$, then R fails to be selectively elicitable on \mathcal{M}_0^d .*

Proof The proof follows from [25, Proposition 3.3 and Theorem 3.5]. Note that while [25, Theorem 3.5] uses the finiteness assumption on the expected exhaustive scoring function, [25, Proposition 3.3] does not need such a finiteness assumption on the expected selective scoring function. □

4 Elicitability of systemic risk measures based on expected shortfall

As described in Sect. 1.2, the last decade has seen a lively debate about which risk measure to use in practice. The main focus has been on the dichotomy between the law-invariant measures value at risk (VaR_α) and expected shortfall (ES_α) at probabilistic level $\alpha \in (0, 1)$. Recall that

$$\begin{aligned} \text{VaR}_\alpha(F) &= -\inf\{x \in \mathbb{R}: \alpha \leq F(x)\} \in \mathbb{R}, \\ \text{ES}_\alpha(F) &= \frac{1}{\alpha} \int_0^\alpha \text{VaR}_\beta(F) \, d\beta \\ &= -\frac{1}{\alpha} \int_I x \, dF(x) - \frac{1}{\alpha} \text{VaR}_\alpha(F) \left(F(-\text{VaR}_\alpha(F)) - \alpha \right) \in \mathbb{R}^*, \end{aligned} \tag{4.1}$$

where $I = (-\infty, -\text{VaR}_\alpha(F)]$.

Recall that Theorems 3.1 and 3.10 establish identifiability and elicibility results for systemic risk measures based on a scalar risk measure ρ which is identifiable, and therefore—under weak regularity assumption—elicitable; see Steinwart et al. [58]. Moreover, Proposition 3.9 establishes that under weak regularity conditions, the identifiability of ρ is also necessary for the identifiability and elicibility of the systemic risk measure based on ρ . Therefore, $R^{\text{ES}_\alpha}(Y) = \{k \in \mathbb{R}^d : \text{ES}_\alpha(\Lambda(Y + k)) \leq 0\}$ in general fails to be elicitable. On the other hand, the pair $(\text{VaR}_\alpha, \text{ES}_\alpha)$ is elicitable under weak regularity conditions; see Acerbi and Szekely [1] and Fissler and Ziegel [27]. This might trigger the suspicion that the pair $(R^{\text{VaR}_\alpha}, R^{\text{ES}_\alpha})$ 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 $(R^{\text{VaR}_\alpha}, R^{\text{ES}_\alpha})$ in general fails to have the exhaustive CxLS property for $d \geq 2$, ruling out its exhaustive elicibility. The risk measure $R^{\text{VaR}_\alpha}(Y)$ only encodes information about the sign of $\text{VaR}_\alpha(\Lambda(Y + k))$ for each $k \in \mathbb{R}^d$. Apart from $k \in R_0^{\text{VaR}_\alpha}(Y)$, we know nothing about the actual size of $\text{VaR}_\alpha(\Lambda(Y + k))$. The positive result about the elicibility of the pair $(\text{VaR}_\alpha, \text{ES}_\alpha)$, however, exploits the fact that for the scoring function $S_\alpha(x, y) = -(\mathbb{1}_{\{y \leq -x\}} - \alpha)x/\alpha - \mathbb{1}_{\{y \leq -x\}}y/\alpha$, $x, y \in \mathbb{R}$, $\text{VaR}_\alpha(F)$ is the *minimiser* of the expected score $S_\alpha(x, F)$, while $\text{ES}_\alpha(F)$ is its *minimum*; see Frongillo and Kash [34]. Therefore, we consider the function-valued functional $T^{\text{VaR}_\alpha} : \mathcal{Y}^d \rightarrow \mathbb{R}^{\mathbb{R}^d}$, where for each $Y \in \mathcal{Y}^d$,

$$T^{\text{VaR}_\alpha}(Y) : \mathbb{R}^d \rightarrow \mathbb{R}, \quad \mathbb{R}^d \ni k \mapsto T^{\text{VaR}_\alpha}(Y)(k) = \text{VaR}_\alpha(\Lambda(Y + k)). \tag{4.2}$$

4.1 Identifiability results

Let $\mathcal{M}_{\text{inc,cont}} \subseteq \mathcal{M}$ be the subclass of continuous and strictly increasing distribution functions in \mathcal{M} . Let $\mathcal{M}_{\text{inc,cont}}^d \subseteq \mathcal{M}^d$ be the subclass of distributions such that for any $Y \sim F \in \mathcal{M}_{\text{inc,cont}}^d$, the distribution of $\Lambda(Y + k)$ is in $\mathcal{M}_{\text{inc,cont}}$ for all $k \in \mathbb{R}^d$. A strict $\mathcal{M}_{\text{inc,cont}}$ -identification function $V : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$ for the pair $(\text{VaR}_\alpha, \text{ES}_\alpha) : \mathcal{M} \rightarrow \mathbb{R} \times \mathbb{R}^*$ is given, for $(v, e) \in \mathbb{R}^2$ and $y \in \mathbb{R}$, by

$$V(v, e, y) = \left(\begin{array}{c} \alpha - \mathbb{1}_{\{y+v \leq 0\}} \\ e + \mathbb{1}_{\{y+v \leq 0\}}y/\alpha + (\mathbb{1}_{\{y+v \leq 0\}} - \alpha)v/\alpha \end{array} \right),$$

which can be verified by a straightforward calculation. This function induces a (non-strict) selective \mathcal{M}^d -identification function $U : \mathbb{R}^{\mathbb{R}^d} \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^2$ for the pair $(T^{\text{VaR}_\alpha}, R_0^{\text{ES}_\alpha}) : \mathcal{M}^d \rightarrow \mathbb{R}^{\mathbb{R}^d} \times 2^{\mathbb{R}^d}$. For $v : \mathbb{R}^d \rightarrow \mathbb{R}$, $k \in \mathbb{R}^d$ and $y \in \mathbb{R}^d$, $U(v, k, y)$ is defined as

$$\left(\begin{array}{c} \alpha - \mathbb{1}_{\{\Lambda(y+k) \leq -v(k)\}} \\ \mathbb{1}_{\{\Lambda(y+k) \leq -v(k)\}}\Lambda(y+k)/\alpha + (\mathbb{1}_{\{\Lambda(y+k) \leq -v(k)\}} - \alpha)v(k)/\alpha \end{array} \right). \tag{4.3}$$

