Localizing Strictly Proper Scoring Rules*

Ramon F. A. de Punder
Department of Quantitative Economics
University of Amsterdam and Tinbergen Institute

Cees G. H. Diks[†]
Department of Quantitative Economics
University of Amsterdam and Tinbergen Institute

Roger J. A. Laeven
Department of Quantitative Economics
University of Amsterdam, CentER and EURANDOM

Dick J. C. van Dijk
Department of Econometrics
Erasmus University Rotterdam and Tinbergen Institute

May 26, 2025

---

*We are very grateful to Yacine Ait-Sahalia, Timo Dimitriadis, Tilmann Gneiting, Andrew Harvey, Alexander Jordan, Frank Kleibergen, Siem Jan Koopman, Rutger-Jan Lange, Sebastian Lerch, Xiaochun Meng, Marc-Oliver Pohle, Johannes Resin, Johanna Ziegel and participants at various seminars and conferences, including at the Heidelberg Institute for Theoretical Studies, Tinbergen Institute, University of Copenhagen, the 42nd International Symposium on Forecasting in Oxford (July 2022), the 10th International Workshop on Applied Probability in Thessaloniki (June 2023), the 5th Quantitative Finance and Financial Econometrics International Conference in Marseille (June 2023), the 12th ECB Conference on Forecasting Techniques in Frankfurt (June 2023), the 16th Meeting of the Netherlands Econometric Study Group in Rotterdam (June 2023), the International Association for Applied Econometrics Annual Conference in Oslo (June 2023), the 11th World Congress in Probability and Statistics in Bochum (August 2024) and the 76th European Meeting of the Econometric Society (August 2024), for their comments and suggestions. This research was supported in part by the Netherlands Organization for Scientific Research under grant NWO Vici 2020–2027 (Laeven).

[†] Corresponding author. Mailing Address: PO Box 15867, 1001 NJ Amsterdam, The Netherlands. Phone: +31 (0) 20 525 4252. Email: `C.G.H.Diks@uva.nl`.

Abstract

When comparing predictive distributions, forecasters are typically not equally interested in all regions of the outcome space. To address the demand for focused forecast evaluation, we propose a procedure to transform strictly proper scoring rules into their localized counterparts while preserving the score divergence and strict propriety. This is accomplished by applying the original scoring rule to a censored distribution. Our procedure nests the censored likelihood score as a special case. Among a multitude of others, it also implies a class of censored kernel scores that offers a (possibly multivariate) alternative to the threshold weighted Continuously Ranked Probability Score (twCRPS), extending its local propriety to more general weight functions than single tail indicators. Within this localized framework, we obtain a generalization of the Neyman Pearson lemma, establishing the censored likelihood ratio test as uniformly most powerful. For other tests of localized equal predictive performance, results of Monte Carlo simulations and empirical applications to risk management, inflation and climate data consistently emphasize the excellent power properties of censoring versus other localization methods.

Keywords: Density forecast evaluation; Tests for equal predictive ability; Censoring; Likelihood ratio; CRPS.

1 INTRODUCTION

Over the past decades, probabilistic forecasts have garnered increasing attention across a variety of disciplines, primarily because they provide a more comprehensive understanding of the stochastic nature of a random variable under scrutiny than point forecasts (Dawid 1984). A cornerstone for the effective evaluation of such probabilistic forecasts is the use of strictly proper scoring rules (Gneiting and Raftery 2007; Brehmer and Gneiting 2020; Patton 2020), which have been widely advocated for their ability to ensure fair comparative assessments of different forecast methods. Scoring rules are inherently connected with divergence measures; under the restriction of strict propriety, these measures are subsumed under Bregman divergences (Dawid 2007; Ovcharov 2018; Painsky and Wornell 2020). While the usefulness of unweighted probabilistic forecasting is well-recognized and well-understood, various applications, such as the analysis of large financial portfolio losses, inflation targets or temperature ranges, require a focused, localized evaluation of predictive distributions.

In this paper, we introduce a natural localization mechanism for strictly proper scoring rules that preserves the score divergence and strict propriety. By censoring (Bernoulli 1760; Tobin 1958) the observation and distribution before applying the original scoring rule, we find a sweet spot between retaining and discarding information when focusing on a region of interest. Crucially, unlike existing approaches that employ conditional distributions, our method preserves the overall probability of receiving an observation in (or outside) the target region, obviously relevant when comparing various candidate distributions focused on the same area. Moreover, within the region of interest, our mechanism maintains the original distribution's shape. This is particularly beneficial when evaluating functionals in this region, such as quantiles or conditional expectations. Our procedure can be used to generate a multitude of strictly locally proper scoring rules. These include the censored likelihood (CSL) score, proposed by Diks et al. (2011), and the threshold weighted Continuously Ranked Probability Score (twCRPS), put forward by Gneiting and Ranjan (2011), for weight functions for which Holzmann and Klar (2017a) have shown that the twCRPS is strictly locally proper. On the other hand, for weight functions for which the twCRPS is not strictly locally proper, our analysis provides a strictly locally proper alternative.

The information retained by our censoring approach translates into advantageous power properties of tests aimed at comparing density forecasts in regions of interest. We prove a generalization of the Neyman Pearson (1933) lemma, revealing that the censored likelihood ratio leads to a Uniformly Most Powerful (UMP) test. By contrast, we provide explicit evidence that the conditional likelihood (CL) score does not admit a UMP test. Monte Carlo simulations and empirical applications analyze the power properties of the Diebold and Mariano (2002) (DM) type test statistic, within the framework of Giacomini and White (2006), based on censored vis-à-vis alternative localized scoring rules. Censored scoring rules have competitive power properties in all Monte Carlo experiments conducted. In multiple empirical experiments, involving financial, macroeconomic and climate data, we utilize the DM tests in the Model Confidence Set (MCS) procedure of Hansen et al. (2011). The MCSs resulting from censored scoring rules are typically smaller than those arising from alternative localization procedures, broadly aligning with the power properties displayed by the Monte Carlo results.

Our research contributes to the literature on focused scoring rules, initiated by the weighted likelihood score (WLS) of Amisano and Giacomini (2007). Diks et al. (2011) and Gneiting and Ranjan (2011) sought to correct the (regular) impropriety of this scoring rule by introducing the CL, CSL and twCRPS, respectively. Holzmann and Klar (2017a) substantially advanced focused scoring rules, using conditioning to construct proportionally locally proper scoring rules from unweighted scoring rules other than the logarithmic score. They also showed that strict local propriety of the ensuing scoring rules can be restored by adding an auxiliary weighted scoring rule, based on an arbitrary strictly proper rule for the probability of an observation landing in the region of interest. Our work differs importantly by opting for censoring rather than conditioning as localization mechanism. Through censoring, we enable the direct application of the original scoring rule to the localized measure, thereby avoiding the need for an auxiliary scoring rule and preserving the original Bregman divergence. As detailed by Brehmer and Gneiting (2020, Theorem 1), the conditional scoring rules of Holzmann and Klar (2017a) can also be viewed as an extension of the WLS refined through a 'properization' process. Consequently, properization is not a viable mechanism for retaining strict propriety of the original scoring rule. Recently, Allen et al. (2023) introduced a framework to create strictly locally proper scoring rules from the class of kernel scores. Furthermore, Mitchell and Weale (2023) proposed using censored density forecasts, to perform statistical inference based on a central region of the forecast distribution; they do *not* aim to evaluate competing candidate densities. We provide a detailed comparison of our censoring approach with these alternative localization procedures in Section 3.5.

Our research also rests upon a substantial body of research concerning unweighted strictly proper scoring rules and their associated divergence measures. Although the formalization of strict propriety was rigorously achieved by Gneiting and Raftery (2007), scoring rules satisfying this property date at least to the Quadratic Scoring rule of Brier (1950). The literature in this domain has evolved from an initial focus on discrete settings to a more general treatment. In this vein, we rely on the expanded frameworks of the Power (PowS_α) and PseudoSpherical (PsSphS_α) families as advocated by Gneiting and Raftery (2007) and Ovcharov (2018) rather than their discrete foundations.

Interest in targeting specific regions of predictive distributions has surged across diverse fields, including meteorology, climatology, hydrology, finance, and economics. In financial risk management, attention is particularly concentrated on the left tail of return distributions, according to mandated risk measures such as Value-at-Risk and Expected Shortfall (Cont et al. 2010; Fissler et al. 2015). Analogously, in macroeconomics, 'Growth-at-Risk' and 'Inflation-at-Risk' are emerging concepts, signifying values that deviate significantly from benchmarks established by institutions such as Central Banks (Adrian et al. 2019; Lopez-Salido and Loria 2020; Iacopini et al. 2023). In other scenarios, the emphasis might rest on the central region or on another specific region of the distribution, often dictated by external constraints or objectives. Examples range from optimizing growing conditions for specific crops such as tubers, to calibrating wind speeds for peak wind turbine performance, and regulating blood sugar levels for effective diabetes management. All these applications require region-specific performance evaluations aligned with the interest in particular outcomes. Accordingly, as illustrated by Lerch et al. (2017), it is crucial to distinguish between strict propriety and strict local propriety; failing to do so can result in misleading results.

The paper is organized as follows. Section 2 provides the foundational concepts. Section 3 introduces the censored scoring rule and establishes its strict local propriety. This section also contains guidance for the practical use of the censoring procedure, a generalization of the Neyman Pearson lemma, and a comparison with alternative weighted scoring rules. Section 4 discusses the empirical performance of our approach. Section 5 concludes. Proofs, derivations of theoretical properties, results of the Monte Carlo study, and additional empirical results are provided in the accompanying Supplementary Material. Data and codes are available from the paper's GitHub: `https://anonymous.4open.science/r/LSPS/`.

2 SCORING RULES AND DIVERGENCES

2.1 Unweighted scoring rules and score divergences

Consider a random variable $Y : \Omega \to \mathcal{Y}$ from a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space $(\mathcal{Y}, \mathcal{G})$. Denote by $\mathcal{P}$ a convex class of probability distributions on $(\mathcal{Y}, \mathcal{G})$. A *scoring rule* S assigns numerical values (scores) to observations $y \in \mathcal{Y}$ and distributions $\mathrm{F} \in \mathcal{P}$, through a mapping $S : \mathcal{P} \times \mathcal{Y} \to \mathbb{R} \cup \{-\infty\} =: \bar{\mathbb{R}}$. Following Holzmann and Klar (2017a), we assume that S is measurable w.r.t. $\mathcal{G}$ and quasi-integrable w.r.t. all $\mathrm{P} \in \mathcal{P}$, for all $\mathrm{F} \in \mathcal{P}$, and such that $\mathbb{E}_{\mathrm{P}} S(\mathrm{F}, Y) < \infty$ and $\mathbb{E}_{\mathrm{P}} S(\mathrm{P}, Y) \in \mathbb{R}$. The latter condition guarantees that the *score divergence*, $\mathbb{D}_S(\mathrm{P} \| \mathrm{F}) := \mathbb{E}_{\mathrm{P}} S(\mathrm{P}, Y) - \mathbb{E}_{\mathrm{P}} S(\mathrm{F}, Y)$, exists and maps onto $(-\infty, \infty]$. Adhering to Gneiting and Raftery (2007), a minimal requirement for S is that it is *strictly proper*.

Definition 1 (Strictly proper scoring rule)**.** *A scoring rule* $S : \mathcal{P} \times \mathcal{Y} \to \bar{\mathbb{R}}$ *is proper relative to* $\mathcal{P}$ *if* $\mathbb{D}_S(\mathrm{P} \| \mathrm{F}) \geq 0,\ \forall \mathrm{P}, \mathrm{F} \in \mathcal{P}$*, and strictly proper if, additionally,* $\mathbb{D}_S(\mathrm{P} \| \mathrm{F}) = 0$ *if and only if* $\mathrm{P} = \mathrm{F}$*,* $\forall \mathrm{P}, \mathrm{F} \in \mathcal{P}$*.*

Equivalently, a score divergence is a *divergence measure* (see, e.g., Eguchi, 1985) if and only if S is strictly proper; here, a divergence measure $\mathbb{D} : \mathcal{P} \times \mathcal{P} \to (-\infty, \infty]$ satisfies (i) $\mathbb{D}(\mathrm{P} \| \mathrm{F}) \geq 0,\ \forall \mathrm{P}, \mathrm{F} \in \mathcal{P}$, and (ii) $\mathbb{D}(\mathrm{P} \| \mathrm{F}) = 0$ if and only if $\mathrm{P} = \mathrm{F},\ \forall \mathrm{P}, \mathrm{F} \in \mathcal{P}$, by definition. For distributions on $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$, where $\mathcal{B}(\mathcal{Y})$ denotes the Borel σ-algebra on $\mathcal{Y}$, such a score divergence is known to be a Bregman (1967) divergence (Ovcharov 2018). This excludes f-divergences other than the Kullback-Leibler divergence (Kullback and Leibler 1951). Two remarks are in place. First, distributions $\mathrm{F} \in \mathcal{P}$ are compared in terms of their P-expected score differences, so that uniqueness of members in $\mathcal{P}$ should formally be interpreted in terms of P-a.s. equivalence classes of P. Similarly, $\mathrm{P} = \mathrm{F}$ is formally defined as $\mathrm{P}(E) = \mathrm{F}(E)$, $\forall E \in \mathcal{G}$. For ease of exposition, we henceforth omit technicalities about P-a.s. equivalence. Second, if there exists a σ-finite measure μ such that $\mathrm{F} \ll \mu$, $\forall \mathrm{F} \in \mathcal{P}$, with $\ll$ denoting absolute continuity, then scoring rules and associated definitions and results can easily be formulated relative to the class of induced μ-densities $f := \frac{\mathrm{dF}}{\mathrm{d}\mu}$, also denoted by $\mathcal{P}$, like classes of distribution functions F.

Gneiting and Raftery (2007) provide an extensive list of strictly proper scoring rules, which can be divided into *local* scoring rules and *distance-sensitive* scoring rules (Ehm and Gneiting 2012). We use the same distinction when discussing examples, yet allowing local scoring rules to also depend on the density via a global norm of the density, and refer to these henceforth as *semi-local*. In this subcategory, our focus lies on the Logarithmic (LogS), Quadratic (QS) and Spherical (SphS) scoring rules, along with their extensions to the Power (PowS_α) and PseudoSpherical (PsSphS_α) families; see Table 1. For distance-sensitive scoring rules we confine ourselves to the rich class of kernel scores. This class has been shown to be strictly proper under known conditions (Gneiting and Raftery 2007; Steinwart and Ziegel 2021), nesting the multivariate Energy Score family, which in turn includes the univariate Continuously Ranked Probability Score (CRPS) as a special case.

2.2 Minimal localization and censoring

In this paper, we suppose the application at hand introduces a region of interest $A \subseteq \mathcal{Y}$, which is assumed to be measurable, i.e., $A \in \mathcal{G}$. Following Holzmann and Klar (2017a), we adopt the strict perspective that outcomes y in the complement $A^c \equiv \mathcal{Y} \backslash A$ are of no interest. In terms of events $E \in \mathcal{G}$, we interpret focusing on A such that only the information generated by events intersecting with A are relevant. To formalize this, we consider the *smallest* σ-algebra on $\mathcal{Y}$ containing all events intersecting with A: $\mathcal{G}_A^{\min} := \sigma(\{E \cap A : E \in \mathcal{G}\}) \subseteq \mathcal{G}$. Note that $\mathcal{G}_A^{\min}$ also includes A^c since σ-algebras are closed under complements. We refer to the restriction of F to the minimal σ-algebra $\mathcal{G}_A^{\min}$, denoted $\mathrm{F}_{|\mathcal{G}_A^{\min}}$, as the coarsest restriction or *minimal localization* of F to A.

Minimal localization naturally gives rise to *censoring*, formally via a pushforward measure of F, from $\mathcal{G}$ to $\mathcal{G}_A^{\min}$. Censoring (Bernoulli 1760) refers to the statistical concept used to model a variable under scrutiny whose value, upon measurement or observation, is only partially known (Tobin 1958). Under censoring, for realizations of a random variable Y that occur in A^c, it is only known that they are not in A. Realizations in A^c are hence indistinguishable under censoring and 'A^c' may therefore be viewed as a single realization of the censored random variable. We censor all outcomes in A^c to a single abstract outcome '$*$', which may be interpreted as 'NaN', uniquely identifying A^c. The importance of allowing $*$ to be outside $\mathcal{Y}$ becomes clear in Section 3.1. In Section 3.2, we discuss distance-sensitive scoring rules, which replace $*$ by a suitable $y_0 \in \mathcal{Y}$. Formally, the *censored random variable* $Y_A^\flat : \mathcal{Y} \to \mathcal{Y}_A^\flat$ is defined as the $\mathcal{G}/\mathcal{G}_A^\flat$-measurable function

$$Y_A^\flat \equiv Y_A^\flat(y) := \begin{cases} y, & y \in A, \\ *, & y \in A^c, \end{cases} \tag{1}$$

with $\mathcal{G}_A^\flat := \sigma(\{E \cap A : E \in \mathcal{G}\} \cup \{*\})$ on $\mathcal{Y}_A^\flat := A \cup \{*\}$. The distribution $\mathrm{F}_A^\flat$ of $Y_A^\flat$ is referred to as the *censored distribution* of F. It is the pushforward measure of F by $Y_A^\flat$, equal to $\mathrm{F}_{|\mathcal{G}_A^{\min}}$ up to relabeling the event $A^c \in \mathcal{G}_A^{\min}$ by $\{*\} \in \mathcal{G}_A^\flat$. In concise form,

$$\mathrm{F}_A^\flat := \mathrm{F}_{|\mathcal{G}_A^\flat} = \mathrm{F}_A + \bar{F}_A \delta_*, \tag{2}$$

where $\mathrm{F}_A(E) := \mathrm{F}(A \cap E)$, $\forall E \in \mathcal{G}^* := \sigma(\{\mathcal{G}, *\})$, $\bar{F}_A := \mathrm{F}(A^c)$ and δ_* denotes the Dirac measure at $*$, i.e., $\delta_*(E) = \mathbb{1}_E(*)$, $\forall E \in \mathcal{G}$. The indicator function $\mathbb{1}_A(y)$ equals unity if

$y \in A$ and zero otherwise. An illustration of minimal localization is given in Example 1.

Example 1 (Minimal localization). *Jim draws a ball from a vase containing six balls, one of which is silver, two are orange, and three are blue. He wins a car if, and only if, he draws the silver ball. This game can be described by a random variable* Y*, on the outcome space* $\mathcal{Y} = \{s, o, b\}$ *with power set* σ*-algebra* $\mathcal{G} = \{\emptyset, \{s\}, \{o\}, \{b\}, \{s, o\}, \{s, b\}, \{o, b\}, \{s, o, b\}\}$*, having probability mass function (pmf)* $p(y) = \frac{1}{6}\mathbb{1}_{\{s\}}(y) + \frac{1}{3}\mathbb{1}_{\{o\}}(y) + \frac{1}{2}\mathbb{1}_{\{b\}}(y),\ y \in \mathcal{Y}$*. Suppose Jim only cares about whether he wins the car or not (and not how he loses). Thus, his region of interest is* $A = \{s\}$*, which induces* $\mathcal{Y}^\flat_A = \{s, *\}$*, where* $\{*\}$ *corresponds to* $\{o, b\}$*, and* $\mathcal{G}^\flat_A = \sigma(\{s, *\}) = \{\emptyset, \{s\}, \{*\}, \{s, *\}\}$*, corresponding to* $\sigma(\{s\})$ *on* $\mathcal{Y}$*. The pmf* $p(y)$ *localizes to* $p^\flat_A(y) = \frac{1}{6}\mathbb{1}_{\{s\}}(y) + \frac{5}{6}\mathbb{1}_{\{*\}}(y),\ y \in \{s, *\}$*.*

2.3 Local and localized divergences and weighted scoring rules

Example 2 (The need to focus). *Continuing Example 1, suppose Jim and his friend Pam know that the vase contains six balls colored silver, orange and blue, but do not know their exact numbers. Jim suspects one silver, four orange, and one blue ball, while Pam believes there are two silver, one orange, and three blue balls. Let Jim's and Pam's implied pmfs be* f *and* g*. One finds* $\mathsf{KL}(p\|f) - \mathsf{KL}(p\|g) = \frac{1}{2}\log(\frac{3}{2}) > 0$*, where* $\mathsf{KL}(p\|f) := \mathbb{E}_p(\log p(Y) - \log f(Y))$ *denotes the Kullback-Leibler (KL) divergence from* p *to* f*. Hence, Pam's belief is statistically closer to the truth in absence of a region of interest. However, since Jim only cares about winning the car, Pam's accurate belief of the blue balls' count, with a relatively high true probability* $p(\{b\}) = 1/2$*, is irrelevant and even misleading. Her close fit outside the silver outcome obscures her inaccuracy where Jim is correct. This illustrates the need to localize the KL divergence to align with Jim's focus on winning.*

As demonstrated by Example 2, it is imperative to adapt the divergence when particular outcomes are of importance. Otherwise, an excellent fit in non-critical regions of the outcome space may obscure a poor fit in regions of relevance. We describe the relative importance of outcomes $y \in \mathcal{Y}$ by a *weight function* $w \in \mathcal{W}$, where $\mathcal{W}$ is a set of $\mathcal{G}$-measurable mappings $w : \mathcal{Y} \to [0, 1]$. Then the question arises how to accordingly transform the divergence and the scoring rule.

Censoring as defined in Section 2.2 pertains to the weight function $w(y) = \mathbb{1}_A(y)$. To simplify the notation, we often use the subscript A in place of $\mathbb{1}_A$ when referring to indicator functions. Censoring transforms the class of distributions from $\mathcal{P}$ to $\mathcal{P}^{\flat}_A$ where $\mathcal{P}^{\flat}_A := \{\mathrm{F}^{\flat}_A, \mathrm{F} \in \mathcal{P}\}$. As we will see in the next section, censoring induces a divergence on $\mathcal{P}^{\flat}_A \times \mathcal{P}^{\flat}_A$ that equals 0 if and only if $\mathrm{P}^{\flat}_A = \mathrm{F}^{\flat}_A$. This, in turn, is equivalent to the measures coinciding (only) locally on A, i.e., $\mathrm{P}(A \cap E) = \mathrm{F}(A \cap E),\ \forall E \in \mathcal{G}$, for which we introduce the short-hand notation $\mathrm{P} \overset{A}{=} \mathrm{F}$.

Let us consider a general divergence measure $\mathbb{D}$, i.e., not necessarily a *score* divergence. In Definitions 2 and 3 below, we introduce a *local divergence* and (the more specific) *localized divergence*. Both definitions are given for general weight functions, corresponding to the region of interest A_w defined as

$$A_w := \{y \in \mathcal{Y} : w(y) > 0\}.$$

On A_w, we again use the short-hand notation $\mathrm{F} \overset{A_w}{=} \mathrm{G}$ for $\mathrm{F}(A_w \cap E) = \mathrm{G}(A_w \cap E),\ \forall E \in \mathcal{G}$. Censoring will yield not just a local divergence but even a localized divergence.

Definition 2 (Local divergence). *A map* $\mathbb{D}_w : \mathcal{P} \times \mathcal{P} \to (-\infty, \infty]$ *is called a* local divergence *(w.r.t.* A_w*) if (i)* $\mathrm{F} \overset{A_w}{=} \mathrm{G}$ *implies* $\mathbb{D}_w(\mathrm{P} \| \mathrm{F}) = \mathbb{D}_w(\mathrm{P} \| \mathrm{G}),\ \forall \mathrm{P}, \mathrm{F}, \mathrm{G} \in \mathcal{P}$*, (ii)* $\mathbb{D}_w(\mathrm{P} \| \mathrm{F}) \geq 0,\ \forall \mathrm{P}, \mathrm{F} \in \mathcal{P}$*, and (iii)* $\mathbb{D}_w(\mathrm{P} \| \mathrm{F}) = 0$ *if and only if* $\mathrm{P} \overset{A_w}{=} \mathrm{F},\ \forall \mathrm{P}, \mathrm{F} \in \mathcal{P}$.