Proposition 4.1 *For any $F \in \mathcal{M}^d$, the component U_2 of U defined in (4.3) is oriented in the sense that*

$$\bar{U}_2(T^{\text{VaR}_\alpha}(F), k, F) \begin{cases} < 0, & \text{if } k \notin R^{\text{ES}_\alpha}(F), \\ = 0, & \text{if } k \in R_0^{\text{ES}_\alpha}(F), \\ > 0, & \text{if } k \in R^{\text{ES}_\alpha}(F) \setminus R_0^{\text{ES}_\alpha}(F) \end{cases} \tag{4.4}$$

for any $k \in \mathbb{R}^d$. Moreover, U is a selective $\mathcal{M}_{\text{inc,cont}}^d$ -identification function for the pair $(T^{\text{VaR}_\alpha}, R_0^{\text{ES}_\alpha}) : \mathcal{M}^d \rightarrow \mathbb{R}^{\mathbb{R}^d} \times 2^{\mathbb{R}^d}$.

Proof Let $F \in \mathcal{M}^d$ and $k \in \mathbb{R}^d$. Then $\bar{U}_2(T^{\text{VaR}_\alpha}(F), k, F)$ equals

$$\begin{aligned} & \frac{1}{\alpha} \mathbb{E}_F [\Lambda(Y+k) \mathbb{1}_{\{\Lambda(Y+k) \leq -T^{\text{VaR}_\alpha}(F)(k)\}}] \\ & \quad + \frac{1}{\alpha} \text{VaR}_\alpha(\Lambda(Y+k)) \left(F_{\Lambda(Y+k)}(-\text{VaR}_\alpha(\Lambda(Y+k))) - \alpha \right) \\ & = -\text{ES}_\alpha(\Lambda(Y+k)) \begin{cases} < 0, & \text{if } k \notin R^{\text{ES}_\alpha}(F), \\ = 0, & \text{if } k \in R_0^{\text{ES}_\alpha}(F), \\ > 0, & \text{if } k \in R^{\text{ES}_\alpha}(F) \setminus R_0^{\text{ES}_\alpha}(F), \end{cases} \end{aligned}$$

where $\Lambda(Y+k) \sim F_{\Lambda(Y+k)}$. If $F \in \mathcal{M}_{\text{inc,cont}}^d$, then $\bar{U}_1(T^{\text{VaR}_\alpha}(F), k, F) = 0$. □

4.2 Elicitability results

We introduce the following regularity assumption on T^{VaR_α} defined in (4.2).

Assumption 4.2 The functional $T^{\text{VaR}_\alpha} : \mathcal{M}_{\text{inc,cont}}^d \rightarrow \mathbb{R}^{\mathbb{R}^d}$ takes values in $\mathcal{C}(\mathbb{R}^d; \mathbb{R})$, the space of continuous functions from \mathbb{R}^d to \mathbb{R} .

Clearly, Assumption 4.2 is satisfied if Λ is continuous. For any increasing function $g : \mathbb{R} \rightarrow \mathbb{R}$, introduce $S_{\alpha,g}(x, y) = (\mathbb{1}_{\{y \leq x\}} - \alpha)(g(x) - g(y))$. Recall from Gneiting [36, Theorem 2.6] that $S_{\alpha,g}$ is a nonnegative consistent selective scoring function for the α -quantile. Moreover, if g is strictly increasing, $S_{\alpha,g}$ 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 [36].

Theorem 4.3 (i) Under Assumption 2.2, for every $k \in \mathbb{R}^d$, the function

$$\begin{aligned} S_k : \mathbb{R}^{\mathbb{R}^d} \times \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d) \times \mathbb{R}^d &\rightarrow [0, \infty), \\ S_k(v, A, y) &= -\mathbb{1}_A(k)U_2(v, k, y) - \mathbb{1}_{R^{\text{ES}_\alpha}(y)}(k)\Lambda(y+k) \end{aligned} \tag{4.5}$$

is a nonnegative \mathcal{M}^d -consistent exhaustive scoring function for the pair

$$(T^{\text{VaR}_\alpha}, R^{\text{ES}_\alpha}) : \mathcal{M}^d \rightarrow \mathbb{R}^{\mathbb{R}^d} \times \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d).$$

(ii) Under Assumption 2.2 and if π_1, π_2 are σ -finite nonnegative measures on $\mathcal{B}(\mathbb{R}^d)$, the map

$$\begin{aligned} S_{\pi_1, \pi_2} : \mathbb{R}^{\mathbb{R}^d} \times \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d) \times \mathbb{R}^d &\rightarrow [0, \infty], \\ S_{\pi_1, \pi_2}(v, A, y) &= \int_{\mathbb{R}^d} S_{\alpha, gk}(-v(k), \Lambda(y+k)) \pi_1(dk) \\ &\quad + \int_{\mathbb{R}^d} S_k(v, A, y) \pi_2(dk), \end{aligned} \tag{4.6}$$

where the function $g_k: \mathbb{R} \rightarrow \mathbb{R}$ is increasing and S_k is given in (4.5) for each $k \in \mathbb{R}^d$, is a nonnegative \mathcal{M}^d -consistent exhaustive scoring function for

$$(T^{\text{VaR}_\alpha}, R^{\text{ES}_\alpha}): \mathcal{M}^d \rightarrow \mathbb{R}^{\mathbb{R}^d} \times \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d).$$

(iii) If Assumptions 2.3 and 4.2 hold, if g_k is strictly increasing for all $k \in \mathbb{R}^d$ and if π_1, π_2 are strictly positive, then the restriction of S_{π_1, π_2} defined in (4.6) to $\mathcal{C}(\mathbb{R}^d; \mathbb{R}) \times \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d) \times \mathbb{R}^d$ is a nonnegative strictly $\mathcal{M}_{\text{inc,cont};0}^d$ -consistent exhaustive scoring function for $(T^{\text{VaR}_\alpha}, R^{\text{ES}_\alpha}): \mathcal{M}_{\text{inc,cont};0}^d \rightarrow \mathcal{C}(\mathbb{R}^d; \mathbb{R}) \times \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$, where $\mathcal{M}_{\text{inc,cont};0}^d \subseteq \mathcal{M}_{\text{inc,cont}}^d$ is such that $\bar{S}_{\pi_1, \pi_2}(T^{\text{VaR}_\alpha}(F), R^{\text{ES}_\alpha}(F), F) < \infty$ for all $F \in \mathcal{M}_{\text{inc,cont};0}^d$.

Proof (i) Let $F \in \mathcal{M}^d$, $v \in \mathbb{R}^{\mathbb{R}^d}$, $A \in \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$ and take $v^* = T^{\text{VaR}_\alpha}(F)$, $A^* = R^{\text{ES}_\alpha}(F)$ and $k \in \mathbb{R}^d$. If $\bar{S}_k(v, A, F) = \infty$, there is nothing to show. So we assume that $\bar{S}_k(v, A, F)$ is finite. Consider

$$\begin{aligned} \bar{S}_k(v, A, F) - \bar{S}_k(v^*, A, F) &= \mathbb{1}_A(k) (-\bar{U}_2(v, k, F) + \bar{U}_2(v^*, k, F)) \\ &= \mathbb{1}_A(k) \mathbb{E}_F [S_{\alpha, \text{id}}(-v(k), \Lambda(Y+k)) \\ &\quad - S_{\alpha, \text{id}}(-v^*(k), \Lambda(Y+k))] \\ &\geq 0, \end{aligned}$$

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

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

where the inequality follows from (4.4) and establishes the consistency of S_k . Together with the fact that $S_k(T^{\text{VaR}_\alpha}(\delta_y), R^{\text{ES}_\alpha}(\delta_y), y) = 0$, this implies the nonnegativity of S_k .

(ii) S_{0, π_2} is \mathcal{M}^d -consistent for $(T^{\text{VaR}_\alpha}, R^{\text{ES}_\alpha}): \mathcal{M}^d \rightarrow \mathbb{R}^{\mathbb{R}^d} \times \widehat{\mathcal{P}}(\mathbb{R}^d; \mathbb{R}_+^d)$ due to (i). Since S_{α, g_k} is a consistent selective scoring function for the α -quantile, the assertion follows by invoking Fubini's theorem.

(iii) Let $F \in \mathcal{M}_{\text{inc,cont};0}^d$, $v \in \mathcal{C}(\mathbb{R}^d; \mathbb{R})$, $A \in \mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$ and $v^* = T^{\text{VaR}_\alpha}(F)$, $A^* = R^{\text{ES}_\alpha}(F)$. If $v \neq v^*$, then $K = \{k \in \mathbb{R}^d: v(k) \neq v^*(k)\} \neq \emptyset$ is open. If $\bar{S}_{\pi_1, \pi_2}(v, A, F) = \infty$, there is nothing to show. Otherwise,

$$\begin{aligned} &\mathbb{E}_F [S_{\pi_1, \pi_2}(v, A, Y) - S_{\pi_1, \pi_2}(v^*, A, Y)] \\ &\geq \int_K \mathbb{E}_F [S_{\alpha, g_k}(-v(k), \Lambda(Y+k)) - S_{\alpha, g_k}(-v^*(k), \Lambda(Y+k))] \pi_1(dk) \\ &\quad + \frac{1}{\alpha} \int_{A \cap K} \mathbb{E}_F [S_{\alpha, \text{id}}(-v(k), \Lambda(Y+k)) - S_{\alpha, \text{id}}(-v^*(k), \Lambda(Y+k))] \pi_2(dk) \end{aligned}$$

is positive, where the first integral is strictly positive and the second is nonnegative (and strictly positive if and only if $\pi_2(A \cap K) > 0$).

If $A \neq A^*$, then $\mathbb{E}_F[S_{\pi_1, \pi_2}(v^*, A, Y) - S_{\pi_1, \pi_2}(v^*, A^*, Y)] > 0$, which follows with similar arguments as in the proof of Theorem 3.10 (iii). \square

Theorem 4.3 (ii) suggests that there is again the possibility to consider Murphy diagrams to assess the quality of forecasts for $(T^{\text{VaR}_\alpha}, R^{\text{ES}_\alpha})$ simultaneously over all scoring functions given in (4.6). 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. [19], 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 et al. [65], the measure π_1 only accounts for forecast accuracy in the VaR component. Therefore, if interest focuses on the ES component, it makes sense to set $\pi_1 = 0$ to facilitate the analysis. This implies that one can consider the Murphy diagram $\mathbb{R}^d \ni k \mapsto \mathbb{E}[S_k(v, A, Y)]$ with the elementary scores S_k given in (4.5). The empirical formulation in the spirit of (3.15) is straightforward.

5 Examples and simulations

5.1 Consistency of the exhaustive scoring function for R

In this section, we demonstrate via a simulation study the discrimination ability of the consistent exhaustive scoring functions constructed in Theorem 3.10. We do so in the context of the prediction space setting introduced in Gneiting and Ranjan [38]. This means that we explicitly model the information sets of each forecaster. For the sake of simplicity and following Gneiting et al. [37] and Fissler and Ziegel [30], we choose to consider only one-step-ahead forecasts and prediction–observation sequences that are independent and identically distributed over time. Despite this simplification, there are still a variety of parameters to consider in the simulation study:

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

We confine ourselves to the following choices of these parameters.

(i)–(iii) We work with two different combinations of Y_t and Λ . In both cases, we work with a financial system of $d = 5$ 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 the risk factor μ_t follows a 5-dimensional normal distribution with mean 0, correlations 0.5 and variances 1, and ϵ_t follows a 5-dimensional standard normal distribution. Moreover, μ_t and ϵ_t are independent for all t . Thus conditionally on μ_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 $i, j = 1, \dots, 5, i \neq j$ and 2 otherwise. Following the suggestion of Amini et al. [4], the aggregation function Λ_1 is of the

form $\Lambda_1(Y_t) = (1 - \beta) \sum_{i=1}^d Y_{i,t}^+ - \beta \sum_{i=1}^d Y_{i,t}^-$, and we set $\beta = 0.75$. In this way, both gains and losses influence the value of the aggregation function, but the losses have a higher weight.