Definition 3 (Localized divergence). *Let* $\mathcal{P}_w$ *denote a class of distributions obtained by a map* $[\cdot]_w : \mathcal{P} \to \mathcal{P}_w$ *coinciding with the identity map* $[\mathrm{F}]_w = \mathrm{F}, \forall \mathrm{F} \in \mathcal{P}$ *for* $w = \mathbb{1}_{\mathcal{Y}}$*. A local divergence* $\mathbb{D}_w : \mathcal{P} \times \mathcal{P} \to (-\infty, \infty]$ *is called a* localized divergence *of* $\mathbb{D}$ *(w.r.t.* A_w*)*

if $\forall \mathrm{P}_w, \mathrm{F}_w \in \mathcal{P}_w,\ \exists \mathrm{P}, \mathrm{F} \in \mathcal{P}:$

$$\mathbb{D}_w(\mathrm{P}\|\mathrm{F}) = \mathbb{D}(\mathrm{P}_w\|\mathrm{F}_w).$$

Condition (i) in Definition 2 ensures invariance w.r.t. (information generated by the) events that are not intersecting with A_w, hence are irrelevant. Furthermore, condition (iii) applies only locally on A_w. Definition 3 introduces the subclass of local divergences that preserve the unweighted divergence measure $\mathbb{D}$ by applying it to a weighted transformation of the distribution space.

Just as strictly proper scoring rules give rise to divergence measures, local divergences emerge naturally from *weighted scoring rules* that are *strictly locally proper* relative to some class of distributions $\mathcal{P}$ and weight functions $\mathcal{W}$. For all $w \in \mathcal{W}$, we define a weighted scoring rule as the map $S_w : \mathcal{P} \times \mathcal{Y} \to \bar{\mathbb{R}}$ such that $S_w(\cdot,\cdot)$ is a scoring rule. A weighted scoring rule is said to be *localizing* if measures coinciding on A_w receive the same score for any realization; formally, $\forall \mathrm{P}, \mathrm{F} \in \mathcal{P},\ S_w(\mathrm{P}, y) = S_w(\mathrm{F}, y),\ \forall y \in \mathcal{Y}$, whenever $\mathrm{P} \overset{A_w}{=} \mathrm{F}$. Definition 4 extends strict propriety to localizing weighted scoring rules and is equivalent to that given by Holzmann and Klar (2017a, p. 2414). Here, $\mathbb{D}_{S_w}(\mathrm{P}\|\mathrm{F}) := \mathbb{E}_{\mathrm{P}}(S_w(\mathrm{P}, Y)) - \mathbb{E}_{\mathrm{P}}(S_w(\mathrm{F}, Y))$.

Definition 4 (Strictly locally proper scoring rule)**.** *A weighted scoring rule* $S_w : \mathcal{P}\times\mathcal{Y} \to \bar{\mathbb{R}}$ *is locally proper relative to* $(\mathcal{P}, \mathcal{W})$ *if it is localizing and* $S_w(\cdot,\cdot)$ *is proper for all* $w \in \mathcal{W}$*. Furthermore, it is strictly locally proper relative to* $(\mathcal{P}, \mathcal{W})$ *if, additionally,* $\mathrm{P} \overset{A_w}{=} \mathrm{F}$ *if and only if* $\mathbb{D}_{S_w}(\mathrm{P}\|\mathrm{F}) = 0$*,* $\forall w \in \mathcal{W}$*.*

Clearly, the score divergence $\mathbb{D}_{S_w}(\mathrm{P}\|\mathrm{F})$ is a local divergence for all $w \in \mathcal{W}$ if and only if S_w is strictly locally proper relative to $(\mathcal{P}, \mathcal{W})$. As a general note, we remark that focusing on particular (subsets of) outcomes using a strictly locally proper scoring rule represents a strict modeling perspective; possibly non-localizing scoring rules, such as the twCRPS, can have their own merits in specific applications, by taking additional information into account; see also Example D.3 in the Appendix and its discussion. As such, non-localizing scoring rules may statistically be competitive to strictly locally proper scoring rules as evidenced by our Monte Carlo simulations.

3 THE CENSORED SCORING RULE

In this section, we introduce the censored scoring rule. To this end, we first generalize the censored distribution in Equation (2) to general weight functions, and next to arbitrary nuisance distributions that suitably replace the Dirac measure in Equation (3) below, to accommodate distance-sensitive scoring rules.

For the definition of the censored distribution, it is helpful to first extend w and F to w^* and F^* on $(\mathcal{Y}^*, \mathcal{G}^*)$, where $\mathcal{Y}^* := \mathcal{Y} \cup \{*\}$, such that $w^*(y) = w(y)\mathbb{1}_{\mathcal{Y}}(y)$, $\forall y \in \mathcal{Y}^*$ and $\mathrm{F}^*(E) = \mathrm{F}(E \cap \mathcal{Y})$, $\forall E \in \mathcal{G}^*$. In the original measurable space, $*$ is then already an outcome, associated with the event $\{*\}$, albeit assigned with weight and measure zero. For ease of notation, we henceforth drop the superscripts $*$. The censored distribution in Equation (2) generalizes to

$$\mathrm{dF}^\flat_w := \mathrm{dF}_w + \bar{F}_w \, \mathrm{d}\delta_*, \qquad \text{where} \quad \mathrm{dF}_w := w \, \mathrm{dF}, \qquad \bar{F}_w := 1 - \int_{\mathcal{Y}} w \, \mathrm{dF}, \tag{3}$$

defining a probability measure on $(\mathcal{Y}^\flat_{A_w}, \mathcal{G}^\flat_{A_w})$, with $\mathcal{G}^\flat_{A_w} := \sigma(\{E \cap A_w : E \in \mathcal{G}\} \cup \{*\})$ and $\mathcal{Y}^\flat_{A_w} := A_w \cup \{*\}$. In contrast to indicator functions, the probability of the censoring event $\bar{F}_w$ is generally different from $\mathrm{F}(A^c_w) = 1 - \int_{A_w} \mathrm{dF}$, rendering the identification $Y^\flat_{A_w}(y) = * \iff y \in A^c_w$ infeasible. However, Appendix C shows that identification is recoverable by introducing an auxiliary random variable $Z|(Y = y)$, which is $\mathrm{B}_{w(y)} \equiv$ Bernoulli$(w(y))$-distributed, to define $Y^\flat_w \equiv Y^\flat_w(y, z)$ as being y if $z = 1$ and $*$ otherwise, since then $Y^\flat_w(y, z) = * \iff z = 0$. The distribution $\mathrm{F}^\flat_w$ admits the $(\mu + \delta_*)$-density $f^\flat_w(y) = w(y)f(y)\mathbb{1}_{A_w}(y) + \bar{F}_w \mathbb{1}_{\{*\}}(y)$, $y \in \mathcal{Y}^\flat_{A_w}$, provided that $\mathrm{F} \ll \mu$; see Appendix B.1 for details.

For indicator weight functions, the density simplifies to $f^\flat_A(y) = f(y)\mathbb{1}_A(y) + \bar{F}_A \mathbb{1}_{A^c}(y)$, $y \in \mathcal{Y}$, similar to Borowska et al. (2020).

A critical difference from the (generalized) conditional distribution

$$\mathrm{dF}^\sharp_w := \frac{1}{1 - \bar{F}_w}\,\mathrm{dF}_w,$$

assuming $\bar{F}_w < 1$, is that $\mathrm{F}^\sharp_A(A) \neq \mathrm{F}(A) = \mathrm{F}^\flat_A(A)$. In other words, conditioning does not preserve all probabilities of interest and does accordingly not coincide with the minimal localization of F to A. The symbols 'sharp' ($\sharp$) and 'flat' ($\flat$) reflect their respective operations: conditioning sharpens the density on A by a factor $1/(1 - \bar{F}_A)$, whereas censoring flattens the shape outside A into a point mass. The associated scoring rule $S^\sharp_w(\mathrm{F}, y) := w(y)S(\mathrm{F}^\sharp_w, y)$ fails to be strictly locally proper. Holzmann and Klar (2017a) remedy this by adding an auxiliary scoring rule for the missing information about A^c. However, by this addition, the corresponding score divergence generally fails to be a localized divergence; see Section 3.5.

3.1 Censored scoring

Ideally, the censored scoring rule would be given by the identity $S^\flat_A(\mathrm{F}, y) = S(\mathrm{F}^\flat_A, y^\flat_A)$, as this would fully respect the forecaster's specific choice of the unweighted scoring rule S. The censored scoring rule given by Definition 5 below indeed reduces to this definition for the indicator weight function $w(y) = \mathbb{1}_A(y)$. It is also attractive for general weight functions, for which the randomization perspective based on the auxiliary random variable Z in Appendix C yields the similar identity $S^\flat_w(\mathrm{F}, y) = \mathbb{E}_{\mathrm{B}_{w(y)}} S(\mathrm{F}^\flat_w, Y^\flat_w(y, Z))$.

Definition 5 (Censored scoring rule)**.** *Let $S : \mathcal{P}^\flat \times \mathcal{Y} \to \bar{\mathbb{R}}$, $\mathcal{P}^\flat = \{\mathrm{F}^\flat_w, \mathrm{F} \in \mathcal{P}, w \in \mathcal{W}\}$, denote a scoring rule. Then, for all $w \in \mathcal{W}$, the corresponding censored scoring rule is given by the map $S^\flat_w : \mathcal{P} \times \mathcal{Y} \to \bar{\mathbb{R}}$,*

$$S^\flat_w(\mathrm{F}, y) := w(y)S(\mathrm{F}^\flat_w, y) + \big(1 - w(y)\big)S(\mathrm{F}^\flat_w, *),$$

where the censored distribution $\mathrm{F}_w^\flat$ *is given in Equation* (3)*.*

Theorem 1 establishes that the censored scoring rule is strictly locally proper.

Theorem 1. *If the scoring rule* S *is strictly proper relative to* $\mathcal{P}^\flat$*, the censored scoring rule* $S_w^\flat$ *in Definition 5 is strictly locally proper relative to* $(\mathcal{P}, \mathcal{W})$*. Moreover, its associated score divergence* $\mathbb{D}_{S_w^\flat}$ *is a localized divergence of* $\mathbb{D}_S$*, for all* $w \in \mathcal{W}$*.*

Theorem 1 is a special case of the more general Theorem 2 below, hence its proof is subsumed in the proof of Theorem 2. The assumption required in Theorem 1 ensures that the unweighted scoring rule is well-defined w.r.t. mixed continuous-discrete distributions on measurable spaces extended by $*$. For the PsSphS_α family, this is explicitly verified in Example 3; the same argument applies to all semi-local scoring rules considered, *mutatis mutandis.*

Let us provide some intuition for the result of Theorem 1, relying on the established mathematical identity $\mathbb{D}_{S_w^\flat}(\mathrm{P}\|\mathrm{F}) = \mathbb{D}_S(\mathrm{P}_w^\flat\|\mathrm{F}_w^\flat)$. Since S is assumed to be strictly proper relative to $\mathcal{P}^\flat$, its score divergence $\mathbb{D}_S$ defines a divergence on $\mathcal{P}_w^\flat \times \mathcal{P}_w^\flat$, for all $w \in \mathcal{W}$. However, when viewed as a composite map, first censoring and then computing the divergence, it is no longer a divergence on the space of uncensored measures, as the censoring map is generally non-invertible. Nonetheless, by maintaining the probability of the censoring event, censoring preserves identifiability on A_w, i.e., $\mathrm{P} \overset{A_w}{=} \mathrm{F}$ if and only if $\mathrm{P}_w^\flat = \mathrm{F}_w^\flat$. Hence, under censoring, the strict propriety of the scoring rule and the associated divergence are retained locally. For the Logarithmic scoring rule, the censored scoring rule coincides with that of Diks et al. (2011).

Example 3 (Censored PsSphS)**.** *Consider a class of* μ*-densities* $\mathcal{P}_\alpha$ *on* $(\mathcal{Y}, \mathcal{G}, \mu)$ *with finite* L^α*-norm, i.e.,* $\|f\|_\alpha := \left(\int_\mathcal{Y} f^\alpha \mathrm{d}\mu\right)^{1/\alpha} < \infty$, $\forall f \in \mathcal{P}_\alpha$*. The PseudoSpherical family* $\mathrm{PsSphS}_\alpha(f, y) = f(y)^{\alpha-1}/\|f\|_\alpha^{\alpha-1}$, $\alpha > 1$*, as advocated by Gneiting and Raftery (2007), is strictly proper relative to* $\mathcal{P}_\alpha$*. To verify its strict propriety relative to* $\mathcal{P}_\alpha^\flat$ *as required for*

Theorem 1, we write $\|f_w^\flat\|_\alpha^\alpha \leq 1 + \|f\|_\alpha^\alpha < \infty$, $\forall f \in \mathcal{P}_\alpha$, $\forall w \in \mathcal{W}$, *hence* $S_w^\flat(f, y)$ *is strictly locally proper relative to* $(\mathcal{P}, \mathcal{W})$. *Notably, while* $S_w^\flat(f, *) = \bar{F}_w^{\alpha-1}/(\|wf\|_\alpha + \bar{F}_w^\alpha)^{(\alpha-1)/\alpha}$ *does not depend solely on* $\bar{F}_w$, *it holds that* $S_w^\flat(f, *) = S_w^\flat(g, *)$, *if* $f = g$ *a.s. on* A_w.

In Theorem 1, we assume that the scoring rule S is well defined relative to a class of distributions $\mathcal{P}^\flat$ with a point mass at $*$. Distance-sensitive scoring rules are generally incompatible with such measures, since the distance to $*$ is undefined. Section 3.2 shows that, in typical applications, $*$ can then be replaced by any $y_0 \in \mathcal{Y}$, provided that all $\mathrm{F} \in \mathcal{P}$ assign zero mass to y_0 or all $w \in \mathcal{W}$ assign zero weight to y_0. Then, $\mathrm{F}_w(y_0) = 0$, which ensures identifiability of the censoring event. Here, we introduce the short-hand notation $\mathrm{F}_w(y_0)$ to indicate the measure F_w of the event $\{y_0\}$. In Example 1, the above condition would prevent choosing $y_0 = s$, avoiding $(Y_A^\flat)^{-1}(s) = s \cup A^c = \mathcal{Y}$. When using semi-local scoring rules, labeling the censored outcome as $* \notin \mathcal{Y}$ ensures this identifiability by construction, even if all outcomes in $\mathcal{Y}$ have positive mass under all distributions in $\mathcal{P}$ and $w(y) > 0$ for all $y \in \mathcal{Y}$.

3.2 Generalized censored scoring

This subsection introduces a more flexible censoring framework. Suppose that a weight function introduces k pivotal points $r_1, \ldots, r_k \in \mathcal{Y}$. Then a natural generalization of the censored distribution in Equation (3) reads

$$\mathrm{dF}_{w,\mathcal{R}_k}^\flat := \mathrm{dF}_w + \bar{F}_w \sum_{i=1}^{k} \gamma_i \mathrm{d}\delta_{r_i}, \qquad \boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_k)' \in \Delta(k), \tag{4}$$

where $\Delta(k)$ denotes the unit $(k-1)$-simplex and $\mathcal{R}_k := \{r_i\}_{i=1}^k$, with $r_i \in \mathcal{Y}, \forall i$. Section 3.3 provides guidance on choosing (r_i, γ_i). Definition 6 formalizes the adaptation of the censored scoring rule to generalized censored measures, nesting $\mathrm{F}_{w,\mathcal{R}_k}^\flat$ for $\mathrm{H} = \sum_{i=1}^k \gamma_i \delta_{r_i}$. Here, we refer to H as a *nuisance* distribution since its sole role is to suitably allocate the probability mass $\bar{F}_w$.

Definition 6 (Generalized censored scoring rule)**.** *Let* $S : \mathcal{P}^\flat \times \mathcal{Y} \to \bar{\mathbb{R}}$*, denote a scoring rule, where* $\mathcal{P}^\flat = \{\mathrm{F}^\flat_{w,\mathrm{H}}, \mathrm{F} \in \mathcal{P}, w \in \mathcal{W}, \mathrm{H} \in \mathcal{H}\}$*, in which* $\mathrm{dF}^\flat_{w,\mathrm{H}} := \mathrm{dF}_w + \bar{F}_w \mathrm{dH}$ *denotes the generalized censored distribution and* $\mathcal{H} \subseteq \mathcal{P}$ *a class of nuisance distributions. Then for all* $w \in \mathcal{W}$ *and* $\mathrm{H} \in \mathcal{H}$ *the associated generalized censored scoring rule is given by the map* $S^\flat_{w,\mathrm{H}} : \mathcal{P} \times \mathcal{Y} \to \bar{\mathbb{R}}$*,*

$$S^\flat_{w,\mathrm{H}}(\mathrm{F}, y) := w(y) S(\mathrm{F}^\flat_{w,\mathrm{H}}, y) + \big(1 - w(y)\big) \mathbb{E}_\mathrm{H} S(\mathrm{F}^\flat_{w,\mathrm{H}}, Q),$$

where H *denotes the distribution of the random variable* Q*, distributed independently of* y*.*

Since both F and H are defined relative to $(\mathcal{Y}, \mathcal{G})$, $\mathrm{F}^\flat_{w,\mathrm{H}}$ is defined relative to $(\mathcal{Y}, \mathcal{G})$ too, for all $w \in \mathcal{W}, \mathrm{H} \in \mathcal{H}$. The dependence of $\mathcal{P}^\flat$ on $(\mathcal{P}, \mathcal{W}, \mathcal{H})$ in Definition 6 is notationally suppressed.

Assumption 1. *The weight function* $w \in \mathcal{W}$ *and nuisance distribution* $\mathrm{H} \in \mathcal{H} \subseteq \mathcal{P}$ *are such that* $\exists E \in \mathcal{G} : \mathrm{F}_w(E) = 0$ *and* $\mathrm{H}(E) > 0,\ \forall \mathrm{F} \in \mathcal{P}, \mathrm{H} \in \mathcal{H}$*.*

The following theorem, the proof of which is contained in Appendix A.1, establishes the strict local propriety of the generalized scoring rule.

Theorem 2. *Suppose that: (i) the unweighted scoring rule* S *in Definition 6 is strictly proper relative to* $\mathcal{P}^\flat$*, and (ii)* $\mathcal{W}$ *and* $\mathcal{H}$ *are such that Assumption 1 is satisfied. Then, the generalized censored scoring rule* $S^\flat_{w,\mathrm{H}}$ *in Definition 6 is strictly locally proper relative to* $(\mathcal{P}, \mathcal{W}, \mathcal{H})$*. Moreover, its associated score divergence* $\mathbb{D}_{S^\flat_{w,\mathrm{H}}}$ *is a localized divergence of* $\mathbb{D}_S$*, for all* $w \in \mathcal{W}$*,* $\mathrm{H} \in \mathcal{H}$*.*

The intuition behind Theorem 2 builds on that of Theorem 1, by the extended identity

$$\mathbb{D}_{S^\flat_{w,\mathrm{H}}}(\mathrm{P} \| \mathrm{F}) = \mathbb{D}_S(\mathrm{P}^\flat_{w,\mathrm{H}} \| \mathrm{F}^\flat_{w,\mathrm{H}}). \tag{5}$$

Assumption 1 implies that $\mathrm{P}^\flat_{w,\mathrm{H}} = \mathrm{F}^\flat_{w,\mathrm{H}}$ if and only if $\mathrm{P} \overset{A_w}{=} \mathrm{F}$. For any weight function, this holds trivially for $\mathrm{H} = \delta_*$, while $\mathrm{H} = \delta_{y_0}$, with $y_0 \in \mathcal{Y}$, requires the measure F to satisfy

$\mathrm{F}_w(y_0) = 0$. For $\mathrm{H} = \gamma\delta_{r_1} + (1-\gamma)\delta_{r_2}$, $r_1, r_2 \in \mathcal{Y}$, and $\gamma \in [0,1]$, Assumption 1 demands $\mathrm{F}_w(r_i) = 0$ for at least one $i \in \{1,2\}$; so, if both r_1 and r_2 belong to A_w, then at least one must not carry mass under F. This readily generalizes to finitely many points r_i. Indeed, Appendix C shows that by writing $S^\flat_{w,\mathrm{H}}(\mathrm{F}, y) = \mathbb{E}_{\mathrm{B}_{w(y)},\mathrm{H}} S\big(\mathrm{F}^\flat_{w,\mathrm{H}}, Y^\flat_w(y, Z, Q)\big)$, identifiability of the censoring event under a single realization of Q suffices, since strict local propriety is retained if at least one scoring rule in a sum of proper scoring rules is strictly locally proper.

A rich class of scoring rules obtained by the generalization of the censored scoring rule is that of kernel scores (Gneiting and Raftery 2007). Kernel scores depend on a negative definite kernel $\rho : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$, i.e., a symmetric function satisfying $\sum_{i=1}^n \sum_{j=1}^n a_i a_j \rho(y_i, y_j) \leq 0$, for all $n \in \mathbb{N}$, $a_1, \ldots, a_n \in \mathbb{R}$ that sum to 0 and all $y_1, \ldots, y_n \in \mathcal{Y}$. These include popular examples such as the Euclidean distance on $\mathbb{R}^d$ and the angular distance between two points on a circle. Example 4 displays the general expression for the generalized censored kernel score and corresponding localized divergence for $\mathrm{H} = \delta_r$.