(b) We consider an extended model of Eisenberg and Noe [20]; see Feinstein et al. [23]. The participants have liabilities towards each other, and $L_{ij,t}$ represents the nominal liability of participant i towards participant j at time point t , for $i, j = 1, \dots, 5$. 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} and L_{ij} instead of $L_{is,t}$ and $L_{ij,t}$. Moreover, we denote by $\bar{L}_s = \sum_{i=1}^d L_{is}$ the sum of all payments promised to society. The vector Y_t represents the endowments of the participants at time point t . As suggested in [20], 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 Λ_2 corresponds to the sum of all payments society obtains in the clearing process as described in [20], lowered by 90% of the amount promised to society to ensure that the aggregation function can attain both positive and negative values. This ensures that the system is acceptable if ρ applied to the sum of all payments society obtains does not exceed $-0.9\bar{L}_s$; see also [23]. To simulate the endowments Y_t of the participants, we assume that $Y_{it} = (\mu_{it} + \epsilon_{it})^2$ for $i = 1, \dots, 5$ with μ_t and ϵ_t 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 society.

(iv) For ρ , we consider the scalar risk measures VaR_α and ES_α , $\alpha \in (0, 1)$, as well as the expectile-based version of VaR defined as $\text{EVaR}_\tau(X) = -e_\tau(X)$, $\tau \in (0, 1)$, where the expectile e_τ is the unique solution to the equation

$$\tau \mathbb{E}[(X - e_\tau)^+] = (1 - \tau) \mathbb{E}[(X - e_\tau)^-]$$

for $X \in L^1(\Omega, \mathbb{R})$; see Newey and Powell [53]. For the interpretation of expectile-based risk measures in finance, we refer to Bellini and Di Bernardino [10]; for a novel economic angle, see Ehm et al. [19]. Using the standard identification functions for VaR_α and EVaR_τ from Gneiting [35] and the explicit construction in (3.1), the selective identification function for R_0 is given by $V_{R_0}(k, y) = \alpha - \mathbb{1}_{\{\Lambda(k+y) \leq 0\}}$ for $\rho = \text{VaR}_\alpha$, and by $V_{R_0}(k, y) = \tau(\Lambda(k+y))^+ - (1 - \tau)(\Lambda(k+y))^-$ for $\rho = \text{EVaR}_\tau$. As mentioned in Sect. 4, if the focus is on ES_α , we can and do set the measure π_1 to 0 in (4.6), meaning that no particular choice of a scoring or identification function need be made for this risk measure.

(v) We consider two ideal forecasters with different information sets. Anne has access to the risk factor μ_t and uses the correct conditional distribution of Y_t given μ_t for her predictions. That is, for $t = 1, \dots, N$, she issues forecasts

$$A_t = R(\mathcal{N}_5(\mu_t, I_5)) = R(\mathcal{N}_5(0_5, I_5)) - \mu_t \quad \text{in case (a),}$$

$$A_t = R(\mathcal{N}_5(\mu_t, I_5)^2) \quad \text{in case (b).}$$

Here, $\mathcal{N}_d(m, \Sigma)^2$ denotes the distribution of a random vector $(X_1^2, \dots, X_d^2)^\top$ with $(X_1, \dots, X_d)^\top \sim \mathcal{N}_d(m, \Sigma)$. On the other hand, Bob does not use information about

the risk factor μ_t . Therefore he uses the correct unconditional distribution of Y_t for his forecasts and predicts, for $t = 1, \dots, N$,

$$\begin{aligned}
 B_t &= R(\mathcal{N}(0_5, \Sigma)) && \text{in case (a),} \\
 B_t &= R(\mathcal{N}(0_5, \Sigma)^2) && \text{in case (b).}
 \end{aligned}$$

(vi) We choose π to be a 5-dimensional Gaussian measure with mean $m \in \mathbb{R}^5$ and covariance I_5 . We also set $\pi_2 = \pi$ in (4.6). To enhance the discrimination ability of the score $S_{R,\pi}$, we choose m close to the boundary of $R(Y_t)$. Here we work with $m = 2 \times \mathbf{1}$ as this value appears to be fairly close to Bob’s unconditional forecasts in all cases. This choice of π turns out to be beneficial with respect to integrability considerations and renders our scores finite. Indeed, since V_{R_0} for $\rho = \text{VaR}_\alpha$ is bounded, it is $\pi \otimes F$ -integrable for any finite measure π . In the case of $\rho = \text{EVaR}_\tau$, more considerations are necessary. From the construction of Λ_2 , it is clear that it is a bounded function; in particular, the values lie in the interval $[-0.9\bar{L}_S, \sum_{i=1}^d L_{iS} - 0.9\bar{L}_S]$. 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 π . Finally, since Λ_1 only grows linearly and both π and Y_t are Gaussian, the integrability is also guaranteed in this case. Similar considerations confirm that the sufficient integrability conditions are also satisfied for $\rho = \text{ES}_\alpha$.

(vii) We work with sample sizes $N = 250$, a good proxy for the number of working (and trading) days in a year. In a financial context, the two forecasters Anne and Bob could be both portfolio managers or regulators. While Λ_1 would be an appropriate aggregation function if Anne and Bob are portfolio managers, Λ_2 would be more appropriate for regulators. The choices of including or omitting information contained in the risk factor μ_t might stem from different data access, or deliberate different choices of risk factors. From a different angle, Bob might deliberately choose to use the unconditional profit and loss distribution to come up with a more prudent risk measurement approach than Anne who uses the conditional distribution; see McNeil et al. [51, Sect. 9.2.1].

To compare Anne’s with Bob’s forecast performance, we employ the classical Diebold–Mariano test [17] based on the scoring functions $S_{R,\pi}$ of the form (3.13) and (4.6). We repeat the experiment 1000 times for setting (a) and 100 times for setting (b) since due to the presence of clearing, the computation time tends to be quite lengthy in setting (b). We approximate π with a Monte Carlo draw of size 100 000. The computations are performed with the statistics software R, and in particular its Rcpp package to also integrate parts of C++ code to enhance the computational speed.