Example 4. *Consider a class of distributions* $\mathcal{P}_r$ *on some measurable space* $(\mathcal{Y}, \mathcal{G})$*, such that* $\mathrm{F}(r) = 0, \forall \mathrm{F} \in \mathcal{P}_r$*, where* $r \in \mathcal{Y}$*, including all (absolutely) continuous distributions on* $\mathcal{Y}$*. Consider the kernel score family* $S_\rho(\mathrm{F}, y) := \frac{1}{2}\mathbb{E}_\mathrm{F}\rho(X, X') - \mathbb{E}_\mathrm{F}\rho(X, y) + \frac{1}{2}\rho(y, y)$ *with divergence* $\mathbb{D}_{S_\rho}(\mathrm{P}\|\mathrm{F}) = \mathbb{E}_{\mathrm{P},\mathrm{F}}\rho(X, Y) - \frac{1}{2}\mathbb{E}_\mathrm{P}\rho(X, X') - \frac{1}{2}\mathbb{E}_\mathrm{F}\rho(Y, Y')$*, nesting the CRPS*$(F, y) := -\int_\mathbb{R} \big(F(s) - \mathbb{1}_{[y,\infty)}(s)\big)^2 \mathrm{d}s$ *as a special case for* $\rho(x, x') = |x - x'|$ *on* $\mathbb{R}$*. The associated generalized censored scoring rule for* $\mathrm{H} = \delta_r$ *reads* $S^\flat_{\rho,w,r}(\mathrm{F}, y) = \frac{1}{2}\mathbb{E}_{\mathrm{F}^\flat_{w,r}}\rho(X, X') - w(y)\Big(\mathbb{E}_{\mathrm{F}^\flat_{w,r}}\rho(X, y) - \frac{1}{2}\rho(y, y)\Big) - \big(1 - w(y)\big)\Big(\mathbb{E}_{\mathrm{F}^\flat_{w,r}}\rho(X, r) - \frac{1}{2}\rho(r, r)\Big)$*. Assumption 1 is clearly satisfied for all weight functions* $w \in \mathcal{W}$ *and distributions* $\mathrm{F} \in \mathcal{P}_r$*. Therefore, the score divergence* $\mathbb{D}_{S^\flat_{\rho,w,r}}(\mathrm{P}\|\mathrm{F}) = \mathbb{E}_{\mathrm{P}^\flat_{w,r},\mathrm{F}^\flat_{w,r}}\rho(X, Y) - \frac{1}{2}\mathbb{E}_{\mathrm{P}^\flat_{w,r}}\rho(X, X') - \frac{1}{2}\mathbb{E}_{\mathrm{F}^\flat_{w,r}}\rho(Y, Y')$ *is a localized divergence if* S_ρ *is strictly proper relative to* $\mathcal{P}^\flat_r$*, which follows from the conditions under which* S_ρ *is strictly proper relative to* $\mathcal{P}_r$*. For one-sided indicator weight functions* $\mathbb{1}_{(-\infty,r)}(y)$ *and* $\mathbb{1}_{(r,\infty)}(y)$*, the censored CRPS coincides with the* $\mathrm{twCRPS}(F, y) :=$

$-\int_{\mathbb{R}} w(s)\big(F(s) - \mathbb{1}_{[y,\infty)}(s)\big)^2 \mathrm{d}s$ *(Gneiting and Ranjan 2011). More details are given in Appendix E.4.*

From Example 4, we have that $\mathrm{CRPS}^{\flat}_{w,r} = \mathrm{twCRPS}$ for all weight functions for which Holzmann and Klar (2017a, Theorem 5) proved that the twCRPS is strictly locally proper. For $A = (r_1, r_2)$, $A_1^c = (-\infty, r_1]$, $A_2^c = [r_2, \infty)$, $r_1, r_2 \in \mathbb{R}$, the twCRPS decomposes the mass $\bar{F}_A$ into $\mathrm{F}(A_1^c)$ and $\mathrm{F}(A_2^c)$, hence would coincide with $\mathrm{F}^{\flat}_{A,r_1,r_2} := \mathrm{F}_A + \bar{F}_A\big(\gamma\delta_{r_1} + (1-\gamma)\delta_{r_2}\big)$, $\gamma \in [0,1]$, when $\gamma = \mathrm{F}(A_1^c)/\bar{F}_A$. However, this choice of γ — which we do not allow — introduces a dependence on outcomes outside A, rendering the twCRPS non-localizing. For weight functions for which the twCRPS loses its strict local propriety due to its non-localizing nature, $\mathrm{CRPS}^{\flat}_{w,\mathcal{R}_k}$ may serve as a strictly locally proper alternative.

3.3 Practical guidance

The censoring procedure proposed in this paper is sufficiently general to enable researchers who aim to compare competing forecasts to construct censored counterparts of their preferred scoring rules, for practically any chosen family of weight functions. With competing density forecasts at hand, the (generalized) censored scoring rules may be readily applied to obtain scores that can serve as input in a DM type test statistic; see Section 4. The GitHub page associated with this paper provides code for all 18 focused scoring rules considered in our simulations and empirical applications. In particular, for *semi-local scoring rules*, Definition 5 provides an analytic expression for the censored scoring rule for any choice of weight function. Table 1 lists the resulting (simplified) formulas for the unweighted, conditional and censored LogS, PowS_α and PsSphS_α families of scoring rules, as well as their localized divergences derived from the identity $\mathbb{D}_{S^{\flat}_w}(\mathrm{P}\|\mathrm{F}) = \mathbb{D}_S(\mathrm{P}^{\flat}_w\|\mathrm{F}^{\flat}_w)$. The table also displays the generalized censored scoring rules of Definition 6, which for semi-local scoring rules exhibit insensitivity to the nuisance distribution, as long as the nuisance distribution is normalized to $\|h\|_\alpha = 1$.

For *distance-sensitive scoring rules*, such as the kernel scores discussed in Example 4, we use the generalized censoring approach in Definition 6 based on the censored measure in Equation (4). The localized versions of distance-sensitive scoring rules thereby depend on the choice of the pivotal points and associated weights. As illustrated by Examples D.1 and D.2 in Appendix D, weight functions often suggest natural choices for pivotal points, and it is these points we recommend incorporating into the censored measure in practice. Specifically, for the real-valued weight functions $I_{\mathrm{L}}(y;r) := \mathbb{1}_{(-\infty,r)}(y)$, $I_{\mathrm{R}}(y;r) := \mathbb{1}_{(r,\infty)}(y)$, $\Lambda_{a,\mathrm{L}}(y;r) := \frac{1}{1+\exp(a(y-r))}$, $a > 0$, the choice $r \in \mathbb{R}$ is considered pivotal. Similarly, for the weight functions on $\mathbf{y} \in \mathbb{R}^2$ given by $I^2_{\mathrm{L}}(\mathbf{y};\mathbf{r}) := I_{\mathrm{L}}(y_1;r_1) \times I_{\mathrm{L}}(y_2;r_2)$, $I^2_{\mathrm{R}}(\mathbf{y};\mathbf{r}) := I_{\mathrm{R}}(y_1;r_1) \times I_{\mathrm{R}}(y_2;r_2)$ and $\Lambda^2_{a,\mathrm{L}}(\mathbf{y};\mathbf{r}) := \Lambda_{a,\mathrm{L}}(y_1;r_1) \times \Lambda_{a,\mathrm{L}}(y_2;r_2)$, the use of $\mathbf{r} \in \mathbb{R}^2$ is considered natural. For the center indicator, $I_{\mathrm{C}}(y;\ell,r) := \mathbb{1}_{(\ell-r,\ell+r)}(y)$, there are two pivotal points $r_{1,2} = \ell \pm r$. In Section 4, we use the fraction of observations smaller than r_1 to tune the weight γ in (4). For its complement, $I^{\mathrm{c}}_{\mathrm{C}}(y;\ell,r) := 1 - I_{\mathrm{C}}(y;\ell,r)$, we adopt the single pivotal point ℓ.

As clarified in Section 3.2, Assumption 1 is easily satisfied by the censored distribution in Equation (4). For instance, if the underlying distributions are (absolutely) continuous, any value of r is valid; this includes Gaussian density and distribution functions. If the distribution is discrete or discrete-continuous, any r at which the distributions under consideration exhibit no point mass may be chosen.

3.4 Localized Neyman Pearson

Anticipating the applications in the next section, we now consider an explicit time-series context. Specifically, we consider a stochastic process $\{Y_t : \Omega \to \mathcal{Y}\}_{t=1}^T$ from a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space $(\mathcal{Y}^T, \mathcal{G}^T)$, where $\mathcal{Y}^T$ and $\mathcal{G}^T$ denote the product outcome space and σ-algebra of the individual outcome spaces $\mathcal{Y}$ and σ-algebras

Table 1: Examples of semi-local scoring rules.

Name	Logarithmic	Power family	PseudoSpherical family
Unweighted			
$S(f,y)$	$\mathrm{LogS}(f,y)=\log f(y)$	$\mathrm{PowS}_\alpha(f,y)=\alpha f(y)^{\alpha-1}-(\alpha-1)\|f\|_\alpha^\alpha, \quad \alpha>1$	$\mathrm{PsSphS}_\alpha(f,y)=\dfrac{f(y)^{\alpha-1}}{\|f\|_\alpha^{\alpha-1}}, \quad \alpha>1$
Special cases	–	$\mathrm{QS}(f,y)=\mathrm{PowS}_2(f,y)$ $\mathrm{LogS}(f,y)=\lim\limits_{\alpha\downarrow 1}\dfrac{\mathrm{PowS}_\alpha(f,y)-1}{\alpha-1}$	$\mathrm{SphS}(f,y)=\mathrm{PsSphS}_2(f,y)$ $\mathrm{LogS}(f,y)=\lim\limits_{\alpha\downarrow 1}\dfrac{\mathrm{PsSphS}_\alpha(f,y)-1}{\alpha-1}$
$\mathbb{D}_S(p\\|f)$	$\mathrm{KL}(p\\|f)=\mathbb{E}_p\log\left(\frac{p}{f}\right)$	$\|p\|_\alpha^\alpha-\alpha\int f^{\alpha-1}(p-f)\mathrm{d}\mu-\|f\|_\alpha^\alpha$	$\|p\|_\alpha-\frac{\int pf^{\alpha-1}\mathrm{d}\mu}{\|f\|_\alpha^{\alpha-1}}$
$\alpha=2$	–	$\|p-f\|_2^2$	$\|p\|_2\left(1-C(p,f)\right)$
SP class	$\mathcal{P}_{\alpha=1}$	$\mathcal{P}_\alpha$	$\mathcal{P}_\alpha$
$\zeta(t)$	$t\log t$	t^α	–
$S(\tilde{f},\tilde{y})$	$\log f(y)-\log\|b\|$	$\left(\frac{1}{\|b\|}\right)^{\alpha-1}\mathrm{PowS}_\alpha(f,y)$	$\left(\frac{1}{\|b\|}\right)^{\frac{\alpha-1}{\alpha}}\mathrm{PsSphS}_\alpha(f,y)$
Focused			
$S_w^\sharp(f,y)$	$w(y)\log\left(\frac{f(y)}{1-\bar{F}_w}\right)$	$w(y)\Big(\alpha\left(\frac{f_w(y)}{1-\bar{F}_w}\right)^{\alpha-1}-(\alpha-1)\left\\|\frac{f_w(y)}{1-\bar{F}_w}\right\\|_\alpha^\alpha\Big)$	$w(y)\frac{f_w(y)^{\alpha-1}}{\|f_w\|_\alpha^{\alpha-1}}$
$S_w^\flat(f,y)$	$w(y)\log f(y)+\big(1-w(y)\big)\log\bar{F}_w$	$w(y)\alpha f_w(y)^{\alpha-1}+\big(1-w(y)\big)\alpha\bar{F}_w^{\alpha-1}$ $-(\alpha-1)\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)$	$\frac{w(y)f_w(y)^{\alpha-1}+(1-w(y))\bar{F}_w^{\alpha-1}}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{\alpha-1}{\alpha}}}$
$S_{w,h}^\flat(f,y)$	$w(y)\log f(y)+\big(1-w(y)\big)\log\bar{F}_w$	$w(y)\alpha f_w(y)^{\alpha-1}+\big(1-w(y)\big)\alpha\bar{F}_w^{\alpha-1}\|h\|_\alpha^\alpha$ $-(\alpha-1)\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\|h\|_\alpha^\alpha\right)$	$\frac{w(y)f_w(y)^{\alpha-1}+\big(1-w(y)\big)\bar{F}_w^{\alpha-1}\|h\|_\alpha^\alpha}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\|h\|_\alpha^\alpha\right)^{\frac{\alpha-1}{\alpha}}}$
$\mathbb{D}_{S_w^\flat}(p\\|f)$	$\int\log\left(\frac{p_w}{f_w}\right)p_w\mathrm{d}\mu+\log\left(\frac{\bar{P}_w}{\bar{F}_w}\right)\bar{P}_w$	$\|p_w\|_\alpha^\alpha+\bar{P}_w^\alpha-\int p_w f_w^{\alpha-1}\mathrm{d}\mu-\bar{P}_w\bar{F}_w^{\alpha-1}$ $-(\alpha-1)\big(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\big)$	$\left(\|p_w\|_\alpha^\alpha+\bar{P}_w^\alpha\right)^{\frac{1}{\alpha}}-\frac{\int p_w f_w^{\alpha-1}\mathrm{d}\mu+\bar{P}_w\bar{F}_w^{\alpha-1}}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{\alpha-1}{\alpha}}}$

NOTE: This table displays unweighted and focused scoring rules, divergences and associated properties based on two μ-densities, p and f, living on the measurable space $(\mathcal{Y},\mathcal{G},\mu)$, equipped with the L^α-norm $\|p\|_\alpha=(\int_\mathcal{Y}p^\alpha\mathrm{d}\mu)^{1/\alpha}$. The common limiting case of PowS_α and PsSphS_α remains to hold for conditioning and censoring. $\mathbb{D}_S(p\|f)$ denotes the score divergence of f from p and $C(p,f)=\int pf\mathrm{d}\mu/\sqrt{\int p^2\mathrm{d}\mu\int f^2\mathrm{d}\mu}$, the cosine similarity between p and f. The strict propriety class (SP class) is the class of probability measures relative to which the scoring rule is strictly proper. $\mathcal{P}_\alpha$ denotes the class of densities on $(\mathcal{Y},\mathcal{G},\mu)$ such that $\|p\|_\alpha<\infty, \quad \forall p\in\mathcal{P}_\alpha$. The Bregman generator function $\zeta(t)$ parameterizes the subclass of separable Bregman divergences, consisting of the score divergences based on strictly proper scoring rules $S_\zeta:\mathcal{P}\times\mathcal{Y}\to\mathbb{R}$ of the form $S_\zeta(p,y)=\zeta'(p(y))-\int_\mathcal{Y}\zeta'(p(y))p(y)-\zeta(p(y))\mu(\mathrm{d}y)$. $S(\tilde{f},\tilde{y})$ denotes the score of the real-valued random variable $\tilde{Y}=bY+a$, where $a\in\mathbb{R}$ and $b\in\mathbb{R}\backslash\{0\}$, with density $\tilde{f}(\tilde{y})=\frac{1}{|b|}f\big(\frac{\tilde{y}-a}{b}\big)$. The presented results for the focused scoring rules are equivalent in the sense that they yield the same expected score. The generalized censored scoring rule $S_{w,h}^\flat$ departs from a density h of which the support is a subset of $A_w\subseteq\mathcal{Y}$. The weight function is restricted accordingly. Appendix E details the derivations of the results presented in this table.

$\mathcal{G}$, respectively. The process generates the filtration $\{\mathcal{F}_t\}_{t=1}^T$, where $\mathcal{F}_t = \sigma(Y_1, \dots, Y_t)$ is the information set at time t. The random variable of interest is Y_{t+1} conditional on $\mathcal{F}_t$, indicated by a subscript t to the (predictive) distributions, μ-densities and objects related to Q_{t+1}. The regions of interest $A_t \subseteq \mathcal{Y}$ are assumed to be $\mathcal{F}_t$-measurable.

This subsection aims to derive a uniformly most powerful (UMP) test for the hypotheses

$$\mathbb{H}_0 : p_t \mathbb{1}_{A_t} = f_{0t}\mathbb{1}_{A_t}, \quad \forall t \qquad \text{vs.} \qquad \mathbb{H}_1 : p_t \mathbb{1}_{A_t} = f_{1t}\mathbb{1}_{A_t}, \quad \forall t, \tag{6}$$

with f_{0t} and f_{1t}, fixed. Despite the densities f_{0t} and f_{1t} being fixed, the test concerns a multiple versus multiple hypothesis test due to the lacking specification of the densities outside the regions of interest A_t. Moreover, replacing A_t by A_{w_t}, for an $\mathcal{F}_t$-measurable weight function w_t, yields equivalent hypotheses in terms of w_t, specifically, $\mathbb{H}_0 : p_t w_t = f_{0t} w_t$, $\forall t$, versus $\mathbb{H}_1 : p_t w_t = f_{1t} w_t$, $\forall t$. Consequently, the UMP test for these hypotheses is equivalent to that for (6).

Theorem 3 reveals that (6) admits a UMP test, reducing to the Neyman and Pearson (1933) lemma when $A_t = \mathcal{Y}$, $\forall t$. A proof of this result is deferred to Appendix A.2.

Theorem 3 (Localized Neyman Pearson)**.** *For any given $\alpha \in (0,1)$, the UMP test of size α for testing problem* (6) *reads*

$$\phi_A^\flat(\mathbf{y}) = \begin{cases} 1, & \textit{if } \lambda(\mathbf{y}) > c \\ \gamma, & \textit{if } \lambda(\mathbf{y}) = c \\ 0, & \textit{if } \lambda(\mathbf{y}) < c, \end{cases} \qquad \lambda(\mathbf{y}) := \frac{f_{1,A}^\flat(\mathbf{y})}{f_{0,A}^\flat(\mathbf{y})}, \quad f_{j,A}^\flat(\mathbf{y}) := \prod_{t=0}^{T-1} f_{jt,A_t}^\flat(y_{t+1}), \quad j \in \{0,1\},$$

where $\phi_A^\flat : \mathcal{Y}^T \to [0,1]$ denotes a test function specifying the rejection probability, c is the largest constant such that $\mathrm{F}_{0,A}^\flat\big(\lambda(\mathbf{y}) \geq c\big) \geq \alpha$ and $\mathrm{F}_{0,A}^\flat\big(\lambda(\mathbf{y}) \leq c\big) \geq 1-\alpha$, and $\gamma \in [0,1]$ is such that $\alpha = \mathrm{F}_{0,A}^\flat\big(\lambda(\mathbf{y}) > c\big) + \gamma \mathrm{F}_{0,A}^\flat\big(\lambda(\mathbf{y}) = c\big)$.

For $T \equiv 1$, the test reduces to the UMP test of Holzmann and Klar (2017b). Corollary 1 reveals that it can alternatively be formulated in terms of the CSL introduced by Diks et al.

(2011). Corollary 2 endorses that conditioning does not bear a UMP test. The proofs of Corollaries 1 and 2 are deferred to Appendices B.2 and B.3, respectively.

Corollary 1. *An alternative formulation of the UMP test for testing problem* (6) *is given by the test defined in Theorem 3 with* $\lambda(\mathbf{y})$ *replaced by* $\tilde{\lambda}^{\flat}(\mathbf{y}) := \sum_{t=0}^{T-1} \big(LogS^{\flat}_{A_t}(f_{1t}, y_{t+1}) - LogS^{\flat}_{A_t}(f_{0t}, y_{t+1})\big)$*, i.e., in terms of the CSL.*

Corollary 2. *For testing problem* (6)*, the test defined in Theorem 3 with* $\lambda(\mathbf{y})$ *replaced by* $\tilde{\lambda}^{\sharp}(\mathbf{y}) := \sum_{t=0}^{T-1} \big(LogS^{\sharp}_{A_t}(f_{1t}, y_{t+1}) - LogS^{\sharp}_{A_t}(f_{0t}, y_{t+1})\big)$ *is not UMP.*

Insisting on fully specified models under $\mathbb{H}_0$ and $\mathbb{H}_1$ as in the classical Neyman and Pearson (1933) framework, even if only on the restriction to A, can be too demanding in practice. In the empirical applications in Section 4, for instance, we consider a hypothesis where, under the null, two models are 'equally wrong'. Then, generally, no UMP test is available, motivating the adoption of alternative scoring rules in addition to the CSL.

3.5 Related weighted scoring rules

In this subsection, we compare our approach to three alternative localization procedures. First, we consider Holzmann and Klar (2017a), who base their procedure on the conditional distribution $\mathrm{F}^{\sharp}_{w}$. As $\mathrm{P}^{\sharp}_{w} = \mathrm{F}^{\sharp}_{w}$ if and only if $\mathrm{P} \overset{A_w}{=} \mathrm{F}$ does *not* hold, the score divergence $\mathbb{D}_{S^{\sharp}_{w}}(\mathrm{P}\|\mathrm{F}) = (1 - \bar{P}_w)\mathbb{D}_S(\mathrm{P}^{\sharp}_{w}\|\mathrm{F}^{\sharp}_{w})$ generally fails to satisfy condition (ii) of Definition 2, unless $\bar{P}_w = \bar{F}_w$, hence is not a local divergence; see Example F.1 for a specific case. To resolve this, Holzmann and Klar (2017a) add an auxiliary scoring rule s, enforcing the score divergence to be zero if and only $\mathrm{P} \overset{A_w}{=} \mathrm{F}$. Example 5 describes their composite scoring rule, for which the score divergence is a local divergence but not in general a localized divergence.

Example 5. *Holzmann and Klar (2017a) propose a class of weighted scoring rules defined as* $\tilde{S}_{w,s}(\mathrm{F}, y) := S^{\sharp}_{w}(\mathrm{F}, y) + w(y)s(\mathrm{B}_{1-\bar{F}_w}, 1) + \big(1 - w(y)\big)s(\mathrm{B}_{1-\bar{F}_w}, 0)$*, assuming* $\bar{F}_w < 1$*, and where* B_θ *denotes the Bernoulli(*θ*) distribution with pmf* $b(z;\theta) = \theta^z(1-\theta)^{1-z}$*,* $z \in \{0, 1\}$*.*

For any s, $\tilde{S}_{w,s}$ is strictly locally proper (Holzmann and Klar 2017a, Theorem 2) and hence $\mathbb{D}_{\tilde{S}_{(w,s)}}(\mathrm{P}\|\mathrm{F}) = (1-\bar{P}_w)\mathbb{D}_S(\mathrm{P}^{\sharp}_w\|\mathrm{F}^{\sharp}_w) + \mathbb{D}_s(\mathrm{B}_{1-\bar{P}_w}\|\mathrm{B}_{1-\bar{F}_w})$ is a local divergence, but generally not a localized divergence due to the dependence on $\mathbb{D}_s$. Moreover, $\mathbb{D}_S(\mathrm{P}^{\sharp}_w\|\mathrm{F}^{\sharp}_w) = 0$ if P *and* F *are proportional on A_w; then, $\mathbb{D}_{\tilde{S}_{(w,s)}}(\mathrm{P}\|\mathrm{F}) = \mathbb{D}_s(\mathrm{B}_{1-\bar{P}_w}\|\mathrm{B}_{1-\bar{F}_w})$, which depends only on the auxiliary scoring rule.*

Holzmann and Klar (2017a) provide two specific choices for s: (i) $\mathrm{slog}(\mathrm{B}_{1-\bar{F}_w}, z) := z\log(1-\bar{F}_w) + (1-z)\log\bar{F}_w$ and (ii) $\mathrm{sbar}(\mathrm{B}_{1-\bar{F}_w}, z) := z(\log(1-\bar{F}_w)+1) - (1-\bar{F}_w)$. The combination $S = \mathrm{LogS}$ and $s = \mathrm{slog}$ recovers $\mathrm{LogS}^{\flat}_w$ and hence a localized divergence, whereas $S = \mathrm{LogS}$ and $s = \mathrm{sbar}$ leads to the weighted likelihood score by Pelenis (2014), for which the score divergence $\mathbb{D}_{\mathrm{pwl}_w}(\mathrm{P}\|\mathrm{F}) = \mathbb{D}_{\mathrm{LogS}}(\mathrm{P}^{\flat}_w\|\mathrm{F}^{\flat}_w) - \bar{P}_w\log\frac{\bar{P}_w}{\bar{F}_w} + (1-\bar{P}_w)\log\frac{1-\bar{P}_w}{1-\bar{F}_w} + (\bar{P}_w - \bar{F}_w)$ is *not* a localized divergence of $\mathbb{D}_{\mathrm{LogS}}$.