We consider tests with two different one-sided null hypotheses. The null hypothesis $H_0: \mathbb{E}[S_{R,\pi}(A_1, Y_1)] \geq \mathbb{E}[S_{R,\pi}(B_1, Y_1)]$, or in short $H_0: A \geq B$, means that Bob has a better forecast performance than Anne on average, evaluated in terms of $S_{R,\pi}$. On the other hand, the null hypothesis $H_0: \mathbb{E}[S_{R,\pi}(A_1, Y_1)] \leq \mathbb{E}[S_{R,\pi}(B_1, Y_1)]$ or $H_0: A \leq B$ stands for asserting that Anne’s forecasts are superior to Bob’s on average, in terms of $S_{R,\pi}$. In Table 1, we report the relative frequencies of rejections for the respective null hypotheses. Invoking the sensitivity of consistent scoring functions with respect to increasing information sets established in Holzmann and Eulert

Table 1 Ratios of rejections of the null hypotheses at significance level 0.05

	H_0	VaR _{0,01}	VaR _{0,05}	ES _{0,01}	ES _{0,05}	EVaR _{0,01}	EVaR _{0,05}
Λ_1	$A \geq B$	0.995	0.940	0.948	1.000	1.000	1.000
	$A \leq B$	0.000	0.000	0.000	0.000	0.000	0.000
Λ_2	$A \geq B$	0.740	0.870	0.670	0.770	0.790	0.900
	$A \leq B$	0.000	0.000	0.000	0.000	0.000	0.000

[41], we expect that Anne’s forecasts are deemed superior to Bob’s predictions. And in fact, the null $A \leq B$ is never rejected for either scenario, while $A \geq B$ is rejected in between 67% and 100% of all experiments over the various scenarios. In particular, with rejection rates for $H_0: A \geq B$ between 0.94 and 1, we observe that the discrimination ability between Bob and Anne is considerably higher for model (a) than for (b), where we obtain rejection rates ranging from 0.67 to 0.90. This might be due to the fact that Λ_1 is unbounded whereas Λ_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, for fixed probability level $\alpha = \tau$, both in case (a) and (b), the number of instances when Anne’s forecasts are preferred to Bob’s is the highest for $\rho = \text{EVaR}_\alpha$ and the lowest for $\rho = \text{ES}_\alpha$, with the exception of $\text{ES}_{0,05}$ in case (a). This ordering might be in line with the observation that a normally distributed X is deemed differently risky with respect to the three risk measures considered when evaluated at the same probability level: $\text{EVaR}_\alpha(X) \leq \text{VaR}_\alpha(X) \leq \text{ES}_\alpha(X)$ for $\alpha \in (0, 0.5)$, see Nolde and Ziegel [54, Table 1].

5.2 Murphy diagrams and comparative backtests

We illustrate the use of Murphy diagrams, following Corollary 3.15. We work within case (a) of Sect. 5.1 with $d = 2$. In particular, we have $Y_t = \mu_t + \epsilon_t$, where μ_t follows a 2-dimensional normal distribution with mean 0, variances 1 and correlations 0.5, and ϵ_t follows a 2-dimensional standard normal distribution. As the scalar risk measure ρ , we only consider $\text{VaR}_{0,05}$, and we use the aggregation function $\Lambda_1: \mathbb{R}^2 \rightarrow \mathbb{R}$, i.e., $\Lambda(x) = 0.25(x_1^+ + x_2^+) - 0.75(x_1^- + x_2^-)$. We should like to combine this with an illustration of the Murphy diagrams in the context of comparative backtests as introduced in Fissler et al. [31] and further developed in Nolde and Ziegel [54]. To this end, suppose Bob’s ideal unconditional risk measure forecasts $B_t = R(\mathcal{N}_2(0_2, \Sigma))$ play the role of the regulator’s standardised procedure. Then we have two internal risk measurement procedures making use of the additional risk factor given by μ_t . The first are Anne’s ideal forecasts $A_t = R(\mathcal{N}_2(\mu_t, I_2)) = R(\mathcal{N}_2(0, I_2)) - \mu_t$. The second are Celia’s forecasts. Confused about different sign conventions in the literature, she issues sign-reversed forecasts C_t assuming that $Y_t \sim \mathcal{N}_2(-\mu_t, I_2)$, resulting in the forecasts $C_t = R(\mathcal{N}_2(-\mu_t, I_2)) = R(\mathcal{N}_2(0, I_2)) + \mu_t$. Again, we consider a time horizon of $N = 250$.

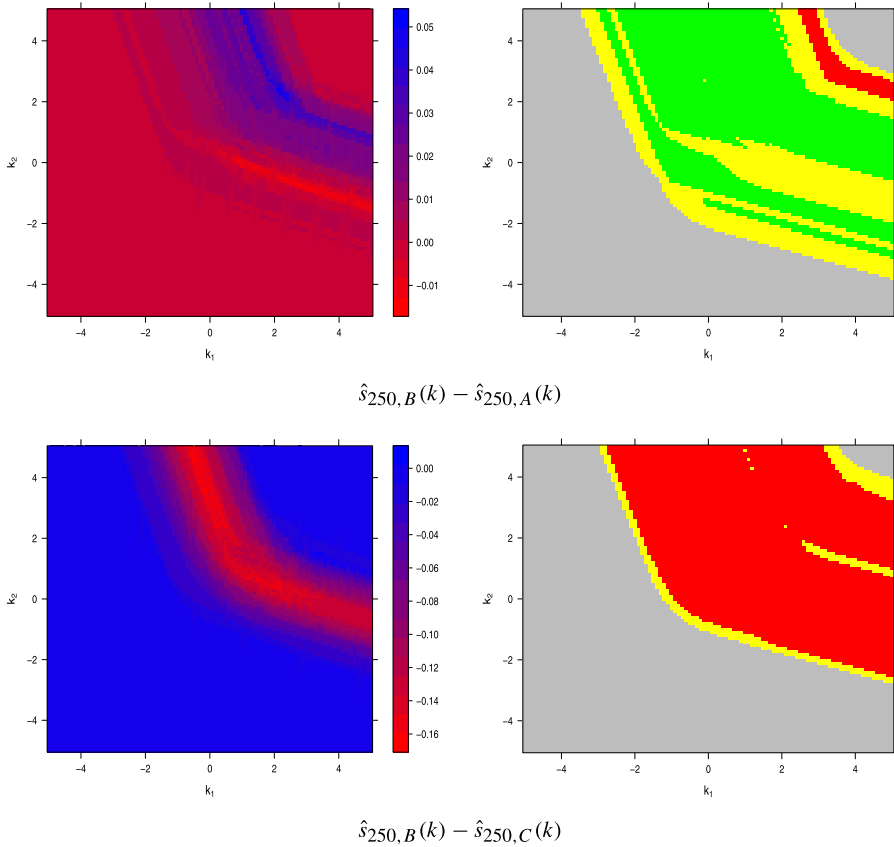


Fig. 3 (Left) Differences $\hat{s}_{250,B}(k) - \hat{s}_{250,f}(k)$ of empirical Murphy diagrams from (3.15) versus $k \in \mathbb{R}^2$ for $f = A$ (top) and $f = C$ (bottom). (Right) Three-zone traffic light illustration of pointwise comparative backtests following Fissler et al. [31]. The green area corresponds to the region where the null $H_0^+ : B \leq f$ is rejected, the red one is where $H_0^- : B \geq f$ is rejected, at level 0.05, respectively. Yellow means that neither H_0^+ nor H_0^- is rejected. In the grey region, the two Murphy diagrams coincide identically

In the left panel of Fig. 3, we illustrate the differences of empirical Murphy diagrams,

$$[-5, 5]^2 \ni k \mapsto \hat{s}_{250,B}(k) - \hat{s}_{250,f}(k) = \frac{1}{250} \sum_{t=1}^{250} (S_{R,k}(B_t, Y_t) - S_{R,k}(f_t, Y_t)),$$

where $f_t \in \{A_t, C_t\}$ stands for either of the two internal procedures produced by Anne or Celia. For the comparison of Bob vs. Anne, we can see that the Murphy diagram $\hat{s}_{250,B}(k) - \hat{s}_{250,A}(k)$ is mostly nonnegative, indicating the superiority of Anne’s more informed forecast over Bob’s prudential unconditional one. On the other hand, comparing Bob vs. Celia, the Murphy diagram $\hat{s}_{250,B}(k) - \hat{s}_{250,C}(k)$ is nonpositive everywhere. This means that all elementary scores indicate the inferiority of Celia’s sign-reversed forecasts. The regions with the highest score differences (in absolute

values) seem to correspond to a blurred version of the boundary of the considered risk measure forecasts. Quite intuitively, the magnitude of the score difference with a maximum of approximately 0.05 is smaller when comparing the two ideal forecasts issued by Bob and Anne in comparison to the situation involving the sign-reversed Celia, where the maximal difference between the Murphy diagrams is in magnitude larger than 0.15. We have performed this experiment several times and observed that the stylised facts are qualitatively stable, though there is still recognisable sample variation present. For transparency reasons, we have depicted the first experiment performed, reporting some more experiments in [26, Supplementary Material].

In the right panel of Fig. 3, we depict the results of pointwise comparative backtests using the traffic-light illustration suggested in [31], which is the analogue for comparative backtests to the three-zone approach for traditional backtests in [7, Appendix B, Sect. III]. In other words, we perform two Diebold–Mariano tests using the elementary score $S_{R,k}$ for each k in a grid of $[-5, 5]^2$ at a significance level of 0.05. The grid cell is coloured in green if the conservative null hypothesis $H_0^+ : B \leq f$ is rejected. This means that the internal procedure f performs significantly better than Bob's standardised procedure in terms of $S_{R,k}$. In contrast, the grid cell is depicted in red if the null $H_0^- : B \geq f$ is rejected, which indicates that the standardised procedure is significantly superior to the internal risk measurement procedure. For all k in the yellow region, none of the two nulls is rejected, meaning that the procedure is indecisive at significance level 0.05. Finally, the grey area corresponds to those points where the score difference is constantly zero for all $t = 1, \dots, N$. Due to the vanishing variance, a Diebold–Mariano test is 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 Fig. 3. For both pairwise comparisons and for k close to the four corners of the area $[-5, 5]^2$, the score differences vanish identically, resulting in a grey colouration. Again, in both cases, there is a 'continuous' behaviour in that the grey region adjoins a yellow stripe before turning into a fairly broad green or red stripe. For the comparison involving Celia, it is reassuring that a substantial region is coloured in red. In this region, the procedure is decisive, deeming Celia significantly inferior to Bob. Moreover, for this particular simulation, there is no green region. The situation comparing the two ideal forecasters Anne and Bob is somewhat more involved. Most points k are coloured in green or yellow/grey, so that either Anne's more informed forecasts are deemed significantly superior to Bob's standardised forecasts or the procedure is indecisive between the two methods. However, 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. 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 [26, Supplementary Material] 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 [41]), but 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 located roughly in a similar area.

6 Summary and discussion

We briefly summarise the various elicibility and identifiability results established in this paper for the set-valued systemic risk measure R defined in (2.1) and its derived quantities. If the underlying scalar risk measure ρ is identifiable, R_0 of the form (2.3) is selectively identifiable. Using a somewhat generalised concept of identifiability and exploiting an orientation argument, this implies that R itself is selectively identifiable. Under further regularity conditions, it even leads to the selective identifiability of EAR_w introduced in (2.5), with function-valued identification functions. On the other hand, the selective identifiability of R_0 together with an orientation argument leads to the exhaustive elicibility of R under appropriate conditions. Moreover, since (under mild conditions) there is a one-to-one relationship between R and R_0 , this also implies the exhaustive elicibility of R_0 . In view of the mutual exclusivity result in Fissler et al. [25], the exhaustive elicibility of R rules out its selective elicibility in the sense of [25]; see Theorem 3.16. The findings of [25] plus the fact that the functional EAR_w possesses the selective CxLS* property imply that EAR_w cannot be exhaustively elicitable; the question of selective elicibility remains open in this case.

Following the idea of joint elicibility that led to the remedy for the non-elicitability problem of expected shortfall, Sect. 4 presents a way of achieving selective identifiability and exhaustive elicibility of systemic risk measures based on ES_α , by using a different functional based on VaR_α .