The arbitrariness of s allows for weighted scoring rules beyond the scope of (generalized) censored scoring rules. For example, $S = \mathrm{QS}$ and $s = \mathrm{slog}$, yields $\mathbb{D}_{\mathrm{QS}_{(w,\mathrm{slog})}}(p\|f) = (1-\bar{P}_w)\|p^{\sharp}_w - f^{\sharp}_w\|^2_2 + \mathbb{D}_{\mathrm{slog}}(\mathrm{B}_{1-\bar{P}_w}\|\mathrm{B}_{1-\bar{F}_w})$. A key distinction from $\mathbb{D}_{\mathrm{QS}^{\flat}_{w,\mathrm{H}}}(p\|f) = \|p^{\flat}_{w,\mathrm{H}} - f^{\flat}_{w,\mathrm{H}}\|^2_2$ is the role of the *auxiliary* KL-divergence, $\mathbb{D}_{\mathrm{slog}}$ in $\mathbb{D}_{\mathrm{QS}_{(w,\mathrm{slog})}}$, which will become *dominant* for large $\bar{F}_w$. Censoring is also not nested within the framework of Holzmann and Klar (2017a). To see this, note that unlike censoring (see Example 3), $\tilde{S}_{w,s}(\mathrm{F}, y)$ in Example 5 enforces $\tilde{S}_{w,s}(\mathrm{F}, y)$ to be only a function of $\bar{F}_w$ if $y \in A^c_w$.

The second comparison concerns the threshold weighted kernel score $\mathrm{tw}S_\rho$ introduced by Allen et al. (2023), generalizing the twCRPS by introducing the kernel $\rho(v(y), v(y'))$ based on a *measurable chaining function* $v : \mathcal{Y} \mapsto \mathcal{Y}$. With formalities presented in their Propositions 4.4 and 4.5, the $\mathrm{tw}S_\rho$ is strictly locally proper if v is injective on A_w and $\rho(v(y), v(\cdot)) = \rho(v(y'), v(\cdot))$, $\forall y, y' \in A^c_w$. For indicator weight functions, both conditions are easily satisfied with $v(y) = y\mathbb{1}_{A_w}(y) + y_0\mathbb{1}_{A^c_w}(y)$, for any $y_0 \in \mathcal{Y}$ such that $\mathrm{F}_w(y_0) = 0$. In that case, the $\mathrm{tw}S_\rho$ reduces to the $S^{\flat}_{\rho,y_0}$ score given in Example 4.

Our censoring approach also admits center indicator functions with two pivotal points, as in Example D.2; these are not considered in the kernel score framework of Allen et al. (2023). For non-indicator weight functions, Allen et al. (2023) provide a specific multivariate example that meets their requirements for strict local propriety, and which can be extended to other non-indicator weight functions that are specified as a product of marginal weight functions, including the multivariate logistic weight functions considered here (see Section E.4 for details). For more general non-indicator weight functions, specifying the chaining function as required for the approach of Allen et al. (2023) is a less trivial task. Our procedure foregoes specifying the chaining function and, moreover, is not restricted to kernel scores.

Third and finally, we compare with Mitchell and Weale (2023), who, for real-valued, unimodal densities and with the center as the region of interest, also consider censored density forecasts based on LogS. However, unlike our setting, they do *not* aim to evaluate multiple candidate densities. As illustrated by Example 6, the dependence of their region of interest on the candidate distribution renders the resulting scoring rule improper, thus not suitable for our aim of evaluating multiple density forecasts.

Example 6. *Mitchell and Weale (2023) consider the alternative censored likelihood score* $\mathrm{Log}_\alpha^{\mathrm{MW}}(f, y) := \log f(y)\mathbb{1}_{A(F;\alpha)}(y) + \log(\alpha)\mathbb{1}_{A(F;\alpha)^c}(y)$*, where* $A(F;\alpha)$ *is the central region of interest. A key difference with the censored likelihood score* $\log f_A^\flat(y)$ *is the dependence of* $A(F;\alpha)$ *on* f*, by which* $\mathrm{Log}_\alpha^{\mathrm{MW}}(f, y)$ *is improper. Indeed, for symmetric densities,* $A(F;\alpha) = [F^{-1}(\alpha/2), F^{-1}(1-\alpha/2)]$*. Then, letting* p *and* f *be the* $\mathcal{N}(0,1)$ *and* $\mathcal{N}(0,\frac{1}{2})$ *density, respectively, we have* $\mathbb{E}_p\mathrm{LogS}_\alpha^{\mathrm{MW}}(p, Y) - \mathbb{E}_p\mathrm{LogS}_\alpha^{\mathrm{MW}}(f, Y) < 0$*, for all* $\alpha > \alpha_0$ *with* $\alpha_0 \approx 0.052$*.*

4 EMPIRICAL PERFORMANCE

We assess the empirical performance of the censoring approach to focus scoring rules on regions of interest by evaluating its ability to discriminate between different forecast methods. We compare the performance of censored scoring rules with conditional scoring rules and with the composite scoring rules proposed by Holzmann and Klar (2017a). For the latter, we augment the conditional scoring rule with the auxiliary rule sbar or slog, see Section 3.5. We consider applications in financial risk management, macroeconomics, and climate, evaluating the focused scoring rules in a similar manner, as described below.

Following Giacomini and White (2006), we treat all components underlying a density forecast, including its estimation procedure, as integral parts of the forecast method. Let $\hat{f}_t$ and $\hat{g}_t$ denote density forecasts resulting from competing methods, each estimated with a rolling window of length m. We test the null hypothesis of equal predictive ability, $\mathbb{H}_0 : \mathbb{E}_{p_t} S_w(\hat{f}_t, Y_{t+1}) = \mathbb{E}_{p_t} S_w(\hat{g}_t, Y_{t+1}), \forall t$, by means of the DM type test statistic $t_{m,n} := \frac{1}{n} \sum_{t=m}^{T-1} \left(S_w(\hat{f}_t, Y_{t+1}) - S_w(\hat{g}_t, Y_{t+1}) \right) / \sqrt{\hat{\sigma}^2_{m,n}/n}$, where $n = T - m$ is the number of observations used for evaluation and $\hat{\sigma}^2_{m,n}$ is a heteroskedasticity and autocovariance-consistent (HAC) variance estimator.

The null hypothesis is equivalent to $\mathbb{D}_{S_w}(p_t \| \hat{f}_t) = \mathbb{D}_{S_w}(p_t \| \hat{g}_t)$ and is rejected if it is sufficiently unlikely that the weighted score divergences from p_t to $\hat{f}_t$ and p_t to $\hat{g}_t$ coincide. This null differs from that in (6), obstructing theoretical results on the power properties of the test. However, because censoring preserves more information than conditioning, we generally expect higher power for test statistics based on censored scoring rules. This is supported by the Monte Carlo results in Appendix G. In these simulation experiments, we also find that the composite scoring rules of Holzmann and Klar (2017a) generally result in test statistics with comparable power. However, further analysis shows that this cannot be attributed to the localization method but rather to the advantageous properties of the logarithmic score that forms the basis for the auxiliary rules sbar and slog.

In practice, including the empirical applications discussed below, one commonly has more than two candidate forecast methods. We therefore start with a collection $\mathcal{M}_0$ of forecast methods, and then use the iterative procedure proposed by Hansen et al. (2011) to reduce $\mathcal{M}_0$ to a Model Confidence Set (MCS) of methods for which the null of equal predictive ability cannot be rejected. Elimination in round k is based on the statistic TR:= $\max_{i,j \in \mathcal{M}_k} |t_{m,n}^{(i,j)}|$, where $t_{m,n}^{(i,j)}$ corresponds to the pairwise $t_{m,n}$-statistic between forecast methods i and j introduced above. Favorable power properties of censoring in the pairwise tests intuitively accelerate elimination in the MCS procedure, resulting in smaller p-values and, consequently, reduced MCS cardinality. We present MCS results at the 0.90 confidence level, with results for the 0.75 confidence level deferred to Appendix I.1, using a block bootstrap with block length $b = 5$ and $B = 10,000$ replications, unless stated otherwise. Our results are robust to variations in these parameters, see Table I.3.

In each application below, the unweighted scoring rules are given by LogS, QS, SphS and S_{ρ_1}, with kernel $\rho_1(\mathbf{x}, \mathbf{x}') := \|\mathbf{x} - \mathbf{x}'\|$, where $\|\cdot\|$ the Euclidean norm, i.e., S_{ρ_1} is the Energy Score that reduces to the CRPS in univariate examples. These scoring rules are localized by (i) conditioning, (ii) censoring, (iii) conditioning with sbar and (iv) conditioning with slog. The twCRPS and twS_{ρ_1} are included only in cases where they differ from CRPS$^\flat$. Hence, we consider 16 or 17 weighted scoring rules per application. There are $|\mathcal{M}_0| = 6$ candidate forecast methods, with specifications differing by application. For reproducibility, Appendix H includes details on the specification of the individual methods. Moreover, Appendix I reports the MCS p-values per individual scoring rule and weight function underlying the summary results presented in this section.

4.1 Financial risk management

Measuring and forecasting the downside risk of asset returns is crucial in risk management, particularly for compliance with regulatory requirements related to measures such as Value-at-Risk and Expected Shortfall. We evaluate density forecasts constructed for daily log-returns y_t on the S&P500 index over the period from January 2, 1996, to December 30, 2022 ($6,777$ observations), sourced from Yahoo Finance. To achieve the required focus on the left tail of the density forecast, we use the indicator weight function $I_{\mathrm{L}}(y_t; \hat{r}_t^q)$, where $\hat{r}_t^q$ denotes the q-th empirical quantile of y_t, based on the same fixed rolling window of length $m = 1,000$ used for estimation of the forecast methods.

All forecast methods used conform to $Y_t | \mathcal{F}_{t-1} \sim \mathcal{D}(\mu, \sigma_t^2, \boldsymbol{\vartheta})$, denoting a parametric family of distributions with constant mean μ, time-varying variance σ_t^2, and any additional parameters collected in $\boldsymbol{\vartheta}$. Although we tested AR(1) and AR(5) models for the conditional mean, neither improved significantly over a constant mean. We consider three conditional variance models: GARCH, threshold GARCH (TGARCH) and realized GARCH (RGARCH), proposed by Bollerslev (1986), Glosten et al. (1993) and Hansen et al. (2012), respectively. We combine each of the volatility models with standard normal and Student-t_ν distributions. Density forecasts are constructed for horizons $\tau = 1$ and 5 days.

Table 2 reveals stark differences in the cardinality of MCS$^\flat$ and MCS$^\sharp$, particularly at $\tau = 1$. In case no correction is applied to the conditional scoring rules, MCS$^\flat$ is strictly smaller than MCS$^\sharp$ in 75% of all replications, while MCS$^\sharp$ contains more than twice the number of methods compared to MCS$^\flat$ on average. These results moderate when the conditional scoring rules are appended with a Holzmann-Klar correction term. Nevertheless, MCS$^\flat$ remains strictly smaller than MCS$^\sharp$ in nearly 40% of the cases, with an average difference in cardinality of 20%. For $\tau = 5$, the differences become smaller but remain in favor of censoring, ranging between 10 and 30%.

We extend the univariate setting to the evaluation of bivariate density forecasts for the vector of log-returns $\mathbf{y}_t \in \mathbb{R}^2$ for the Energy Select Sector SPDR Fund (XLE) and Financial Select Sector SPDR Fund (XLF), for the period January 5, 1999 to December 29, 2023 (6,218 observations). We consider the approximated bivariate empirical q-th quantile of $\mathbf{y}_t$ given by $\tilde{\mathbf{r}}_t^q := (\hat{r}_{1,t}^{q_2}, \hat{r}_{2,t}^{q_2})$, with $q_2 = \sqrt{q}$ to formulate the weight functions $I_{\mathrm{L}}^2(\mathbf{y}_t; \tilde{\mathbf{r}}_t^q)$ and $\Lambda_{a,\mathrm{L}}^2(\mathbf{y}_t; \tilde{\mathbf{r}}_t^q)$, with $a = 3$, while having verified stability of results for $a \in \{2, 4\}$. The individual mean (μ_i) and volatility $(\sigma_{i,t}^2)$ specifications are as in the S&P500 models. We use the Dynamic Conditional Correlation (DCC) approach of Engle (2002) to map the univariate specifications into a bivariate conditional covariance matrix. The univariate distributions are replaced by bivariate standard normal and Student-t_ν distributions.

Table 2 shows that the results for the bivariate density forecasts corroborate the main findings for the univariate S&P500 application. First, the MCS obtained with censored scoring rules is not larger than the MCS resulting from conditional scoring rules in the large majority of cases, namely between 62% and 92%. For about one-third of cases, $\mathrm{MCS}^\flat$ is even strictly smaller than $\mathrm{MCS}^\sharp$. Second, the former percentages are hardly affected by adding the Holzmann-Klar correction terms to the conditional scoring rule, while the latter decline but not substantially. The largest reduction occurs when using the slog correction term for the scoring rules focused with weight function $I_{\mathrm{L}}^2(\mathbf{y}_t; \tilde{\mathbf{r}}_t^q)$, from 33% to 17% for $\tau = 1$. Hence, the correction terms result in equally large MCSs for the censored and augmented conditional scoring rules more frequently. Closer inspection reveals that their compositions almost always are identical as well. Third, the results for the logistic weight function $\Lambda_{a,\mathrm{L}}^2(\mathbf{y}_t; \tilde{\mathbf{r}}_t^q)$ with $a = 3$ do not differ much from those for the indicator weight function $I_{\mathrm{L}}^2(\mathbf{y}_t; \tilde{\mathbf{r}}_t^q)$, which is recovered as $a \to \infty$. However, the discrepancies in cardinality are somewhat moderated, particularly for $\tau = 5$, aligning with the observation by Diks et al. (2011) that the score distribution of the weighted scoring rules becomes more alike for smaller values of a. Finally, in contrast to the univariate setting, the (relative) performance of the censored scoring rules does not decline at longer forecast horizons. If anything, the percentages and average cardinality ratio improve for $\tau = 5$ compared to $\tau = 1$.

Table 2: MCS cardinality of censored and (un)corrected conditional scoring rules

			no correction			sbar			slog		
Sec.	w_t	τ	$\leq$	$<$	$\sharp/\flat$	$\leq$	$<$	$\sharp/\flat$	$\leq$	$<$	$\sharp/\flat$
4.1	$I_{\mathrm{L}}(y_t; \hat{r}_t^q)$	1	96%	75%	2.38	71%	38%	1.20	71%	38%	1.20
		5	62%	29%	1.29	54%	25%	1.08	62%	25%	1.10
	$I_{\mathrm{L}}^2(\mathbf{y}_t; \tilde{\mathbf{r}}_t^q)$	1	62%	33%	1.20	67%	25%	1.06	62%	17%	0.98
		5	92%	29%	1.23	92%	25%	1.28	92%	25%	1.26
	$\Lambda_{3,\mathrm{L}}^2(\mathbf{y}_t; \tilde{\mathbf{r}}_t^q)$	1	62%	33%	1.06	62%	21%	0.95	62%	21%	0.95
		5	71%	33%	1.16	83%	25%	1.16	88%	25%	1.18
4.2	$I_{\mathrm{C}}(y_t; 2, r_1)$	6	100%	92%	2.67	100%	83%	2.42	100%	67%	1.69
		24	92%	75%	2.27	58%	42%	1.24	67%	25%	1.23
	$I_{\mathrm{C}}^c(y_t; 2, r_1)$	6	100%	92%	2.04	83%	50%	1.13	67%	33%	1.19
		24	92%	75%	2.86	50%	33%	1.18	75%	17%	1.04
4.3	$I_{\mathrm{R}}(y_t; \hat{r}_t^q)$	1	83%	58%	1.92	92%	67%	1.92	83%	33%	1.29
		3	75%	46%	1.54	79%	42%	1.40	75%	4%	0.96
	$I_{\mathrm{C}}(y_t; 18, r_2)$	1	100%	58%	2.21	100%	42%	1.42	100%	42%	1.42
		3	100%	58%	1.58	100%	0%	1.00	100%	0%	1.00
Total average			85%	56%	1.82	78%	37%	1.32	79%	27%	1.18

NOTE: This table presents changes in cardinality of the MCS in absolute and relative terms, at 0.90, across different forecast horizons τ, corresponding to the forecasting applications in risk management (Section 4.1), inflation (Section 4.2) and temperature (Section 4.3). sbar and slog refer to the correction terms for conditional scoring rules proposed by Holzmann and Klar (2017a). Columns labeled $\leq$ ($<$) display the percentage of cases where MCS$^\flat$ contains (strictly) fewer forecast methods than MCS$^\sharp$ and the column labeled $\sharp/\flat$ reports the ratio $|\mathrm{MCS}^\sharp|/|\mathrm{MCS}^\flat|$. Each result represents an average over a set of scoring rules $S \in \{\mathrm{LogS}, \mathrm{QS}, \mathrm{SphS}, \mathrm{CRPS}/S_{\rho_1}\}$ and quantile levels $q \in \{0.01, 0.05, 0.10, 0.15, 0.20, 0.25\}$ or levels $r_1 \in \{1, 1.5, 2\}$ and $r_2 \in \{1, 2, 4\}$. The empirical q-th quantiles $\hat{r}_t^q$ of y_t are based on the parameter estimation window of m observations, and $\tilde{\mathbf{r}}_t^q := (\hat{r}_{1,t}^{q_2}, \hat{r}_{2,t}^{q_2})$, with $q_2 = \sqrt{q}$, approximates a bivariate empirical q-th quantile of $\mathbf{y}_t$. The p-values are obtained via a block bootstrap of $B = 10{,}000$ replications, with block length $b = 5$, or $b = 200$ for the climate data. Complete MCS details and associated p-values are provided in Appendix I.

4.2 Macroeconomics

We next consider forecasting inflation, a subject with a long history in macroeconomics that recently has regained prominence. Given that many central banks, including the

Federal Reserve System and the European Central Bank target an annual inflation rate of 2%, we focus on the central range $A_r = (2 - r, 2 + r)$, where $r > 0$, by using the weight function $I_{\mathrm{C}}(y_t; 2, r)$. To address policymakers' concerns for deviations beyond A_r, termed 'Inflation at Risk' (Lopez-Salido and Loria 2020), we additionally consider its complement $I^c_{\mathrm{C}}(y_t; 2, r) = 1 - I_{\mathrm{C}}(y_t; 2, r)$. For the CRPS, we adopt two pivotal points for $I_{\mathrm{C}}(y_t; 2, r)$,while using $\ell = 2$ for its complement, i.e., treating non-tail observations to be on target. Following Stock and Watson (2002), among many others, we construct direct forecasts for annualized τ-month inflation rates $y^{\tau}_{t+\tau} = (1,200/\tau) \log\big(P_{t+\tau}/P_t\big)$, where P_t denotes the U.S. consumer price index (CPI) in month t, for horizons $\tau = 6$ and 24. The sample period runs from January 1960 until December 2015 (672 observations), where density forecasts are obtained for the final 180 months in this time frame.

We consider forecast methods that aim to exploit the 'data-rich environment' in macroeconomic forecasting, with many potentially relevant predictors especially for inflation. Here we follow Medeiros et al. (2021) by using the same 122 variables from the FRED-MD database ($\mathbf{x}_t$). Each of the forecast methods can be represented as $y^{\tau}_{t+\tau} = \mu^{\tau}_{t+\tau}(\mathbf{x}_t) + u^{\tau}_{t+\tau}$, where we consider the following subset of methods listed by Medeiros et al. (2021) for the conditional mean $\mu^{h}_{t+\tau}$: Random Walk, Auto-Regressive model (AR), Bagging, Complete Subset Regression (CSR), Least Absolute Shrinkage and Selection Operator (LASSO), and Random Forest. The error $u^{\tau}_{t+\tau}$ is assumed to follow a two-piece normal distribution, congruent with the statistical model underlying the fan charts published by the Bank of England (Clements 2004; Mitchell and Hall 2005; Gneiting and Ranjan 2011).

The summary results presented in Table 2 reveal a distinct and pronounced preference for censoring, again especially when no correction is applied to the conditional scoring rules. In that case the cardinalities of $\mathrm{MCS}^{\flat}$ are almost always (weakly) smaller than those of $\mathrm{MCS}^{\sharp}$. The relative increase in set cardinality when opting for conditioning over censoring is substantial at more than 100%. Interestingly, the censored scoring rules outperform the conditional rules not only when focusing on the central range around the inflation target of 2%, but also when the interest is on the complementary 'Inflation at Risk' region of more extreme inflation rates. Finally, also in this application the Holzmann-Klar corrections to the conditional scoring rules improve their (relative) performance, although the MCS cardinality results largely remain favorable to censoring.

4.3 Climate

We generate density forecasts for Dutch daily average temperature data, focusing on high temperatures via the weight function $I_{\mathrm{R}}(y_t; \hat{r}_t^q)$ and temperatures near the optimal temperature for tuber growth, approximately 18 degrees Celsius (Struik 2007, Section 18.5.5), using $I_{\mathrm{C}}(y_t; 18, r)$. Extending the data and methodology of Franses et al. (2001) and Tol (1996), we focus on volatility clustering and asymmetries in the relationship between past temperature and volatility, along with seasonal variations in the mean and variance. We use daily observations for the period from February 1, 2003, to January 31, 2023, with a rolling estimation window of $m = 2,922$ days (or 8 years). Our volatility models closely follow the GARCH, QGARCH-I, and QGARCH-II specifications as in Franses et al. (2001), but with alterations in seasonal trend estimation. The GARCH-type models are combined with a standard normal and Student-t_ν distribution.

Using the right-tail weight function $I_{\mathrm{R}}(y_t; \hat{r}_t^q)$ to focus on high daily temperatures, we find results exhibiting pronounced parallels with the left-tail risk management application. In particular, as seen in Table 2, the cardinalities of the censored MCSs are typically much smaller than their uncorrected conditional counterparts for $\tau = 1$ day-ahead forecasts; the differences diminish at the longer forecast horizon $\tau = 3$ or when a Holzmann-Klar correction is appended to the conditional scoring rule.

Focusing on the central range around 18 degrees Celsius with the weight function $I_{\mathrm{C}}(y_t; 18, r)$, we find that there are no instances where conditioning leads to a smaller

MCS for $\tau = 3$ and almost no such cases for $\tau = 1$. Relative to inflation, there is a notable increase in cases where the MCSs possess identical cardinality, also reflected in the smaller ratios $|\text{MCS}^\sharp|/|\text{MCS}^\flat|$.

5 CONCLUSION

In this paper, we propose censoring as a focusing device to accommodate the fact that in many applications, forecasters are particularly interested in specific areas of the outcome space. We demonstrate that a key advantage of censoring is that applying scoring rules to censored distributions results in strictly locally proper scoring rules. To the best of our knowledge, we are the first to derive a transformation of the original scoring rule that preserves both the score divergence and strict propriety, and features high flexibility, being applicable across varied scoring rules, weight functions, and outcome spaces. For specific choices, the censored scoring rule yields intuitively appealing rules apt for practical use. For instance, we recover the twCRPS for tail indicators, while extending its strict local propriety to other weight functions. Our second theoretical contribution, a generalization of the Neyman Pearson lemma, revolves around the censored likelihood score. We have shown that the UMP test of the localized Neyman Pearson hypothesis is a censored likelihood ratio test, reducing to the original lemma if the weight function is positive for all outcomes. By contrast, the conditional likelihood ratio test is not UMP.

We demonstrate the practical relevance of censoring with empirical applications in financial risk management, macroeconomics, and climate. We use the size of the Model Confidence Set (MCS) to gauge the scoring rule's ability to discriminate between competing forecast methods. A common finding in the applications is that the censored MCS is (strictly) smaller than the conditional MCS in a large majority of cases, and often the difference in cardinality is substantial. This conclusion holds across different areas of interest, in particular whether attention is focused on the central range of the distribution or on one (or both) of the tails.

SUPPLEMENTARY MATERIAL

All proofs and additional theoretical results, the Monte Carlo analysis, and full tables on the empirical performance are provided in an online supplementary document. (.pdf)

References

Adrian, T., N. Boyarchenko, and D. Giannone (2019), "Vulnerable Growth", *American Economic Review*, *109*(4), 1263–1289.

Allen, S., D. Ginsbourger, and J. Ziegel (2023), "Evaluating Forecasts for High-Impact Events Using Transformed Kernel Scores", *SIAM/ASA Journal on Uncertainty Quantification*, *11*(3), 906–940.

Amisano, G. and R. Giacomini (2007), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests", *Journal of Business & Economic Statistics*, *25*(2), 177–190.

Bernoulli, D. (1760), "Essai d'une Nouvelle Analyse de la Mortalite Causee par la Petite Verole, et des Avantages de l'Inoculation Pour la Prevenir", *Histoire de l'Acad.*, *Roy. Sci.(Paris) avec Mem*, 1–45.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, *31*(3), 307–327.

Borowska, A., L. Hoogerheide, S. J. Koopman, and H. K. Van Dijk (2020), "Partially Censored Posterior for Robust and Efficient Risk Evaluation", *Journal of Econometrics*, *217*(2), 335–355.

Bregman, L. (1967), "The Relaxation Method of Finding the Common Point of Convex Sets and its Application to the Solution of Problems in Convex Programming", *USSR Computational Mathematics and Mathematical Physics*, *7*(3), 200–217.

Brehmer, J. R. and T. Gneiting (2020), "Properization: Constructing Proper Scoring Rules via Bayes Acts", *Annals of the Institute of Statistical Mathematics*, *72*(3), 659–673.

Brier, G. W. (1950), "Verification of Forecasts Expressed in Terms of Probability", *Monthly Weather Review*, *78*(1), 1–3.

Clements, M. P. (2004), "Evaluating the Bank of England Density Forecasts of Inflation", *The Economic Journal*, *114*(498), 844–866.

Cont, R., R. Deguest, and G. Scandolo (2010), "Robustness and Sensitivity Analysis of Risk Measurement Procedures", *Quantitative Finance*, *10*(6), 593–606.

Dawid, A. P. (1984), "Statistical Theory: The Prequential Approach", *Journal of the Royal Statistical Society. Series A (General)*, *147*(2), 278–292.

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

Diebold, F. X. and R. S. Mariano (2002), "Comparing Predictive Accuracy", *Journal of Business & Economic Statistics*, *20*(1), 134–144.

Diks, C., V. Panchenko, and D. Van Dijk (2011), "Likelihood-based Scoring Rules for Comparing Density Forecasts in Tails", *Journal of Econometrics*, *163*(2), 215–230.

Eguchi, S. (1985), "A Differential Geometric Approach to Statistical Inference on the Basis of Contrast Functionals", *Hiroshima Mathematical Journal*, *15*(2), 341–391.

Ehm, W. and T. Gneiting (2012), "Local Proper Scoring Rules of Order Two", *The Annals of Statistics*, *40*(1), 609–637.

Engle, R. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models", *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Fissler, T., J. F. Ziegel, and T. Gneiting (2015). "Expected Shortfall is Jointly Elicitable with Value at Risk - Implications for Backtesting". DOI: 10.48550/ARXIV.1507.00244. Available at `https://arxiv.org/abs/1507.00244`.

Franses, P. H., J. Neele, and D. Van Dijk (2001), "Modeling Asymmetric Volatility in Weekly Dutch Temperature Data", *Environmental Modelling & Software*, *16*(2), 131–137.

Giacomini, R. and H. White (2006), "Tests of Conditional Predictive Ability", *Econometrica*, *74*(6), 1545–1578.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *The Journal of Finance*, *48*(5), 1779–1801.

Gneiting, T. and A. E. Raftery (2007), "Strictly Proper Scoring Rules, Prediction, and Estimation", *Journal of the American Statistical Association*, *102*(477), 359–378.

Gneiting, T. and R. Ranjan (2011), "Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules", *Journal of Business & Economic Statistics*, *29*(3), 411–422.

Hansen, P. R., Z. Huang, and H. H. Shek (2012), "Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility", *Journal of Applied Econometrics*, *27*(6), 877–906.

Hansen, P. R., A. Lunde, and J. Nason (2011), "The Model Confidence Set", *Econometrica*, *79*(2), 453–497.

Holzmann, H. and B. Klar (2017a), "Focusing on Regions of Interest in Forecast Evaluation", *The Annals of Applied Statistics*, *11*(4), 2404–2431.

Holzmann, H. and B. Klar (2017b). "Weighted Scoring Rules and Hypothesis Testing". Available at `https://arxiv.org/abs/1611.07345v2`.

Iacopini, M., F. Ravazzolo, and L. Rossini (2023), "Proper Scoring Rules for Evaluating Density Forecasts with Asymmetric Loss Functions", *Journal of Business & Economic Statistics*, *41*(2), 482–496.

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

Lerch, S., T. L. Thorarinsdottir, F. Ravazzolo, and T. Gneiting (2017), "Forecaster's Dilemma: Extreme Events and Forecast Evaluation", *Statistical Science*, *32*(1), 106–127.

Lopez-Salido, D. and F. Loria (2020). "Inflation at Risk". Finance and Economics Discussion Series 2020-013. Washington: Board of Governors of the Federal Reserve System. Avalaible at `https://doi.org/10.17016/FEDS.2020.013`.

Medeiros, M. C., G. F. R. Vasconcelos, A. Veiga, and E. Zilberman (2021), "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods", *Journal of Business & Economic Statistics*, *39*(1), 98–119.

Mitchell, J. and S. G. Hall (2005), "Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR 'Fan' Charts of Inflation", *Oxford Bulletin of Economics and Statistics*, *67*(s1), 995–1033.

Mitchell, J. and M. Weale (2023), "Censored Density Forecasts: Production and Evaluation", *Journal of Applied Econometrics*, *38*(5), 714–734.

Neyman, J. and E. Pearson (1933), "IX. On the Problem of the Most Efficient Tests of Statistical Hypotheses", *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *231*(694-706), 289–337.

Ovcharov, E. Y. (2018), "Proper Scoring Rules and Bregman Divergence", *Bernoulli*, *24*(1), 53–79.

Painsky, A. and G. W. Wornell (2020), "Bregman Divergence Bounds and Universality Properties of the Logarithmic Loss", *IEEE Transactions on Information Theory*, *66*(3), 1658–1673.

Patton, A. J. (2020), "Comparing Possibly Misspecified Forecasts", *Journal of Business & Economic Statistics*, *38*(4), 796–809.

Pelenis, J. (2014). "Weighted scoring rules for comparison of density forecasts on subsets of interest". Available at `https://sites.google.com/site/jpelenis/`.

Steinwart, I. and J. F. Ziegel (2021), "Strictly proper kernel scores and characteristic kernels on compact spaces", *Applied and Computational Harmonic Analysis*, *51*, 510–542.

Stock, J. H. and M. W. Watson (2002), "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business & Economic Statistics*, *20*(2), 147–162.

Struik, P. C. (2007). "Chapter 18 - Responses of the Potato Plant to Temperature". In D. Vreugdenhil, J. Bradshaw, C. Gebhardt, F. Govers, D. K. L. Mackerron, M. A. Taylor, and H. A. Ross (Eds.), *Potato Biology and Biotechnology*, pp. 367–393. Amsterdam: Elsevier Science B.V.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, *26*(1), 24–36.

Tol, R. S. (1996), "Autoregressive Conditional Heteroscedasticity in Daily Temperature Measurements", *Environmetrics*, *7*(1), 67–75.

Supplementary Material for "Localizing Strictly Proper Scoring Rules"

Ramon F. A. de Punder
Department of Quantitative Economics
University of Amsterdam and Tinbergen Institute

Cees G. H. Diks*
Department of Quantitative Economics
University of Amsterdam and Tinbergen Institute

Roger J. A. Laeven
Department of Quantitative Economics
University of Amsterdam, CentER and EURANDOM

Dick J. C. van Dijk
Department of Econometrics
Erasmus University Rotterdam and Tinbergen Institute

May 26, 2025

Abstract

This supplementary material complements the main paper with proofs, additional derivations, Monte Carlo analyses and empirical results. Part A provides comprehensive proofs of the main theoretical results in Theorems 2 and 3. Part B contains additional proofs for other findings highlighted in the main text. Part C discusses an alternative formulation of the (generalized) censored scoring rules in terms of a randomization procedure. Part D elaborates on the complications and issues encountered when applying the censoring approach to distance-sensitive scoring rules. Part E presents derivations of the results for the semi-local scoring rules as summarized in Table 1 and specific results for kernel scores. Part F showcases an example of weighted scoring rules violating the conditions for rendering a local divergence. Part G presents the Monte Carlo simulation experiments, examining the size and power properties of Diebold-Mariano statistics based on censored scoring rules and on composite scoring rules following the approach of Holzmann and Klar (2017a). Finally, Parts H and I contain further details on the model specifications and the underlying results for the empirical applications.

---

*Corresponding author. Mailing Address: PO Box 15867, 1001 NJ Amsterdam, The Netherlands. Phone: +31 (0) 20 525 4252. Email: `C.G.H.Diks@uva.nl`.

Contents

A Proofs

A.1 Proof of Theorem 2

We start by showing that the generalized censored scoring rule $S^{\flat}_{\cdot,\cdot}$ in Definition 6 is strictly locally proper relative to $(\mathcal{P}, \mathcal{W}, \mathcal{H})$, assuming $\mathcal{W}$ and $\mathcal{H}$ satisfy Assumption 1. For clarity of exposition, we first prove the main ingredients of the proof via two lemmas and a corollary. We also reintroduce the extended notation for coinciding measures.

Lemma A1. *Consider the generalized censored scoring rule defined in Definition 6, based on the unweighted scoring rule* $S : \mathcal{P}^{\flat} \times \mathcal{Y} \to \bar{\mathbb{R}}$, *where* $\mathcal{P}^{\flat} = \{\mathrm{F}^{\flat}_{w,\mathrm{H}}, \mathrm{F} \in \mathcal{P}, w \in \mathcal{W}, \mathrm{H} \in \mathcal{H}\}$. *Then*

$$\int_{\mathcal{Y}} S^{\flat}_{w,\mathrm{H}}(\mathrm{F}, y)\mathrm{P}(\mathrm{d}y) = \int_{\mathcal{Y}} S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, y)\mathrm{P}^{\flat}_{w,\mathrm{H}}(\mathrm{d}y),$$

$\forall w \in \mathcal{W}$, $\mathrm{H} \in \mathcal{H} \subseteq \mathcal{P}$ *and* $\mathrm{F}, \mathrm{P} \in \mathcal{P}$.

Proof. The result follows by rewriting the integral on the left-hand side. Specifically, fix an arbitrary $w \in \mathcal{W}$, $\mathrm{H} \in \mathcal{H} \subseteq \mathcal{P}$ and $\mathrm{F}, \mathrm{P} \in \mathcal{P}$. Then

$$\begin{aligned}
\int_{\mathcal{Y}} S^{\flat}_{w,\mathrm{H}}(\mathrm{F}, y)\mathrm{P}(\mathrm{d}y) &= \int_{\mathcal{Y}} \Big(w(y)S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, y) + \big(1 - w(y)\big) \int_{\mathcal{Y}} S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, q)\mathrm{H}(\mathrm{d}q)\Big)\mathrm{P}(\mathrm{d}y) \\
&= \int_{\mathcal{Y}} w(y)S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, y)\mathrm{P}(\mathrm{d}y) + \int_{\mathcal{Y}} S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, q) \int_{\mathcal{Y}} \big(1 - w(y)\big)\mathrm{P}(\mathrm{d}y)\mathrm{H}(\mathrm{d}q) \\
&= \int_{\mathcal{Y}} S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, y)\mathrm{P}_w(\mathrm{d}y) + \int_{\mathcal{Y}} S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, y)\bar{P}_w\mathrm{H}(\mathrm{d}y) \\
&= \int_{\mathcal{Y}} S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, y)\big(\mathrm{P}_w(\mathrm{d}y) + \bar{P}_w\mathrm{H}(\mathrm{d}y)\big) \\
&= \int_{\mathcal{Y}} S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, y)\mathrm{P}^{\flat}_{w,\mathrm{H}}(\mathrm{d}y).
\end{aligned}$$

□

Lemma A2. *Consider a class of distributions $\mathcal{P}$, weight functions $\mathcal{W}$ and nuisance distributions $\mathcal{H}$ such that Assumption 1 is satisfied. Define the associated class of censored distributions $\mathcal{P}^\flat = \{\mathrm{F}^\flat_{w,\mathrm{H}}, \mathrm{F} \in \mathcal{P}, w \in \mathcal{W}, \mathrm{H} \in \mathcal{H}\}$ as in Definition 6. Then the equivalence*

$$\mathrm{P}^\flat_{w,\mathrm{H}}(E) = \mathrm{F}^\flat_{w,\mathrm{H}}(E),\ \forall E \in \mathcal{G} \iff \mathrm{P}(E \cap \{w > 0\}) = \mathrm{F}(E \cap \{w > 0\}),\ \forall E \in \mathcal{G},$$

holds $\forall w \in \mathcal{W}$, $\mathrm{H} \in \mathcal{H} \subseteq \mathcal{P}$ and $\mathrm{P}, \mathrm{F} \in \mathcal{P}$.

Proof. " $\implies$ " We start with the most challenging direction, for which Assumption 1 is of critical importance. Let $\tilde{E}$ be an element of $\mathcal{G}$ satisfying the conditions given in Assumption 1. First, note that $\mathrm{P}^\flat_{w,\mathrm{H}}(E) = \mathrm{F}^\flat_{w,\mathrm{H}}(E)$, $\forall E \in \mathcal{G}$, implies $\bar{P}_w = \bar{F}_w$, since $\forall E \in \mathcal{G}$ we have

$$\begin{aligned}
\mathrm{P}^\flat_{w,\mathrm{H}}(E) = \mathrm{F}^\flat_{w,\mathrm{H}}(E) &\implies \mathrm{P}^\flat_{w,\mathrm{H}}\left(E \cap \tilde{E}\right) = \mathrm{F}^\flat_{w,\mathrm{H}}\left(E \cap \tilde{E}\right) \\
&\implies \int_{\mathcal{Y}} (1-w)\,\mathrm{dP} \cdot \mathrm{H}\left(E \cap \tilde{E}\right) = \int_{\mathcal{Y}} (1-w)\,\mathrm{dF} \cdot \mathrm{H}\left(E \cap \tilde{E}\right) \\
&\implies \int_{\mathcal{Y}} (1-w)\,\mathrm{dP} \cdot \mathrm{H}\left(\tilde{E}\right) = \int_{\mathcal{Y}} (1-w)\,\mathrm{dF} \cdot \mathrm{H}\left(\tilde{E}\right) \\
&\implies \int_{\mathcal{Y}} (1-w)\,\mathrm{dP} = \int_{\mathcal{Y}} (1-w)\,\mathrm{dF},
\end{aligned}$$

where we have used the closure of σ-algebras under countable intersections. Then, we can exploit this equality to obtain, $\forall E \in \mathcal{G}$,

$$\begin{aligned}
\mathrm{P}^\flat_{w,\mathrm{H}}(E) = \mathrm{F}^\flat_{w,\mathrm{H}}(E) &\iff \int_{\mathcal{Y}} w(y)\mathbb{1}_{y \in E}\mathrm{P}(\mathrm{d}y) + \bar{P}_w\mathrm{H}(E) = \int_{\mathcal{Y}} w(y)\mathbb{1}_{y \in E}\mathrm{F}(\mathrm{d}y) + \bar{F}_w\mathrm{H}(E) \\
&\implies \int_{\mathcal{Y}} w(y)\mathbb{1}_{y \in E}\mathrm{P}(\mathrm{d}y) = \int_{\mathcal{Y}} w(y)\mathbb{1}_{y \in E}\mathrm{F}(\mathrm{d}y) \\
&\implies \mathrm{P}(E \cap \{w > 0\}) = \mathrm{F}(E \cap \{w > 0\}).
\end{aligned}$$

“ $\Longleftarrow$ ” The other direction is straightforward. Indeed, $\forall E \in \mathcal{G}$,

$$\begin{aligned}\mathrm{P}(E \cap \{w > 0\}) = \mathrm{F}(E \cap \{w > 0\}) &\implies \int_{\mathcal{Y}} w(y)\mathbb{1}_{y \in E}\mathrm{P}(\mathrm{d}y) = \int_{\mathcal{Y}} w(y)\mathbb{1}_{y \in E}\mathrm{F}(\mathrm{d}y) \\ &\implies \int_{\mathcal{Y}} (1-w)\mathrm{dP} = \int_{\mathcal{Y}} (1-w)\mathrm{dF},\end{aligned}$$

and, consequently, $\mathrm{P}^{\flat}_{w,\mathrm{H}}(E) = \mathrm{F}^{\flat}_{w,\mathrm{H}}(E),\ \forall E \in \mathcal{G}, \forall \mathrm{H} \in \mathcal{H}$. □

Corollary A3. *Consider a class of distributions $\mathcal{P}$, weight functions $\mathcal{W}$ and nuisance distributions $\mathcal{H}$ such that Assumption 1 is satisfied. Then, $\forall w \in \mathcal{W}$, the generalized censored scoring rule defined in Definition 6 is localizing $\forall \mathrm{H} \in \mathcal{H}$.*

Proof. Suppose that $\mathrm{P}(E \cap \{w > 0\}) = \mathrm{F}(E \cap \{w > 0\}),\ \forall E \in \mathcal{G}$. Then, by Lemma A2, $\mathrm{P}^{\flat}_{w,\mathrm{H}}(E) = \mathrm{F}^{\flat}_{w,\mathrm{H}}(E), \forall E \in \mathcal{G}$, whence it follows that $S^{\flat}_{w,\mathrm{H}}(\mathrm{P}, y) = S^{\flat}_{w,\mathrm{H}}(\mathrm{F}, y), \forall y \in \mathcal{Y}$. □

We now turn to the main body of the proof of Theorem 2. The definition of a strictly locally proper scoring rule (Definition 4) and the underlying concepts it relies on, namely, a localizing weighted scoring rule and propriety (Definition 1), require us to verify three conditions. Specifically, $\forall \mathrm{H} \in \mathcal{H}$: (i) $S^{\flat}_{w,\mathrm{H}}(\mathrm{P}, y)$ must be localizing relative to $\mathcal{W}$, (ii) $S^{\flat}_{w,\mathrm{H}}(\mathrm{P}, y)$ must be proper relative to $\mathcal{P}$, $\forall w \in \mathcal{W}$, and (iii) the ‘if and only if’ statement in Definition 4. We prove them one by one.

(i) $S^{\flat}_{w,\mathrm{H}}(\mathrm{P}, y)$ is localizing relative to $\mathcal{W}$, $\forall \mathrm{H} \in \mathcal{H}$, by Corollary A3.

(ii) Fix an arbitrary $w \in \mathcal{W}$ and $\mathrm{H} \in \mathcal{H}$. Since $\mathcal{P}^{\flat}_{w,\mathrm{H}} \subseteq \mathcal{P}^{\flat}$, S is strictly proper relative to $\mathcal{P}^{\flat}_{w,\mathrm{H}}$, by which

$$\int_{\mathcal{Y}} S(\mathrm{P}^{\flat}_{w,\mathrm{H}}, y)\mathrm{P}^{\flat}_{w,\mathrm{H}}(\mathrm{d}y) \geq \int_{\mathcal{Y}} S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, y)\mathrm{P}^{\flat}_{w,\mathrm{H}}(\mathrm{d}y), \quad \forall \mathrm{P}^{\flat}_{w,\mathrm{H}}, \mathrm{F}^{\flat}_{w,\mathrm{H}} \in \mathcal{P}^{\flat}_{w,\mathrm{H}}. \tag{A.1}$$

By definition of the class $\mathcal{P}^{\flat}_{w,\mathrm{H}} \equiv \{[\mathrm{P}]^{\flat}_{w,\mathrm{H}}, \mathrm{P} \in \mathcal{P}\}$, where $[\cdot]^{\flat}_{w,\mathrm{H}} : \mathcal{P} \to \mathcal{P}^{\flat}_{w,\mathrm{H}}$, denotes the map given by the censored measure in Definition 6, this is equivalent to

$$\int_{\mathcal{Y}} S([\mathrm{P}]^{\flat}_{w,\mathrm{H}}, y)[\mathrm{P}]^{\flat}_{w,\mathrm{H}}(\mathrm{d}y) \geq \int_{\mathcal{Y}} S([\mathrm{F}]^{\flat}_{w,\mathrm{H}}, y)[\mathrm{P}]^{\flat}_{w,\mathrm{H}}(\mathrm{d}y), \quad \forall \mathrm{P}, \mathrm{F} \in \mathcal{P}, \tag{A.2}$$

and hence, by Lemma A1, also to

$$\int_{\mathcal{Y}} S^{\flat}_{w,\mathrm{H}}(\mathrm{P}, y)\mathrm{P}(\mathrm{d}y) \geq \int_{\mathcal{Y}} S^{\flat}_{w,\mathrm{H}}(\mathrm{F}, y)\mathrm{P}(\mathrm{d}y), \quad \forall \mathrm{P}, \mathrm{F} \in \mathcal{P}. \tag{A.3}$$

Therefore, $S^{\flat}_{w,\mathrm{H}}(\mathrm{P}, y)$ is proper relative to $\mathcal{P}$ by Definition 1.

(iii) Since S is strictly proper relative to $\mathcal{P}^{\flat}$ and hence $\mathcal{P}^{\flat}_{w,\mathrm{H}}$, $\forall w \in \mathcal{W}$, $\mathrm{H} \in \mathcal{H}$, it also follows that, $\forall w \in \mathcal{W}$, $\mathrm{H} \in \mathcal{H}$ and $\mathrm{P}^{\flat}_{w,\mathrm{H}}, \mathrm{F}^{\flat}_{w,\mathrm{H}} \in \mathcal{P}^{\flat}_{w,\mathrm{H}}$,

$$\int_{\mathcal{Y}} S(\mathrm{P}^{\flat}_{w,\mathrm{H}}, y)\mathrm{P}^{\flat}_{w,\mathrm{H}}(\mathrm{d}y) = \int_{\mathcal{Y}} S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, y)\mathrm{P}^{\flat}_{w,\mathrm{H}}(\mathrm{d}y) \iff \mathrm{P}^{\flat}_{w,\mathrm{H}}(E) = \mathrm{F}^{\flat}_{w,\mathrm{H}}(E),$$

$\forall E \in \mathcal{G}$, and thus, by Lemma A2,

$$\int_{\mathcal{Y}} S(\mathrm{P}^{\flat}_{w,\mathrm{H}}, y)\mathrm{P}^{\flat}_{w}(\mathrm{d}y) = \int_{\mathcal{Y}} S(\mathrm{F}^{\flat}_{w,\mathrm{H}}, y)\mathrm{P}^{\flat}_{w,\mathrm{H}}(\mathrm{d}y) \iff \mathrm{P}(E \cap \{w > 0\}) = \mathrm{F}(E \cap \{w > 0\}),$$

$\forall E \in \mathcal{G}$, and hence, by Lemma A1, also

$$\int_{\mathcal{Y}} S^{\flat}_{w,\mathrm{H}}(\mathrm{P}, y)\mathrm{P}(\mathrm{d}y) = \int_{\mathcal{Y}} S^{\flat}_{w,\mathrm{H}}(\mathrm{F}, y)\mathrm{P}(\mathrm{d}y) \iff \mathrm{P}(E \cap \{w > 0\}) = \mathrm{F}(E \cap \{w > 0\}),$$

$\forall E \in \mathcal{G}$, which is the desired 'if and only if' statement of Definition 4.