The identifiability and elicibility results presented in this paper open the way to traditional and comparative backtests as described in Nolde and Ziegel [54], employing Diebold–Mariano tests and calibration tests as demonstrated in Sect. 5. This might be interesting both in a regulatory framework and for internal risk assessment of companies with different units. In an even more statistical direction, one can employ strictly consistent scores to perform M -estimation for set-valued systemic risk measures. This would require an optimisation over a collection of subsets, such as $\mathcal{F}(\mathbb{R}^d; \mathbb{R}_+^d)$, which is computationally challenging. This might be alleviated by transitioning to a parametric (auto-)regression framework for set-valued systemic risk measures, see Dimitriadis and Bayer [18], so that the optimisation can be performed over a finite-dimensional parameter space.

Acknowledgements We thank two anonymous referees and the Editor Martin Schweizer for their careful assessment and constructive feedback, which improved the quality of the paper and made it more accessible. We should like to express our sincere gratitude to Tilmann Gneiting and Johanna Ziegel for insightful discussions and persistent encouragement. We are indebted to Timo Dimitriadis and to Peter Barančok who provided helpful comments in the context of equivariant scores, to Yuan Li for his careful proofreading of an earlier version of this paper, and to Lukáš Šablica for helpful programming advice on the simulation part of this project.

Funding Note Open Access funding provided by Vienna University of Economics and Business (WU).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons

licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Acerbi, C., Szekely, B.: Backtesting expected shortfall. *Risk Mag.* **December**, 76–81 (2014)
2. Acharya, V.V., Pedersen, L.H., Philippon, T., Richardson, M.: Measuring systemic risk. *Rev. Financ. Stud.* **30**, 2–47 (2016)
3. Adrian, T., Brunnermeier, M.K.: CoVaR. *Am. Econ. Rev.* **106**, 1705–1741 (2016)
4. Amini, H., Filipovic, D., Minca, A.: Systemic risk and central clearing counterparty design. *SIAM J. Financ. Math.* **11**, 60–98 (2020)
5. Armenti, Y., Crépey, S., Drapeau, S., Papapantoleon, A.: Multivariate shortfall risk allocation and systemic risk. *SIAM J. Financ. Math.* **9**, 90–126 (2018)
6. Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. *Math. Finance* **9**, 203–228 (1999)
7. Bank for International Settlements. Consultative Document: Fundamental review of the trading book: a revised market risk framework. Basel Committee on Banking Supervision (2013). Available online at <https://www.bis.org/publ/bcbs265.pdf>
8. Bank for International Settlements. Consultative Document: Fundamental review of the trading book: Outstanding issues. Basel Committee on Banking Supervision (2014). Available online at <https://www.bis.org/bcbs/publ/d305.htm>
9. Bellini, F., Bignozzi, V.: On elicitable risk measures. *Quant. Finance* **15**, 725–733 (2015)
10. Bellini, F., Di, E.: Bernardino. Risk management with expectiles. *Eur. J. Finance* **23**, 487–506 (2017)
11. Biagini, F., Fouque, J.-P., Frittelli, M., Meyer-Brandis, T.: A unified approach to systemic risk measures via acceptance sets. *Math. Finance* **29**, 329–367 (2019)
12. Bignozzi, W., Burzoni, M., Munari, C.: Risk measures based on benchmark loss distributions. *J. Risk Insur.* **87**, 437–475 (2020)
13. Chen, C., Iyengar, G., Moallemi, C.C.: An axiomatic approach to systemic risk. *Manag. Sci.* **59**, 1373–1388 (2013)
14. Cont, R., Deguest, R., Scandolo, G.: Robustness and sensitivity analysis of risk measurement procedures. *Quant. Finance* **10**, 593–606 (2010)
15. Davis, M.H.A.: Verification of internal risk measure estimates. *Stat. Risk. Model.* **33**, 67–93 (2016)
16. Dawid, P.: 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**, 534–535 (2016)
17. Diebold, F.X., Mariano, R.S.: Comparing predictive accuracy. *J. Bus. Econ. Stat.* **13**, 253–263 (1995)
18. Dimitriadis, T., Bayer, S.: A joint quantile and expected shortfall regression framework. *Electron. J. Stat.* **13**, 1823–1871 (2019)
19. Ehm, W., Gneiting, T., Jordan, A., Krüger, F.: Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **78**, 505–533 (2016)
20. Eisenberg, L., Noe, T.H.: Systemic risk in financial networks. *Manag. Sci.* **47**, 236–249 (2001)
21. Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., Beleraj, A.: An academic response to Basel 3.5. *Risks* **2**(1), 25–48 (2014)
22. Emmer, S., Kratz, M., Tasche, D.: What is the best risk measure in practice? A comparison of standard risk measures. *J. Risk* **8**, 31–60 (2015)
23. Feinstein, Z., Rudloff, B., Weber, S.: Measures of systemic risk. *SIAM J. Financ. Math.* **8**, 672–708 (2017)
24. Fissler, T.: On Higher Order Elicitability and Some Limit Theorems on the Poisson and Wiener Space. PhD Thesis, University of Bern (2017). Available online at http://biblio.unibe.ch/download/eldiss/17fissler_t.pdf