But then, as we have verified each of the listed conditions (i) to (iii), we have shown that $S^{\flat}_{w,\mathrm{H}}(\mathrm{P}, y)$ is strictly locally proper relative to $(\mathcal{P}, \mathcal{W}, \mathcal{H})$.

The score divergence $\mathbb{D}_{S^{\flat}_{w,\mathrm{H}}}$ is a local divergence because $S^{\flat}_{w,\mathrm{H}}$ is strictly locally proper.

Finally, we show that $\mathbb{D}_{S^\flat_{w,\mathrm{H}}}$ is a localized divergence of $\mathbb{D}_S$. For all $\mathrm{H} \in \mathcal{H}$, the map $[\cdot]^\flat_{w,\mathrm{H}} : \mathcal{P} \to \mathcal{P}^\flat_{w,\mathrm{H}}$ is such that the original measure F is recovered for $w(y) = \mathbb{1}_{\mathcal{Y}}(y)$. Furthermore, a direct consequence of Lemma A1 is that for all $w \in \mathcal{W}$, $\mathrm{H} \in \mathcal{H}$,

$$\begin{aligned}
\mathbb{D}_{S^\flat_{w,\mathrm{H}}}(\mathrm{P}\|\mathrm{F}) &= \int_{\mathcal{Y}} S^\flat_{w,\mathrm{H}}(\mathrm{F}, y)\mathrm{P}(\mathrm{d}y) - \int_{\mathcal{Y}} S^\flat_{w,\mathrm{H}}(\mathrm{F}, y)\mathrm{P}(\mathrm{d}y) \\
&= \int_{\mathcal{Y}} S(\mathrm{P}^\flat_{w,\mathrm{H}}, y)\mathrm{P}^\flat_{w,\mathrm{H}}(\mathrm{d}y) - \int_{\mathcal{Y}} S(\mathrm{F}^\flat_{w,\mathrm{H}}, y)\mathrm{P}^\flat_{w,\mathrm{H}}(\mathrm{d}y), \\
&= \mathbb{D}_S(\mathrm{P}^\flat_{w,\mathrm{H}}\|\mathrm{F}^\flat_{w,\mathrm{H}}).
\end{aligned}$$

Additionally observing that the map $[\cdot]^\flat_{w,\mathrm{H}} : \mathcal{P} \to \mathcal{P}^\flat_{w,\mathrm{H}}$ is surjective $\forall w \in \mathcal{W}, \mathrm{H} \in \mathcal{H}$, we conclude that $\forall \mathrm{P}^\flat_{w,\mathrm{H}}, \mathrm{F}^\flat_{w,\mathrm{H}} \in \mathcal{P}^\flat_{w,\mathrm{H}}, \exists \mathrm{P}, \mathrm{F} \in \mathcal{P} : \mathbb{D}_{S^\flat_{w,\mathrm{H}}}(\mathrm{P}\|\mathrm{F}) = \mathbb{D}_S(\mathrm{P}^\flat_{w,\mathrm{H}}\|\mathrm{F}^\flat_{w,\mathrm{H}})$. Therefore, $\mathbb{D}_{S^\flat_{w,\mathrm{H}}}$ is a localized divergence of $\mathbb{D}_S$, for all $w \in \mathcal{W}, \mathrm{H} \in \mathcal{H}$. □

A.2 Proof of Theorem 3

We start by rephrasing the hypotheses. Since the densities f_{jt}, where $j \in \{0, 1\}$ must integrate to one on $A_t \cup A_t^c$, the hypotheses imply that these densities integrate to $\bar{F}_{jt} := \mathrm{F}_{jt}(A_t^c)$ on A_t^c. Therefore, the implied specification on A_t^c can be summarized as

$$\frac{\bar{F}_{jt}}{\bar{H}_{jt}} h_{jt}\mathbb{1}_{A_t^c} = \bar{F}_{jt}[h_{jt}]^\sharp_{A_t^c}, \quad j \in \{0, 1\},$$

where the unknown densities $h_{jt} = \frac{\mathrm{dH}_{jt}}{\mathrm{d}\mu}$ can be seen as infinite-dimensional nuisance parameters, and $\bar{H}_{jt} := \mathrm{H}_{jt}(A_t^c)$. Explicating the implied assumption on A_t^c in the hypotheses and phrasing them in terms of a statement about the whole sample distribution leads to the equivalent hypotheses

$$\mathbb{H}_j : p(\mathbf{y}) = f_j(\mathbf{y}) := \prod_{t=0}^{T-1} \left(f_{jt}(y_{t+1}) \mathbb{1}_{A_t}(y_{t+1}) + \bar{F}_{jt}[h_{jt}]^{\sharp}_{A_t^c}(y_{t+1}) \right), \quad j \in \{0, 1\}.$$

Since the densities f_{jt} are fixed, and the densities h_{jt} are unrestricted under both hypotheses, the class of densities satisfying hypothesis $\mathbb{H}_j$ can alternatively be written as

$$\mathbb{F}_j = \left\{ \prod_{t=0}^{T-1} \left(f_{jt}(y_{t+1}) \mathbb{1}_{A_t}(y_{t+1}) + \bar{F}_{jt}[h_{jt}]^{\sharp}_{A_t^c}(y_{t+1}) \right), h_j \in \mathcal{H} \right\}, \quad j \in \{0, 1\},$$

in which $\mathcal{H}$ denotes the space of all densities on $\mathcal{Y}^T$. In terms of the index set of all observations $\mathcal{I} = \{1, \ldots, T\}$, this space can also be denoted as $\mathcal{Y}(\mathcal{I}) = \prod_{t \in \mathcal{I}} \mathcal{Y}_t$.

Fixing an $\alpha \in (0, 1)$, the aim is to find a uniformly most powerful (UMP) test ϕ^* of size α for testing problem (6), i.e., a solution to the maximization problem

$$\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{f_1} \phi, \qquad \Phi(\alpha) = \{\phi : \sup_{f_0 \in \mathbb{F}_0} \mathbb{E}_{f_0} \phi \leq \alpha\}. \tag{A.4}$$

Now fix an $h_1 \in \mathcal{H}$ so that the distribution under the alternative is completely known. Given the fact that the hypotheses are, in the end, silent about the shape of the densities on A_t^c, we conjecture that a UMP test neglects the information about the shape of the densities on A_t^c, $\forall t$. If $T = 2$, for example, and we consider the optimal test on $A_0 \times A_1^c$, our intuition is that an optimal test is not concerned about the shape of $[h_{11}]^{\sharp}_{A_1^c}$, that is, the specific values $[h_{11}]^{\sharp}_{A_1^c}(y_2)$ for all $y_2 \in A_1^c$, but only about the total probability of an outcome falling into A_1^c. In other words, we expect that a solution to problem (A.4) has integrated out the dependence on the nuisance densities.

Although it is obvious that marginalizing out the (still assumed to be fixed) density

$h_1 \in \mathcal{H}$ is harmless in terms of power, it is non-trivial that this is an affordable strategy in terms of size for all $h_0 \in \mathcal{H}$. Lemma A4 and its proof show that the subclass of tests disregarding information about the shape of h_1 is guaranteed to be size correct. In our search for the UMP test, Corollary A5 then formalizes the idea that we can restrict our attention to tests of the conjectured kind.

Lemma A4. *Consider testing problem* (6) *and suppose that the outcomes* $(y_t)_{t\in\mathcal{I}_A}$ *are in* A_t*, and the remaining* $T-k$*, with* $k = |\mathcal{I}_A|$*, observations* $(y_t)_{t\in\mathcal{I}_{A^c}}$ *in* A_t^c*. For an arbitrary but fixed density* $h_1 \in \mathcal{H}$*, the test*

$$\psi_{h_1} : \mathcal{Y}^T \to [0,1], \quad \psi_{h_1} = \int_{\mathcal{Y}(\mathcal{I}_{A^c})} \phi^*_{h_1} \prod_{t\in\mathcal{I}_{A^c}} [h_{1t}]^\sharp_{A_t^c} \mathrm{d}\mu^{\otimes|\mathcal{I}_{A^c}|},$$

where $\phi^*_{h_1}$ *denotes a solution to problem* (A.4)*, is such that* $\psi_{h_1} \in \Phi(\alpha)$*.*

Proof. Due to the integral over $\mathcal{Y}(\mathcal{I}_{A^c})$, any test ψ_{h_1} is constant in arguments varying in $\mathcal{Y}(\mathcal{I}_{A^c})$. We can use this observation to simplify the size of a test ψ_{h_1}. In particular, $\forall h_1 \in \mathcal{H}$, we have that

$$\begin{aligned}
\sup_{f_0\in\mathbb{F}_0} \mathbb{E}_{f_0}\psi_{h_1} &= \left(\prod_{t\in\mathcal{I}_{A^c}} \bar{F}_{0t}\right) \sup_{h_0\in\mathcal{H}} \int_{\mathcal{Y}^T} \psi_{h_1} \prod_{t\in\mathcal{I}_A} f_{0t}\mathbb{1}_{A_t} \prod_{t\in\mathcal{I}_{A^c}} [h_{0t}]^\sharp_{A_t^c} \mathrm{d}\mu^{\otimes T} \\
&= \left(\prod_{t\in\mathcal{I}_{A^c}} \bar{F}_{0t}\right) \int_{\mathcal{Y}(\mathcal{I}_A)} \psi_{h_1} \prod_{t\in\mathcal{I}_A} f_{0t}\mathbb{1}_{A_t} \mathrm{d}\mu^{\otimes|\mathcal{I}_A|} \\
&= \left(\prod_{t\in\mathcal{I}_{A^c}} \bar{F}_{0t}\right) \int_{\mathcal{Y}^T} \phi^*_{h_1} \prod_{t\in\mathcal{I}_{A^c}} [h_{1t}]^\sharp_{A_t^c} \mathrm{d}\mu^{\otimes|\mathcal{I}_{A^c}|} \prod_{t\in\mathcal{I}_A} f_{0t}\mathbb{1}_{A_t} \mathrm{d}\mu^{\otimes|\mathcal{I}_A|} \\
&\leq \left(\prod_{t\in\mathcal{I}_{A^c}} \bar{F}_{0t}\right) \sup_{h_0\in\mathcal{H}} \int_{\mathcal{Y}^T} \phi^*_{h_1} \prod_{t\in\mathcal{I}_{A^c}} [h_{0t}]^\sharp_{A_t^c} \mathrm{d}\mu^{\otimes|\mathcal{I}_{A^c}|} \prod_{t\in\mathcal{I}_A} f_{0t}\mathbb{1}_{A_t} \mathrm{d}\mu^{\otimes|\mathcal{I}_A|} \\
&= \sup_{f_0\in\mathbb{F}_0} \mathbb{E}_{f_0}\phi^*_{h_1} \\
&\leq \alpha,
\end{aligned}$$

since $\phi^*_{h_1} \in \Phi(\alpha)$. Hence, $\psi_{h_1} \in \Phi(\alpha)$. $\square$

Corollary A5. *Consider testing problem* (6) *and assume that outcomes* $y_t \in A_t, \forall t \in \mathcal{I}_A$, *and* $y_t \in A_t^c, \forall t \in \mathcal{I}_{A^c}$. *Let* $\Psi(\alpha) \subseteq \Phi(\alpha)$ *denote the class of size* α *tests on* $\mathcal{Y}^T$ *that are constant in arguments varying in* $\mathcal{Y}(\mathcal{I}_{A^c})$. *Then,*

$$\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{f_1}\phi = \max_{\psi \in \Psi(\alpha)} \mathbb{E}_{f_1}\psi, \qquad \forall h_1 \in \mathcal{H}.$$

Proof. Fix an arbitrary $h_1 \in \mathcal{H}$. Since $\Psi(\alpha) \subseteq \Phi(\alpha)$, we trivially have that $\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{f_1}\phi \geq \max_{\psi \in \Psi(\alpha)} \mathbb{E}_{f_1}\psi$. Now suppose that $\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{f_1}\phi < \max_{\psi \in \Psi(\alpha)} \mathbb{E}_{f_1}\psi$. Then, we can always define the test $\tilde{\psi} = \int_{\mathcal{Y}(\mathcal{I}_{A^c})} \phi^* \prod_{t \in \mathcal{I}_{A^c}} [h_{1t}]^\sharp_{A_t^c} \mathrm{d}\mu^{\otimes |\mathcal{I}_{A^c}|}$, with $\phi^* \in \arg\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{f_1}\phi$, satisfying $\mathbb{E}_{f_1}\phi^* = \mathbb{E}_{f_1}\tilde{\psi}$. But, by Lemma A4, $\tilde{\psi} \in \Psi(\alpha)$, in which case $\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{f_1}\phi = \max_{\psi \in \Psi(\alpha)} \mathbb{E}_{f_1}\tilde{\psi}$, contradicting the assumed strict inequality. $\square$

For any fixed $h_1 \in \mathcal{H}$, the reduced optimization problem resulting from Corollary A5 simplifies to a simple versus simple hypothesis in terms of the censored measures $\mathrm{dF}^\flat_{jt,A_t} = \mathbb{1}_{A_t}\mathrm{dF}_{jt} + \bar{F}_{jt}\mathrm{d}\delta_*$, allowing us to apply Neyman and Pearson (1933).

Specifically, for any fixed $h_1 \in \mathcal{H}$, the most powerful test of size α is a solution to the restricted maximization problem

$$\begin{aligned}
&\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{f_1}\phi \\
&= \max_{\boldsymbol{\alpha} \in \Delta_{\bar{T}}(\alpha)} \sum_{k=0}^{T} \sum_{s=1}^{\binom{T}{k}} \max_{\phi_{k,s} \in \Phi(\alpha_{k,s})} \mathbb{E}_{f_1}\left(\phi_{k,s} | y_{t+1} \in A_t, \forall i \in \mathcal{I}_A(k,s) \wedge y_{t+1} \in A_t^c, \forall i \in \mathcal{I}_{A^c}(k,s)\right) \\
&= \max_{\boldsymbol{\alpha} \in \Delta_{\bar{T}}(\alpha)} \sum_{k=0}^{T} \sum_{s=1}^{\binom{T}{k}} \max_{\phi_{k,s} \in \Psi(\alpha_{k,s})} \mathbb{E}_{f_1}\left(\phi_{k,s} | y_t \in A_t, \forall i \in \mathcal{I}_A(k,s) \wedge y_t \in A_t^c, \forall i \in \mathcal{I}_{A^c}(k,s)\right),
\end{aligned}$$

which further simplifies to

$$\begin{aligned}\max_{\phi\in\Phi(\alpha)} \mathbb{E}_{f_1}\phi &= \max_{\boldsymbol{\alpha}\in\Delta_{\bar{T}}(\alpha)} \sum_{k=0}^{T}\sum_{s=1}^{\binom{T}{k}} \max_{\phi_{k,s}\in\Psi(\alpha_{k,s})} \left(\prod_{t\in\mathcal{I}_{A^c}} \bar{F}_{1t}\right) \int_{\mathcal{Y}(\mathcal{I}_A)} \phi_{k,s} \prod_{t\in\mathcal{I}_A} f_{1t}\mathbb{1}_{A_t}\mathrm{d}\mu^{\otimes T} \\ &= \max_{\boldsymbol{\alpha}\in\Delta_{\bar{T}}(\alpha)} \sum_{k=0}^{T}\sum_{s=1}^{\binom{T}{k}} \max_{\phi_{k,s}\in\Psi(\alpha_{k,s})} \int_{\mathcal{Y}^T} \phi_{k,s} \prod_{t=0}^{T-1} \mathrm{dF}^{\flat}_{1t,A_t} \\ &= \max_{\boldsymbol{\alpha}\in\Delta_{\bar{T}}(\alpha)} \sum_{k=0}^{T}\sum_{s=1}^{\binom{T}{k}} \max_{\phi_{k,s}\in\Phi(\alpha_{k,s})} \int_{\mathcal{Y}^T} \phi_{k,s} \prod_{t=0}^{T-1} \mathrm{dF}^{\flat}_{1t,A_t},\end{aligned}$$

where $\bar{T} := \sum_{k=0}^{T}\binom{T}{k}$, $\phi_{k,s}$ the test function and $\mathcal{I}_A(k,s)$ the index set for a combination (k,s), and $\Delta_{\bar{T}}(\alpha) := \{\boldsymbol{\alpha} \in [0,\alpha]^{\bar{T}} : \boldsymbol{\iota}_{\bar{T}}'\boldsymbol{\alpha} = \alpha\}$, with $\boldsymbol{\iota}_{\bar{T}}$ denoting column vector of ones of length $\bar{T}$. The first equality exploits the fact that the test function can be decomposed into test functions operating on a single part of the partitioning of the outcome space $\mathcal{Y}^T$, in which case the maximization problem can be split into finding an optimal test on each of the partitioned parts conditional on the amount of size spent on each part and the optimal distribution of size over the partition of the outcome space. The second equality holds by Corollary A5, the third equality uses that the optimal test is constant in arguments varying in $A^c := \prod_{t=0}^{T-1} A_t^c$, isolating the conditional densities, which integrate to one on A_t^c. The fourth equality holds by definition of the censored measure and the fifth equality uses that all tests that are non-constant in arguments varying in A^c map under the censored measure onto tests that are constant in arguments varying in A^c.

Finally, the result follows by observing that the final maximization problem is equivalent to finding the optimal test $\phi_A^\flat$ for the testing problem $\mathbb{H}_j : p = \prod_{t=0}^{T-1} f^\flat_{jt,A_t}$, $j \in \{0,1\}$, for which $\phi_A^\flat$ is the UMP test by the Fundamental Lemma of Neyman and Pearson (1933). By the equivalence, $\phi_A^\flat$ is, for any $h_1 \in \mathcal{H}$, also the most powerful test for testing problem (6). But, since the test $\phi_A^\flat$ is independent of h_1, it is UMP for testing problem (6). □

B Additional Proofs and Derivations

B.1 The censored density

We recall that we are silently working on the measurable space $(\mathcal{Y}^*, \mathcal{G}^*)$ relative to which we define $\mu^*(E) = \mu(E \cap \mathcal{Y})$, $\forall E \in \mathcal{G}^*$. Again, we suppress the superscript $*$.

Since $(\mu + \delta_*)(E) = 0$ implies that both $\mu(E) = 0$ and $\delta_*(E) = 0$, $\forall E \in \mathcal{G}$, we have that both $\mu \ll \mu + \delta_*$ and $\delta_* \ll \mu + \delta_*$. As a consequence,

$$f_w^\flat := \frac{\mathrm{dF}_w^\flat}{\mathrm{d}(\mu + \delta_*)} = w \frac{\mathrm{dF}}{\mathrm{d}(\mu + \delta_*)} + \bar{F}_w \frac{\mathrm{d}\delta_*}{\mathrm{d}(\mu + \delta_*)}$$

is the censored $(\mu + \delta_*)$-density of $\mathrm{F}_w^\flat$.

We can simplify this density as follows. Understanding that

$$\frac{\mathrm{dF}}{\mathrm{d}(\mu + \delta_*)} = \frac{\mathrm{dF}}{\mathrm{d}\mu} \frac{\mathrm{d}\mu}{\mathrm{d}(\mu + \delta_*)},$$

we recall from the Radon-Nikodym theorem that $\frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\delta_*)}$ is the solution to

$$\int_{\mathcal{Y}} \mathbb{1}_E \mathrm{d}\mu = \int_{\mathcal{Y}} \mathbb{1}_E \frac{\mathrm{d}\mu}{\mathrm{d}(\mu + \delta_*)} \mathrm{d}(\mu + \delta_*) = \int_{\mathcal{Y}} \mathbb{1}_E \frac{\mathrm{d}\mu}{\mathrm{d}(\mu + \delta_*)} \mathrm{d}\mu + \int_{\mathcal{Y}} \mathbb{1}_E \frac{\mathrm{d}\mu}{\mathrm{d}(\mu + \delta_*)} \mathrm{d}\delta_*.$$

By the same theorem, the solution to this equation is guaranteed to exist uniquely.

A glance at this equation reveals that a reasonable candidate is 1 μ-a.e. and 0 δ_*-a.s. We conclude that $\frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\delta_*)} = \mathbb{1}_{\mathcal{Y}\setminus\{*\}}$ is the unique solution for the Radon-Nikodym derivative. By the same token, we conclude from

$$\int_{\mathcal{Y}} \mathbb{1}_E \mathrm{d}\delta_* = \int_{\mathcal{Y}} \mathbb{1}_E \frac{\mathrm{d}\delta_*}{\mathrm{d}(\mu + \delta_*)} \mathrm{d}(\mu + \delta_*) = \int_{\mathcal{Y}} \mathbb{1}_E \frac{\mathrm{d}\delta_*}{\mathrm{d}(\mu + \delta_*)} \mathrm{d}\mu + \int_{\mathcal{Y}} \mathbb{1}_E \frac{\mathrm{d}\delta_*}{\mathrm{d}(\mu + \delta_*)} \mathrm{d}\delta_*,$$

that a reasonable candidate for $\frac{\mathrm{d}\delta_*}{\mathrm{d}(\mu+\delta_*)}$ is 0 μ-a.e. and 1 δ_*-a.s. More specifically, we deduce that $\frac{\mathrm{d}\delta_*}{\mathrm{d}(\mu+\delta_*)} = \mathbb{1}_{\{*\}}$ is the unique solution for the Radon-Nikodym derivative.

Put together, we arrive at

$$f_w^\flat(y) = w(y)\frac{\mathrm{dF}}{\mathrm{d}\mu}(y)\mathbb{1}_{\mathcal{Y}\setminus\{*\}}(y) + \bar{F}_w\mathbb{1}_*(y) = w(y)f(y)\mathbb{1}_{A_w}(y) + \bar{F}_w\mathbb{1}_{\{*\}}(y), \quad y \in \mathcal{Y},$$

where f denotes the μ-density of F. Moreover, given that $f_w^\flat(y) = 0, \forall y \in \mathcal{Y}\setminus\mathcal{Y}_{A_w}^\flat$, recalling $\mathcal{Y} \equiv \mathcal{Y}^*$, we may alternatively write

$$f_w^\flat(y) = w(y)f(y) + \bar{F}_w\mathbb{1}_{\{*\}}(y), \quad y \in \mathcal{Y}_{A_w}^\flat,$$

where $w(*) = 0$ by construction.

B.2 Proof of Corollary 1

The test based on $\tilde{\lambda}(\mathbf{y})$ is equivalent to the UMP test in Theorem 3, since

$$\begin{aligned}\tilde{\lambda}(\mathbf{y}) &= \sum_{t=0}^{T-1}\left(\mathrm{LogS}_{A_t}^\flat(f_{1t}, y_{t+1}) - \mathrm{LogS}_{A_t}^\flat(f_{0t}, y_{t+1})\right)\\ &= \sum_{t=0}^{T-1}\Bigg(\log\left(f_{1t,A_t}^\flat(y_{t+1})\right) - \log\left(f_{0t,A_t}^\flat(y_{t+1})\right)\Bigg)\\ &= \log\lambda(\mathbf{y}),\end{aligned}$$

and hence $\lambda(\mathbf{y}) \gtreqless c \iff \tilde{\lambda}(\mathbf{y}) \gtreqless \tilde{c}$, with $\tilde{c} = \log c$.

B.3 Proof of Corollary 2

We show that $\phi^{\sharp}_{A}$ is not UMP by constructing a counterexample in which the power of $\phi^{\sharp}_{A}$ is strictly smaller than that of $\phi^{\flat}_{A}$. In particular, suppose that $T = 1$ and accordingly drop the time subscripts. Moreover, consider two Lebesgue densities f_0 and f_1 on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that differ on $A \subset \mathbb{R}$. For $T = 1$, the likelihood ratios of the conditional and censored test simplify to

$$\begin{aligned}
\lambda^{\sharp}_{A}(y) &:= \mathrm{e}^{\tilde{\lambda}^{\sharp}_{A}(y)} = \frac{\mathrm{F}_0(A)}{\mathrm{F}_1(A)} \frac{f_1(y)}{f_0(y)} \mathbb{1}_{A}(y) + \frac{1}{1} \mathbb{1}_{A^c}(y) \\
\lambda^{\flat}_{A}(y) &= \mathrm{e}^{\tilde{\lambda}^{\flat}_{A}(y)} = \frac{f_1(y)}{f_0(y)} \mathbb{1}_{A}(y) + \frac{\mathrm{F}_1(A^c)}{\mathrm{F}_0(A^c)} \mathbb{1}_{A^c}(y).
\end{aligned}$$

There exist many examples for which the power of the censored test is strictly larger than the power of the conditional test. For instance, suppose that f_0 and f_1 have identical support in $\mathbb{R}$, and on A are strictly positive, proportional and unequal. Then, by proportionality, $\frac{\mathrm{F}_0(A)}{\mathrm{F}_1(A)} \frac{f_1(y)}{f_0(y)} = 1$ for $y \in A$ and hence $\lambda^{\sharp}_{A}(y) = 1$, so that a test of level $\alpha \in (0, 1)$ based on $\lambda^{\sharp}_{A}$ will reject with probability α regardless of the outome y, and hence only have trivial power α under f_1. In that same situation, a level α test based on the censored likelihood ratio $\lambda^{\flat}_{A}$ will have power strictly larger than α, because $\lambda^{\flat}_{A}(y)$ can take two different values, depending on whether the outcome y falls in A or A^c. By rejecting $\mathbb{H}_0$ with higher probability for those realizations y for which the censored likelihood ratio $\lambda^{\flat}_{A}(y)$ takes the larger of these values, it will achieve power strictly larger than α. Consequently, the conditional test $\phi^{\sharp}_{A}$ is not UMP.

As a concrete example, consider the above proportional situation with $\mathrm{F}_0(A) = \eta_0 \in (0, 1)$ and $\mathrm{F}_1(A) = \eta_1 \in (0, 1)$, where $\eta_1 > \eta_0$ and hence $\frac{\eta_1}{\eta_0} > 1 > \frac{1-\eta_1}{1-\eta_0}$, so the likelihood ratio in favor of $\mathbb{H}_1$ is larger than 1 for $y \in A$ and smaller than 1 for $y \in A^c$. Under these conditions it can be verified that the level α test based on $\lambda_A^\flat$ for $\alpha \in (0, \eta_0)$ rejects $\mathbb{H}_0$ in favor of $\mathbb{H}_1$ with probability $\gamma_A = \frac{\alpha}{\eta_0}$ for $y \in A$ and with probability $\gamma_{A^c} = 0$ for $y \in A^c$. The power of this test is $\frac{\alpha\eta_1}{\eta_0} > \alpha$. For $\alpha \in [\eta_0, 1)$, the level α test based on $\lambda_A^\flat$ rejects with probability $\gamma_A = 1$ for $y \in A$ and with probability $\gamma_{A^c} = \frac{\alpha-\eta_0}{1-\eta_0}$ for $y \in A^c$. The power of the test is $\eta_1 + \frac{\alpha-\eta_0}{1-\eta_0}(1-\eta_1)$, which is also strictly larger than α, as can be seen by inspecting the complementary probability: $1-\left(\eta_1 + \frac{\alpha-\eta_0}{1-\eta_0}(1-\eta_1)\right) = (1-\eta_1)\left(1 - \frac{\alpha-\eta_0}{1-\eta_0}\right) = \frac{1-\eta_1}{1-\eta_0}(1-\alpha) < 1-\alpha$. □

C Z, Q-Randomization

The (generalized) censored scoring rules in Definition 5 and Definition 6 can alternatively be formulated in terms of a randomization procedure. This is appealing, since it generalizes the identity $S_A^\flat(\mathrm{F}, y) = S(\mathrm{F}_A^\flat, y_A^\flat)$ obtained for indicator functions in Section 3.1 to general weight functions. The randomization procedure relies on an auxiliary random variable Z, indicating, conditional on the realization y, whether the observation is censored or not. Specifically, we assume $Z|(Y=y) \sim \mathrm{B}_{w(y)} \equiv \mathrm{Bernoulli}(w(y))$.

Let $\mathcal{Z} := \{0,1\}$, and denote by $\tilde{\mathcal{G}}$ the product σ-algebra of $\mathcal{G}$ and $\sigma(\mathcal{Z})$. The censored random variable in Equation (1) generalizes to the $\tilde{\mathcal{G}}/\mathcal{G}_w^\flat$-measurable function $Y_w^\flat : \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y}_w^\flat$, defined by

$$Y_w^\flat \equiv Y_w^\flat(y, z) := \begin{cases} y, & z = 1, \\ *, & z = 0. \end{cases} \tag{C.1}$$

Correspondingly, the censored distribution $\mathrm{F}_w^\flat$ in Equation (3) is the pushforward measure of the joint distribution of Y and Z by $Y_w^\flat$. For indicator weight functions $w(y) = \mathbb{1}_A(y)$, $Z = 1$ (with probability one) for $y \in A$, and 0 otherwise. Hence, for $w(y) = \mathbb{1}_A(y)$, the map reduces to the censored random variable in Equation (1).

Using the auxiliary random variable Z, the censored scoring rule in Definition 5 can be written as

$$S_w^\flat(\mathrm{F}, y) = \mathbb{E}_{\mathrm{B}_{w(y)}} S(\mathrm{F}_w^\flat, Y_w^\flat(y, Z)). \tag{C.2}$$

A similar representation of the generalized censored scoring rule in Definition 6 can be obtained by additionally introducing a random variable Q, independent of (Y, Z), with distribution $\mathrm{H} \in \mathcal{H} \subseteq \mathcal{P}$. Since all distributions in $\mathcal{P}$ are defined relative to $(\mathcal{Y}, \mathcal{G})$, the nuisance distribution is defined relative to the restricted space $(\mathcal{Y}_\mathrm{H}, \mathcal{G}_\mathrm{H})$, where $\mathcal{Y}_\mathrm{H} \subseteq \mathcal{Y}$ and $\mathcal{G}_\mathrm{H} \subseteq \mathcal{G}$, for all $\mathrm{H} \in \mathcal{H}$. Let $\tilde{\mathcal{G}}_\mathrm{H}$ be the product σ-algebra of $\mathcal{G}$, $\sigma(\mathcal{Z})$ and $\mathcal{G}_\mathrm{H}$.

We define $\mathcal{Y}_{w,\mathrm{H}}^\flat := A_w \cup \mathcal{Y}_\mathrm{H}$ and $\mathcal{G}_{w,\mathrm{H}}^\flat := \sigma(\{E \cap A_w : E \in \mathcal{G}\} \cup \mathcal{G}_\mathrm{H})$, and note that $\mathcal{Y}_{w,\mathrm{H}}^\flat \subseteq \mathcal{Y}$ and $\mathcal{G}_{w,\mathrm{H}}^\flat \subseteq \mathcal{G}, \forall w \in \mathcal{W}, \mathrm{H} \in \mathcal{H}$. Consider the $\tilde{\mathcal{G}}_\mathrm{H}/\mathcal{G}_{w,\mathrm{H}}^\flat$-measurable function $Y_w^\flat : \mathcal{Y} \times \mathcal{Z} \times \mathcal{Y}_\mathrm{H} \to \mathcal{Y}_{w,\mathrm{H}}^\flat$, given by

$$Y_{w,\mathrm{H}}^\flat := Y_{w,\mathrm{H}}^\flat(y, z, q) = \begin{cases} y, & \text{if } z = 1, \\ q, & \text{if } z = 0. \end{cases} \tag{C.3}$$

The generalized censored distribution $\mathrm{F}_{w,\mathrm{H}}^\flat$ in Definition 6 is the pushforward measure of the joint distribution of Y, Z and Q by $Y_{w,\mathrm{H}}^\flat$. Moreover, the generalized censored distribution in Definition 6 can be written as

$$S_{w,\mathrm{H}}^\flat(\mathrm{F}, y) = \mathbb{E}_{\mathrm{B}_{w(y)},\mathrm{H}} S\big(\mathrm{F}_{w,\mathrm{H}}^\flat, Y_{w,\mathrm{H}}^\flat(y, Z, Q)\big). \tag{C.4}$$

For strict local propriety of $S_{w,\mathrm{H}}^\flat(\mathrm{F}, y)$, the weight function $w \in \mathcal{W}$ and nuisance distribution $\mathrm{H} \in \mathcal{H}$ must be such that Assumption 1 is satisfied for all $\mathrm{F} \in \mathcal{P}$. The Z, Q-

randomization formulation of the generalized censored scoring rule in Equation (C.3) helps in providing more intuition for Assumption 1. In particular, if the support of the nuisance distribution $\mathcal{Y}_{\mathrm{H}}$ and A_w, the support of the measure F_w, overlap, then the identifiability of the censoring event is generally lost. Indeed, in that case outcomes in $A_w \cap \mathcal{Y}_{\mathrm{H}}$ can correspond to either $z = 1$ or $z = 0$, in contrast to the Z-randomization procedure in Equation (C.1). However, under Assumption 1, the probability $\bar{F}_w$ can still be identified under H; see Example C.1.

Example C.1 (Intuition Assumption 1)**.** *Reconsider the setting of Example 1. Rather than extending the outcome space and uniquely identifying the censoring event by* $*$*, we now consider the generalized censored distribution* $\mathrm{dF}^\flat_{A,\mathrm{H}} = \mathrm{dF}_A + \bar{F}_A \mathrm{dH}$*, where* H *is the* $\mathrm{B}_{1/2}$ *distribution over* $\mathcal{Y}_{\mathrm{H}} = \{s, b\}$*. The outcome* b *overlaps with the outcome of interest* $A = \{s\}$*, in which case Assumption 1 allows for randomization over* $\{s, b\}$ *since* $\mathrm{F}_A(b) = 0$ *while* $\mathrm{H}(b) = 1/2 > 0$*. Moreover,* $\mathcal{Y}^\flat_{\mathrm{H},A} = \{s, b\}$ *and* $\mathcal{G}^\flat_{A,\mathrm{H}} = \mathcal{G}$*. The event* s *induces* $(Y^\flat_{A,\mathrm{H}})^{-1}(s) = \{(s, 1, q) : q \in \{s, b\}\} \cup \{(y, 0, b) : y \in \{o, b\}\}$ *with probability* $\mathrm{F}^\flat_{w,\mathrm{H}}(s) = \mathrm{F}((Y^\flat_{A,\mathrm{H}})^{-1}(s)) = \mathrm{F}(s) + \bar{F}_A/2$*. The probability* $\bar{F}_A$ *is not identifiable from this expression alone.*

Key to identification is the event b*, for which* $(Y^\flat_{A,\mathrm{H}})^{-1}(b) = \{(y, 0, b) : y \in \{o, b\}\}$ *implies* $\mathrm{F}^\flat_{w,\mathrm{H}}(b) = \mathrm{F}((Y^\flat_{A,\mathrm{H}})^{-1}(\{o, b\})) = \bar{F}_A/2$*. Hence, if two censored distributions coincide for all events, then so do the probabilities of the censoring event, as desired for strict local propriety (see Lemma A2). The same requirement is reflected in the scoring rule representation in Equation* (C.4)*, simplifying to* $S^\flat_{w,\mathrm{H}}(\mathrm{F}, y) = \frac{1}{2}\mathbb{E}_{\mathrm{B}_{w(y)}} S\big(\mathrm{F}^\flat_{w,\mathrm{H}}, Y^\flat_{w,\mathrm{H}}(y, Z, s)\big) + \frac{1}{2}\mathbb{E}_{\mathrm{B}_{w(y)}} S\big(\mathrm{F}^\flat_{w,\mathrm{H}}, Y^\flat_{w,\mathrm{H}}(y, Z, b)\big)$*. Indeed, the expectation concerned with the realization* b *of* Q *does not suffer from identifiability issues, providing strictness in local propriety, and this is sufficient when considering expected score differences based on proper scoring rules.*

The censored pmf $f^{\flat}_{A}(y) = f(y)\mathbb{1}_{A}(y) + \bar{F}_{A}\mathbb{1}_{*}(y)$ defined on $\mathcal{Y}^{\flat}_{A}$ and the generalized censored pmf $f^{\flat}_{A,\mathrm{H}}(y) = (f(y) + \frac{1}{2}\bar{F}_{A})\mathbb{1}_{A}(y) + \frac{1}{2}\bar{F}_{A}\mathbb{1}_{\{b\}}(y)$ defined on $\mathcal{Y}^{\flat}_{A,\mathrm{H}}$ are both valid pmfs for generating strictly proper scoring rules. However, there is a critical distinction that motivates the use of the non-generalized censored pmf when both constructions are feasible. Indeed, the censored pmf adheres to minimal localization, as a result of which $f^{\flat}_{A}(y) = f(y), \forall y \in A$, so the censored pmf coincides with the original pmf on the region of interest. By contrast, the generalized censored pmf does not coincide with the original pmf on A, as it alters the distribution by incorporating mass from the nuisance component.

D Censoring of Distance-Sensitive Scoring Rules

Distance-sensitive scoring rules depend on a distance measure $d : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_{+}$ and are generally less easily localized than (semi-)local scoring rules. For instance, minimal localization is infeasible because $d(y, *)$ is undefined if A^{c} is a subset of the outcome space rather than a specific outcome in $\mathcal{Y}$ itself, which generally will be the case. Hence, by choosing an arbitrary point $y_0 \in \mathcal{Y}$ as censoring value, i.e. $\mathrm{F}^{\flat}_{A,y_0} = \mathrm{F}_{A} + \bar{F}_{A}\delta_{y_0}$, the generalized censored distribution $\mathrm{F}^{\flat}_{A,y_0}$ and the restricted measure $\mathrm{F}^{\flat}_{A} \equiv \mathrm{F}_{|\mathcal{G}^{\flat}_{A}}$ will generally not coincide. Example D.1, however, reveals that if $(\mathcal{Y}, \mathcal{G}) = (\mathbb{R}, \mathcal{B})$, where $\mathcal{B} \equiv \mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra on $\mathbb{R}$, and $A = (-\infty, r)$, then the specific choice $y_0 = r$ induces a censored measure $\mathrm{F}^{\flat}_{A,r}$ coinciding with $\mathrm{F}_{|\mathcal{G}^{\flat}_{A}}$. Therefore, we consider r as a natural choice for y_0 if $w(y) = \mathbb{1}_{(-\infty,r)}(y)$.

Example D.1. *A risk manager models the asset portfolio return Y using a Gaussian distribution $\mathrm{F} \equiv \mathrm{F}_{\mu,\sigma^2}$ on $(\mathbb{R}, \mathcal{B})$, with mean μ and variance σ^2, and $\mathcal{B} \equiv \mathcal{B}(\mathbb{R})$ the Borel σ-algebra on $\mathbb{R}$. Due to regulatory requirements, the risk manager is only concerned with return values below a threshold $r \in \mathbb{R}$, i.e., the region of interest is $A = (-\infty, r)$. As*

$\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, y] : y \in \mathbb{R}\})$, the minimal σ-algebra on $\mathbb{R}$ can be written as $\mathcal{B}^\flat_A = \sigma(\{(-\infty, y] \cap (-\infty, r) : y \in \mathbb{R}\})$, and includes the event $A^c = [r, \infty)$. The minimal localization $\mathrm{F}^\flat_A$ restricts F to $\mathcal{B}^\flat_A$, specifically, $\mathrm{F}^\flat_A((-\infty, y]) = \mathrm{F}((-\infty, y]) = F(y), \forall y < r$, and $\mathrm{F}^\flat_A([r, \infty)) = \mathrm{F}([r, \infty)) = 1 - F(r)$. However, the measure $\mathrm{F}^\flat_A$ does not admit a distribution function $F^\flat_A$ because, e.g. $F^\flat_A(r + 1) \equiv \mathrm{F}^\flat_A((-\infty, r + 1])$ is undefined since $(-\infty, r + 1] \notin \mathcal{B}^\flat_A$. A natural extension of $\mathrm{F}^\flat_A$ on $(\mathbb{R}, \mathcal{B}^\flat_A)$ to a distribution $\mathrm{F}^\flat_A$ on $(\mathbb{R}, \mathcal{B})$ is such that $\mathrm{F}^\flat_A([r + v, \infty)) = 0, \forall v > 0$, which effectively sets the discontinuity of the distribution function $\mathrm{F}^\flat_A(y) \equiv \mathrm{F}^\flat_A((-\infty, y])$ at $y = r$. This choice is considered natural as it ensures that the extended $\mathrm{F}^\flat_A$ and F coincide on $\mathcal{B}^\flat_A$, i.e. $\mathrm{F}^\flat_A(E) = \mathrm{F}(E), \forall E \in \mathcal{B}^\flat_A$, and $\mathrm{F}^\flat_A(E) = 0, \forall E \in \mathcal{B} \backslash \mathcal{B}^\flat_A$.

It is not always possible to choose $y_0 \in \mathcal{Y}$ such that $\mathrm{F}^\flat_{A,y_0}$ coincides with the restriction of F to $\mathcal{G}^\flat_A$. A clear example is the case where we focus on non-tail events by using the weight function $w(y) = \mathbb{1}_{(r_1,r_2)}(y)$ on $\mathbb{R}$, where $r_1, r_2 \in \mathbb{R}$ and $r_1 < r_2$. As illustrated by Example D.2, in this case there does not exist a measure equivalent to the minimal localization $\mathrm{F}^\flat_A$ that admits a distribution function. The root of this problem is that for the distribution function to exist, we need both the events $(A^c)_1 := (-\infty, r_1]$ and $(A^c)_2 := [r_2, \infty)$, while only their union $A^c = (A^c)_1 \cup (A^c)_2$ is a member of the minimal σ-algebra $\mathcal{B}^\flat_A$. As a solution, we suggest to distribute $\mathrm{F}(A^c)$ according to the F-independent weights γ and $1-\gamma$ over the events $(A^c)_1$ and $(A^c)_2$, respectively, where $\gamma \in [0, 1]$. The latter ensures that the scoring rule remains localizing because the transformation can be written as a transformation of the minimal localization $\mathrm{F}^\flat_A$, that is, $\mathrm{F}^\flat_{A,r_1,r_2} := \mathrm{F}_A + \bar{F}_A\big(\gamma\delta_{r_1} + (1 - \gamma)\delta_{r_2}\big)$. Following the reasoning of Example D.1, there exists a measure on $\mathcal{B}$ equivalent to $\mathrm{F}^\flat_{A,r_1,r_2}$ for which the distribution function exists, which is precisely the aim of the second transformation. Since the second transformation does not depend on F, and hence is the same for all distributions in $\mathcal{P}$, the censored scoring rule based on $\mathrm{F}^{\flat}_{A,r_1,r_2}$ remains localizing.

Example D.2. *A tuber grower models the temperature Y in his greenhouse using a continuous distribution F on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Focusing on temperatures near the optimal temperature of 18 degrees Celsius, he considers the interval $A = (r_1, r_2)$, with e.g., $r_1 = 16$ and $r_2 = 20$. The region of interest renders the minimal σ-algebra $\mathcal{B}^{\flat}_A = \mathcal{B}((r_1, r_2))$ on $\mathbb{R}$, which includes $A^c = (-\infty, r_1] \cup [r_2, \infty)$, but does neither include $(-\infty, r_1 - v_1]$ nor $[r_2 + v_2, \infty)$, $\forall v_1, v_2 \geq 0$. Consequently, $\mathrm{F}^{\flat}_A$, the restricted distribution of F on $\mathcal{B}^{\flat}_A$, does not admit a distribution function. Unlike in Example D.1, there is no equivalent measure of $\mathrm{F}^{\flat}_A$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. However, when extending $\mathrm{F}^{\flat}_A$ on $(\mathbb{R}, \mathcal{B}^{\flat}_A)$ to $\mathrm{F}^{\flat}_A$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, it is natural to assign zero probability to the events $[r_2 + v_2, \infty)$, $\forall v_1, v_2 > 0$. The total probability $\bar{F}_w$ assigned to $(-\infty, r_1] \cup [r_2, \infty)$ is then distributed between $(-\infty, r_1]$ and $[r_2, \infty)$ according to a tuning parameter $\gamma \in [0, 1]$: $\mathrm{F}^{\flat}_A((-\infty, r]) = \gamma \bar{F}_w$ and $\mathrm{F}^{\flat}_A([r_2, \infty)) = (1 - \gamma)\bar{F}_w$.*

The associated σ-algebra $\tilde{\mathcal{G}}_A$ of the generalized censored measure $\mathrm{F}^{\flat}_{A,r_1,r_2}$ is closely related to the minimal σ-algebra $\mathcal{G}^{\flat}_A$. However, the measure $\mathrm{F}^{\flat}_{A,r_1,r_2}$ is generally not the restriction of F to $\tilde{\mathcal{G}}_A$ due to the distribution of $\bar{F}_A$. It would have been, were $\gamma_{\mathrm{F}} = \mathrm{F}((A^c)_1)/\bar{f}_A$ but this choice induces a localization bias, meaning that outcomes outside A that are not implied by A affect the scoring rule, which is undesirable if the region of interest is A.

For $w(y) = \mathbb{1}_{(r_1,r_2)}(y)$, one can show that the twCRPS of Gneiting and Ranjan (2011) is equivalent to the generalized censored scoring rule based on $\mathrm{F}^{\flat}_{A,r_1,r_2}$ with $\gamma_{\mathrm{F}} = \mathrm{F}((A^c)_1)/\bar{F}_A$. Example D.3 illustrates the consequences of its non-localizing nature. Even if a candidate measure F coincides with P on the region of interest, the scoring rule can favor a different candidate G different from the true distribution P on the region of interest. As this bias towards G is a result of the scoring rule being non-localizing it is referred to as a localization bias.

Example D.3 (Localization bias). *Let* Y *be a random variable that follows a piecewise uniform distribution across the intervals* $A = [0, 1)$, $B = [1, 2)$ *and* $C = [2, 3]$, *with probabilities* π_A, π_B *and* π_C, *respectively. Figure D.1 displays the densities and distribution functions of the true distribution* P *and two candidates* F *and* G. *Suppose that the region of interest is* B, *with corresponding weight function* $w(y) = \mathbb{1}_B(y)$, *and note that* $\mathrm{P} \overset{B}{=} \mathrm{F}$, *while* P *and* G *differ on* B. *A prevalent weighted version of the CRPS is given by* $\mathrm{twCRPS}(F, y) = \int_B \big(F(s) - \mathbb{1}_{[y,\infty)}(s)\big)^2 \mathrm{d}s$, *with score divergence* $\mathbb{D}_{\mathrm{twCRPS}}(\mathrm{F}\|\mathrm{G}) = \int_B \big(F(s) - G(s)\big)^2 \mathrm{d}s$; *see Gneiting and Ranjan (2011). This weighted variant of the CRPS is clearly non-localizing, for instance, because its value depends on* $\mathrm{F}(A)$, *while* $\mathrm{F}(A)$ *is not implied by* $\mathrm{F}(B)$, *only the sum* $\mathrm{F}(A) + \mathrm{F}(C)$ *is. Its failure to be localizing introduces an inconsistency in evaluating distributions over the region* B. *Indeed, by accounting for behavior of* F *and* G *on* A *(i.e., outside* B*) where* G *is closer to* P *than* F *(see Figure D.1), the twCRPS favors* G *on* B.