25. Fissler, T., Frongillo, R., Hlavinová, J., Rudloff, B.: Forecast evaluation of set-valued functionals. (2020). Preprint, available online at [arXiv:1910.07912v2](https://arxiv.org/abs/1910.07912v2)
26. Fissler, T., Hlavinová, J., Rudloff, B.: Elicitability and identifiability of systemic risk measures (2019). Preprint, available online at [arXiv:1907.01306v2](https://arxiv.org/abs/1907.01306v2)
27. Fissler, T., Ziegel, J.F.: Higher order elicibility and Osband's principle. *Ann. Stat.* **44**, 1680–1707 (2016)
28. Fissler, T., Ziegel, J.F.: Erratum: Higher order elicibility and Osband's principle. *Ann. Stat.* (2021), forthcoming. <https://doi.org/10.1214/20-AOS2014>. Available online at [arXiv:1901.08826v2](https://arxiv.org/abs/1901.08826v2)
29. Fissler, T., Ziegel, J.F.: Order-sensitivity and equivariance of scoring functions. *Electron. J. Stat.* **13**, 1166–1211 (2019)
30. Fissler, T., Ziegel, J.F.: Evaluating range value at risk forecasts (2020). Preprint, available online at [arXiv:1902.04489](https://arxiv.org/abs/1902.04489)
31. Fissler, T., Ziegel, J.F., Gneiting, T.: Expected shortfall is jointly elicitable with value-at-risk: implications for backtesting. *Risk Mag.* **January**, 58–61 (2016)
32. Föllmer, H., Schied, A.: Convex measures of risk and trading constraints. *Finance Stoch.* **6**, 429–447 (2002)
33. Föllmer, H., Schied, A.: *Stochastic Finance: An Introduction in Discrete Time*, 4th edn. de Gruyter, Berlin (2016)
34. Frongillo, R., Kash, I.: On elicitation complexity. In: Cortes, C., et al. (eds.) *Advances in Neural Information Processing Systems*, vol. 28, pp. 3258–3266. Curran Associates, Red Hook (2015)
35. Gneiting, T.: Making and evaluating point forecasts. *J. Am. Stat. Assoc.* **106**, 746–762 (2011)
36. Gneiting, T.: Quantiles as optimal point forecasts. *Int. J. Forecast.* **27**, 197–207 (2011)
37. Gneiting, T., Balabdaoui, F., Raftery, A.E.: Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **69**, 243–268 (2007)
38. Gneiting, T., Ranjan, R.: Combining predictive distributions. *Electron. J. Stat.* **7**, 1747–1782 (2013)
39. Hampel, F.R.: A general qualitative definition of robustness. *Ann. Math. Stat.* **42**, 1887–1896 (1971)
40. Hoffmann, H., Meyer-Brandis, T., Svindland, G.: Risk-consistent conditional systemic risk measures. *Stoch. Process. Appl.* **126**, 2014–2037 (2016)
41. Holzmann, H., Eulert, M.: The role of the information set for forecasting – with applications to risk management. *Ann. Appl. Stat.* **8**, 79–83 (2014)
42. Huber, P.J., Ronchetti, E.M.: *Robust Statistics*, 2nd edn. Wiley, Hoboken (2009)
43. Jordan, A., Mühlemann, A., Ziegel, J.F.: Optimal solutions to the isotonic regression problem (2019). Preprint, available online at [arXiv:1904.04761](https://arxiv.org/abs/1904.04761)
44. Jouini, E., Schachermayer, W., Touzi, N.: Law invariant risk measures have the Fatou property. In: Kusuoka, S., Yamazaki, A. (eds.) *Adv. Math. Econ.*, vol. 9, pp. 49–72. Springer, Tokyo (2006)
45. Koenker, R.: *Quantile Regression*. Cambridge University Press, Cambridge (2005)
46. Koenker, R., Bassett, G.: Regression quantiles. *Econometrica* **46**, 33–50 (1978)
47. Krättschmer, V., Schied, A., Zähle, H.: Comparative and qualitative robustness for law-invariant risk measures. *Finance Stoch.* **18**, 271–295 (2014)
48. Kromer, E., Overbeck, L., Zilch, K.: Systemic risk measures on general measurable spaces. *Math. Methods Oper. Res.* **84**, 323–357 (2016)
49. Lambert, N.: Elicitation and evaluation of statistical forecasts (2019). Preprint, available online at <http://ai.stanford.edu/~nlambert/papers/elicibility.pdf>
50. Lambert, N., Pennock, D.M., Shoham, Y.: Eliciting properties of probability distributions. In: *Proceedings of the 9th ACM Conference on Electronic Commerce*, pp. 129–138. Association for Computing Machinery, Chicago (2008)
51. McNeil, A.J., Frey, R., Embrechts, P.: *Quantitative Risk Management: Concepts, Techniques and Tools*, 2nd edn. Princeton University Press, Princeton (2015)
52. Nau, R.F.: Should scoring rules be 'effective'? *Manag. Sci.* **31**, 527–535 (1985)
53. Newey, W.K., Powell, J.L.: Asymmetric least squares estimation and testing. *Econometrica* **55**, 819–847 (1987)
54. Nolde, N., Ziegel, J.F.: Elicitability and backtesting: perspectives for banking regulation. *Ann. Appl. Stat.* **11**(1833–1874), 12 (2017)
55. Osband, K.H.: *Providing Incentives for Better Cost Forecasting*. PhD Thesis, University of California, Berkeley (1985). Available online at <https://doi.org/10.5281/zenodo.4355667>.
56. Patton, A.J.: Comparing possibly misspecified forecasts. *J. Bus. Econ. Stat.* **38**, 796–809 (2020)
57. Rogers, L.C.G., Veraart, L.A.M.: Failure and rescue in an interbank network. *Manag. Sci.* **59**, 882–898 (2013)

58. Steinwart, I., Pasin, C., Williamson, R., Zhang, S.: Elicitation and identification of properties. In: Balcan, M.F., et al. (eds.) Proceedings of the 27th Conference on Learning Theory. Proceedings of Machine Learning Research, vol. 35, pp. 482–526 (2014)
59. Strähl, Ch., Ziegel, J.F.: Cross-calibration of probabilistic forecasts. *Electron. J. Stat.* **11**, 608–639 (2017)
60. Svindland, G.: Continuity properties of law-invariant (quasi-)convex risk functions on L^∞ . *Math. Financ. Econ.* **3**, 39–43 (2010)
61. Tsyplakov, A.: Theoretical guidelines for a partially informed forecast examiner (2014). Working paper, available online at https://mpra.ub.uni-muenchen.de/67333/1/MPRA_paper_67333.pdf
62. Weber, S.: Distribution-invariant risk measures, information and dynamic consistency. *Math. Finance* **16**, 419–441 (2006)
63. Ziegel, J.F.: Coherence and elicibility. *Math. Finance* **26**, 901–918 (2016)
64. Ziegel, J.F.: 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**, 555–556 (2016)
65. Ziegel, J.F., Krüger, F., Jordan, A., Fasciati, F.: Robust forecast evaluation of expected shortfall. *J. Financ. Econom.* **18**, 95–120 (2020)