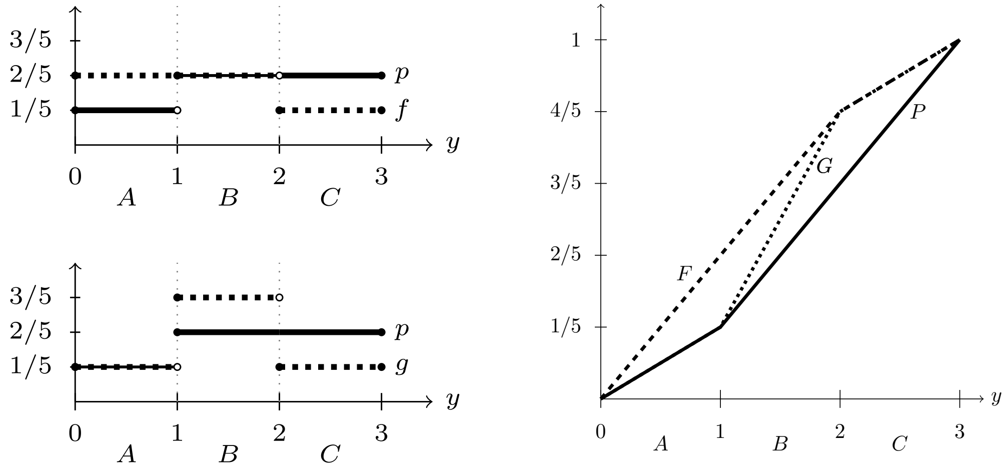


Figure D.1: Densities (left) and distribution functions (right) of distributions F, G and true distribution P, all piecewise uniformly distributed on $[0, 3]$ but with different probabilities $\boldsymbol{\pi} := (\pi_A, \pi_B, \pi_C)'$. Specifically, $\boldsymbol{\pi}_p = (1/5, 2/5, 2/5)'$, $\boldsymbol{\pi}_f = (2/5, 2/5, 1/5)'$ and $\boldsymbol{\pi}_g = (1/5, 3/5, 1/5)'$.

The localization bias towards G in Example D.3 is undesirable if the researcher is solely concerned with outcomes in B. In practice, however, interest might extend beyond region B alone, encompassing outcomes located specifically in either region A or C, rather than their combined complement $B^c = A \cup C$. Put differently, a researcher might have explicit interest in events within the collection $\tilde{\mathcal{B}}_A$, which cannot be adequately represented by a weight function defined on the outcome space alone. Consequently, in those situations non-localizing scoring rules, such as the threshold-weighted continuous ranked probability score (twCRPS), may be better suited for evaluating predictive accuracy when the researcher's focus does not correspond precisely to a predefined weight function w. However, in this paper, consistent with the frameworks presented by Diks et al. (2011), Holzmann and Klar (2017a), and Allen et al. (2023), we maintain the assumption that the researcher's interests are fully captured by a given weight function defined on the outcome space.

E Derivations for Specific Scoring Rules

In the following subsections, we explicitly derive the results corresponding to kernel scores and the results summarized in Table 1. The scoring rules in Table 1 depend on densities and hence their conditional and censored counterparts on the conditional and censored densities, which are given by $f_w^\sharp(y) = w(y)f(y)/(1-\bar{F}_w)$, $y \in \mathcal{Y}_w^\sharp$ and $f_w^\flat(y) = w(y)f(y) + \bar{F}_w \mathbb{1}_{\{*\}}(y)$, $\mathcal{Y}_w^\flat$, respectively. The assumption on the nuisance density h in the caption of Table 1, that is, its support is a subset of $A^c \subseteq \mathcal{Y}$, implies that $w(y)h(y) = 0$, $\forall y \in \mathcal{Y}$. Additionally observing that $f_w(y) = 0, \forall y \in A^c$, this facilitates the simplification of the expressions below. In the derivations, $S(\tilde{f}, \tilde{y})$ denotes the score of the real-valued random variable $\tilde{Y} = bY + a$, where $a \in \mathbb{R}$ and $b \in \mathbb{R}\backslash\{0\}$, with density $\tilde{f}(\tilde{y}) = \frac{1}{|b|} f\big(\frac{\tilde{y}-a}{b}\big)$. Since the results for the focused scoring rules hold by means of having the same expected score differences, a.s.-equivalent scoring rules and candidate distribution independent additive terms can be neglected, denoted by '$\overset{\text{a.s.}}{=}$' and '$\overset{\text{eqv.}}{=}$', respectively. To save space, some obvious results are omitted.

E.1 Logarithmic Score (LogS)

Following the order in which the assertions pertaining to the Logarithmic scoring rule $\mathrm{LogS}(f, y) = \log f(y)$ appear in Table 1, they can be easily verified as follows:

$$
\begin{aligned}
\mathrm{LogS}(\tilde{f}, \tilde{y}) &= \log \tilde{f}(\tilde{y}) = \log f(y) - \log|b| \overset{\text{eqv.}}{=} \log f(y), \\
\mathrm{LogS}^{\sharp}_{w}(f, y) &= w(y) \log\left(\frac{w(y)f(y)}{1 - \bar{F}_w}\right) \\
&\overset{\text{eqv.}}{=} w(y) \log\left(\frac{f(y)}{1 - \bar{F}_w}\right) \\
&= S^{\mathrm{CL}}_{w}(f, y),
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{LogS}^{\flat}_{w}(f, y) &= w(y) \log\big(w(y)f(y) + \bar{F}_w \mathbb{1}_{\{*\}}(y)\big) + \big(1 - w(y)\big) \log \bar{F}_w \\
&\overset{\text{a.s.}}{=} w(y) \log\big(w(y)f(y)\big) + \big(1 - w(y)\big) \log \bar{F}_w \\
&\overset{\text{eqv.}}{=} w(y) \log\big(f(y)\big) + \big(1 - w(y)\big) \log \bar{F}_w \\
&= S^{\mathrm{CSL}}_{w}(f, y), \\
\mathrm{LogS}^{\flat}_{w,h}(f, y) &= w(y) \log f^{\flat}_{w,h}(y) + \big(1 - w(y)\big) \int_{\mathcal{Y}} \log f^{\flat}_{w,h}(q) h(q) \mu(\mathrm{d}q) \\
&= w(y)\Big(\log\big(f_w(y)\big) \mathbb{1}_{A_w}(y) + \log\big(\bar{F}_w h(y)\big) \mathbb{1}_{A^c_w}(y) \Big) \\
&\qquad + \big(1 - w(y)\big) \int_{A^c_w} \Big(\log\big(f_w(q)\big) \mathbb{1}_{A_w}(y) + \log\big(\bar{F}_w h(q)\big) \mathbb{1}_{A^c_w}(y) \Big) h(q) \mu(\mathrm{d}q) \\
&= w(y) \log f_w(y) + \big(1 - w(y)\big) \int_{A^c_w} \log\big(\bar{F}_w h(q)\big)\, h(q) \mu(\mathrm{d}q) \\
&\overset{\text{eqv.}}{=} w(y) \log f(y) + \big(1 - w(y)\big) \log \bar{F}_w \\
&= S^{\mathrm{CSL}}_{w}(f, y).
\end{aligned}
$$

E.2 Power Scores (PowS_α)

For results related to the PowS_α family $\mathrm{PowS}_\alpha(f, y) = \alpha f(y)^{\alpha-1} - (\alpha - 1)\|f\|_\alpha^\alpha$, where $\alpha > 1$, we start by verifying that

$$\begin{aligned}
\mathrm{PowS}_\alpha(\tilde{f}, \tilde{y}) &= \alpha\big(\tilde{f}(\tilde{y})\big)^{\alpha-1} - (\alpha - 1)\|\tilde{f}\|_\alpha^\alpha \\
&= \alpha \left(\frac{1}{|b|}\right)^{\alpha-1} f(y) - (\alpha - 1) \left(\frac{1}{|b|}\right)^{\alpha-1} \|f\|_\alpha^\alpha \\
&= \left(\frac{1}{|b|}\right)^{\alpha-1} \mathrm{PowS}_\alpha(f, y),
\end{aligned}$$

for which we rely on the result expressed in

$$\begin{aligned}
\|\tilde{f}\|_\alpha^\alpha &= \int_{\tilde{\mathcal{Y}}} \tilde{f}(\tilde{y})^\alpha \mu(\mathrm{d}\tilde{y}) \\
&= \left(\frac{1}{|b|}\right)^{\alpha-1} \int_{\tilde{\mathcal{Y}}} \left(f\left(\frac{\tilde{y} - a}{b}\right)\right)^\alpha \frac{1}{|b|} \mu(\mathrm{d}\tilde{y}) \\
&= \left(\frac{1}{|b|}\right)^{\alpha-1} \int_{\mathcal{Y}} \left(f\left(y\right)\right)^\alpha \mu(\mathrm{d}y) \\
&= \left(\frac{1}{|b|}\right)^{\alpha-1} \|f\|_\alpha^\alpha.
\end{aligned}$$

Next, we verify the limit for the non-focused family. Specifically, the following affine transformation of the score satisfies

$$\begin{aligned}
\lim_{\alpha \downarrow 1} \frac{1}{\alpha - 1} \left(\mathrm{PowS}_\alpha(f, y) - 1\right) &= \lim_{\alpha \downarrow 1} \frac{1}{\alpha - 1} \big(\alpha f(y)^{\alpha-1} - (\alpha - 1)\|f\|_\alpha^\alpha - \alpha + (\alpha - 1)\big) \\
&= \lim_{\alpha \downarrow 1} \frac{\alpha f(y)^{\alpha-1} - \alpha}{\alpha - 1} - \lim_{\alpha \downarrow 1} \|f\|_\alpha^\alpha + 1 \\
&= \lim_{\alpha \downarrow 1} \frac{\alpha f(y)^{\alpha-1} - \alpha}{\alpha - 1} \\
&= \left[\frac{f(y)^{\alpha-1} + \alpha f(y)^{\alpha-1} \log f(y) - 1}{1}\right]_{\alpha=1} \\
&= \log f(y),
\end{aligned}$$

where in the before last step l'Hôpital's rule was used. Furthermore, the conditional version of the PowS_α family displayed in Table 1 is nothing but a direct application of the conditioning procedure.

By the linearity of limits, we obtain

$$\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\big(\mathrm{PowS}^{\sharp}_{\alpha}(f,y)-w(y)\big)=w(y)\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\big(\mathrm{PowS}_{\alpha}(f^{\sharp}_{w},y)-1\big)=\mathrm{LogS}^{\sharp}_{w}(f,y).$$

Turning to the censored focusing method, we note that

$$\|f^{\flat}_{w}\|^{\alpha}_{\alpha}=\int_{\mathcal{Y}}\big(f_w(y)+\bar{F}_w\mathbb{1}_{\{*\}}(y)\big)^{\alpha}(\mu+\delta_*)(\mathrm{d}y)=\|f_w(y)\|^{\alpha}_{\alpha}+\bar{F}^{\alpha}_{w}. \tag{E.1}$$

Using this result, we obtain

$$\begin{aligned}\mathrm{PowS}^{\flat}_{\alpha,w}(f,y)&=w(y)\alpha\left(f_w(y)+\bar{F}_w\mathbb{1}_{\{*\}}(y)\right)^{\alpha-1}+\big(1-w(y)\big)\alpha\bar{F}^{\alpha-1}_{w}-(\alpha-1)\|f^{\flat}_{w}\|^{\alpha}_{\alpha}\\&\overset{\text{a.s.}}{=}w(y)\alpha f_w(y)^{\alpha-1}+\big(1-w(y)\big)\alpha\bar{F}^{\alpha-1}_{w}-(\alpha-1)\left(\|f_w\|^{\alpha}_{\alpha}+\bar{F}^{\alpha}_{w}\right),\end{aligned}$$

which bears the following limit

$$\begin{aligned}&\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\Big(\mathrm{PowS}^{\flat}_{\alpha,w}(f,y)-1\Big)\\&=\frac{1}{1-\alpha}\left(w(y)\alpha f_w(y)^{\alpha-1}+\big(1-w(y)\big)\alpha\bar{F}^{\alpha-1}_{w}-(\alpha-1)\left(\|f_w\|^{\alpha}_{\alpha}+\bar{F}^{\alpha}_{w}\right)-\alpha+(\alpha-1)\right)\\&=w(y)\lim_{\alpha\downarrow 1}\frac{\alpha f_w(y)^{\alpha-1}-\alpha}{(\alpha-1)}+(1-w(y))\lim_{\alpha\downarrow 1}\frac{\alpha\bar{F}^{\alpha-1}_{w}-\alpha}{(\alpha-1)}-\lim_{\alpha\downarrow 1}\left(\|f_w\|^{\alpha}_{\alpha}+\bar{F}^{\alpha}_{w}\right)+1\\&=w(y)\left[\frac{f_w(y)^{\alpha-1}+\alpha f_w(y)^{\alpha-1}\log f_w(y)-1}{1}\right]_{\alpha=1}\\&\qquad+(1-w(y))\left[\frac{\bar{F}^{\alpha-1}_{w}+\alpha\bar{F}^{\alpha-1}_{w}\log\bar{F}_w-1}{1}\right]_{\alpha=1}\\&=w(y)\log f_w(y)+(1-w(y))\log\bar{F}_w\\&=\mathrm{LogS}^{\flat}_{w}(f,y),\end{aligned}$$

where in the third equality l'Hôpital's rule was used.

E.3 PseudoSpherical Scores (PsSphS_α)

As for LogS, we follow the order of the table for the derivations concerning the PseudoSpherical family $\mathrm{PsSphS}_\alpha(f,y) = \frac{f(y)^{\alpha-1}}{\|f\|_\alpha^{\alpha-1}}$, where $\alpha > 1$. In particular,

$$\mathrm{PsSphS}_\alpha(\tilde{f},\tilde{y}) = \frac{\tilde{f}(\tilde{y})^{\alpha-1}}{\|\tilde{f}\|_\alpha^{\alpha-1}} = \frac{\left(\frac{1}{|b|}\right)^{\alpha-1} f(y)^{\alpha-1}}{\left(\frac{1}{|b|}\right)^{\frac{(\alpha-1)^2}{\alpha}} \|f\|_\alpha^{\alpha-1}} = \left(\frac{1}{|b|}\right)^{\frac{\alpha-1}{\alpha}} \mathrm{PsSphS}_\alpha(f,y).$$

Next, we consider the limit as $\alpha \downarrow 1$. Shifting the PsSphS_α family by 1 and rescaling it by a factor $\frac{1}{\alpha-1}$, we obtain

$$\begin{aligned}
&\lim_{\alpha\downarrow 1} \frac{1}{\alpha-1}\left(\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-1} - 1\right) \\
&= \left[\left(\log\left(\frac{f(y)}{\|f\|_\alpha}\right) - (\alpha-1)\frac{1}{\|f\|_\alpha}\frac{\partial}{\partial\alpha}\|f\|_\alpha\right)\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-1}\right]_{\alpha=1} \\
&= \log f(y), \qquad \text{(E.2)}
\end{aligned}$$

where in the first step l'Hôpital's rule was used with the derivative

$$\frac{\partial}{\partial\alpha}\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-1} = \left(\log\left(\frac{f(y)}{\|f\|_\alpha}\right) - (\alpha-1)\frac{1}{\|f\|_\alpha}\frac{\partial}{\partial\alpha}\|f\|_\alpha\right)\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-1},$$

and in the second $\|f\|_1 = 1$ and that the partial derivative of $\|f\|_\alpha$ with respect to α is finite at $\alpha = 1$, which is the case if and only if $\int_{\mathcal{Y}} \log f(y) f(y)\,\mu(\mathrm{d}y)$ is finite, because

$$\begin{aligned}
&\left[\frac{\partial}{\partial\alpha}\left(\int_{\mathcal{Y}} f(y)^{\alpha}\,\mu(\mathrm{d}y)\right)^{\frac{1}{\alpha}}\right]_{\alpha=1} \\
&= \left[\frac{\partial}{\partial\alpha}\left(\mathrm{e}^{\frac{1}{\alpha}\log\int_{\mathcal{Y}} f(y)^{\alpha}\,\mu(\mathrm{d}y)}\right)\right]_{\alpha=1} \\
&= \left[\frac{\partial}{\partial\alpha}\left(\frac{1}{\alpha}\log\int_{\mathcal{Y}} f(y)^{\alpha}\,\mu(\mathrm{d}y)\right)\left(\int_{\mathcal{Y}} f(y)^{\alpha}\,\mu(\mathrm{d}y)\right)^{\frac{1}{\alpha}}\right]_{\alpha=1} \\
&= \left[\frac{\partial}{\partial\alpha}\left(\frac{1}{\alpha}\log\int_{\mathcal{Y}} f(y)^{\alpha}\,\mu(\mathrm{d}y)\right)\right]_{\alpha=1} \\
&= \left[-\frac{1}{\alpha^2}\log\left(\int_{\mathcal{Y}} f(y)^{\alpha}\,\mu(\mathrm{d}y)\right)+\frac{1}{\alpha}\frac{\int_{\mathcal{Y}}\log f(y)f(y)^{\alpha}\,\mu(\mathrm{d}y)}{\int_{\mathcal{Y}} f(y)^{\alpha}\,\mu(\mathrm{d}y)}\right]_{\alpha=1} \qquad \text{(E.3)}
\end{aligned}$$

under regularity assumptions permitting us to move the partial derivative under the integral sign using Leibniz' rule. Then, evaluating the derivative in $\alpha = 1$ yields

$$\left[\frac{\partial}{\partial\alpha}\left(\int_{\mathcal{Y}} f(y)^{\alpha}\,\mu(\mathrm{d}y)\right)^{\frac{1}{\alpha}}\right]_{\alpha=1} = -\log 1 + \int_{\mathcal{Y}}\log f(y)f(y)\,\mu(\mathrm{d}y) = \int_{\mathcal{Y}}\log f(y)f(y)\,\mu(\mathrm{d}y),$$

For the conditional PsSphS_{α} family, we find

$$\begin{aligned}
\mathrm{PsSphS}^{\sharp}_{\alpha,w}(f,y) &= w(y)\frac{\left(\frac{f_w(y)}{1-\bar{F}_w}\right)^{\alpha-1}}{\left(\int_{\mathcal{Y}}\left(\frac{f_w}{1-\bar{F}_w}\right)^{\alpha}\mathrm{d}\mu\right)^{\frac{\alpha-1}{\alpha}}} \\
&= w(y)\frac{f_w(y)^{\alpha-1}}{\|f_w\|_{\alpha}^{\alpha-1}} \\
&= w(y)\left(\frac{f_w(y)^{\alpha}}{\|f_w\|_{\alpha}^{\alpha}}\right)^{\frac{\alpha-1}{\alpha}}.
\end{aligned}$$

By the linearity of limits, we find

$$\begin{aligned}\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\big(\mathrm{PsSphS}^{\sharp}_{\alpha}(f,y)-w(y)\big) &= w(y)\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\big(\mathrm{PsSphS}_{\alpha}(f^{\sharp}_{w},y)-1\big)\\ &= w(y)\log f^{\sharp}_{w}(y)\\ &= \mathrm{LogS}^{\sharp}_{w}(f,y).\end{aligned}$$

Moreover, for the censored PsSphS_{α} family, we recall Equation (E.1) and find that

$$\begin{aligned}\mathrm{PsSphS}^{\flat}_{w}(f,y) &= \frac{w(y)\left(f_w(y)+\bar{F}_w\mathbb{1}_{\{*\}}(y)\right)^{\alpha-1}+\big(1-w(y)\big)\bar{F}_w^{\alpha-1}}{(\|f^{\flat}_{w}\|^{\alpha}_{\alpha})^{\frac{\alpha-1}{\alpha}}}\\ &= \frac{w(y)\left(f_w(y)^{\alpha-1}+\bar{F}_w^{\alpha-1}\mathbb{1}_{\{*\}}(y)\right)+\big(1-w(y)\big)\bar{F}_w^{\alpha-1}}{\big(\|f_w(y)\|^{\alpha}_{\alpha}+\bar{F}_w^{\alpha}\big)^{\frac{\alpha-1}{\alpha}}}\\ &\overset{\text{a.s.}}{=} \frac{w(y)f_w(y)^{\alpha-1}+\big(1-w(y)\big)\bar{F}_w^{\alpha-1}}{\big(\|f_w(y)\|^{\alpha}_{\alpha}+\bar{F}_w^{\alpha}\big)^{\frac{\alpha-1}{\alpha}}}.\end{aligned}$$

For the limit as $\alpha \downarrow 1$, we obtain an analogous result, namely

$$\begin{aligned}\lim_{\alpha\downarrow 1}\tfrac{1}{\alpha-1}\Big(\mathrm{PsSphS}^{\flat}_{w}(f,y)-1\Big) &= w(y)\lim_{\alpha\downarrow 1}\tfrac{1}{\alpha-1}\left(\left(\frac{f_w(y)}{\big(\|f_w\|^{\alpha}_{\alpha}+\bar{F}_w^{\alpha}\big)^{\frac{1}{\alpha}}}\right)^{\alpha-1}-1\right)\\ &\quad+\big(1-w(y)\big)\lim_{\alpha\downarrow 1}\tfrac{1}{\alpha-1}\left(\left(\frac{\bar{F}_w}{\big(\|f_w\|^{\alpha}_{\alpha}+\bar{F}_w^{\alpha}\big)^{\frac{1}{\alpha}}}\right)^{\alpha-1}-1\right).\end{aligned} \tag{E.4}$$

For the first term on the right-hand side of this equation we calculate the derivative

$$\begin{aligned}
&\left[\frac{\partial}{\partial\alpha}\left(\frac{f_w(y)}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)^{\alpha-1}\right]_{\alpha=1} \\
&=\left[\frac{\partial}{\partial\alpha}\log\left(\frac{f_w(y)}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)^{\alpha-1}\right]_{\alpha=1}\left[\left(\frac{f_w(y)}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)^{\alpha-1}\right]_{\alpha=1} \\
&=\left[\frac{\partial}{\partial\alpha}\left((\alpha-1)\left(\log f_w(y)-\frac{1}{\alpha}\log\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)\right)\right)\right]_{\alpha=1} \\
&=\left[\log\left(\frac{f_w(y)}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)-(\alpha-1)\frac{\partial}{\partial\alpha}\left(\log\left(\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}\right)\right)\right]_{\alpha=1} \\
&=\log f_w(y),
\end{aligned}$$

where we used that $\|f_w\|_1+\bar{F}_w=1-\bar{F}_w+\bar{F}_w=1$ and that the partial derivative of $\|f_w^\flat\|_\alpha=\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}$ with respect to α is finite at $\alpha=1$, which by analogy with Equation (E.3), is the case if and only if $\int_{\mathcal{Y}}\log f_w^\flat(y)f_w^\flat(y)\,(\mu+\delta_*)(\mathrm{d}y)$ is finite. Likewise, for the second term in Equation (E.4) we find the derivative

$$\left[\frac{\partial}{\partial\alpha}\left(\frac{\bar{F}_w}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)^{\alpha-1}\right]_{\alpha=1}=\log\bar{F}_w.$$

Using these derivatives when applying l'Hôpital's rule to obtain the limit in Equation (E.4), gives

$$\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\left(\mathrm{PsSphS}_w^\flat(f,y)-1\right)=w(y)\log f_w(y)+(1-w(y))\log\bar{F}_w.$$

E.4 Kernel Scores (S_ρ)

This subsection provides more details on Example 4. Consider a class of distributions $\mathcal{P}_r$ on some measurable space $(\mathcal{Y},\mathcal{G})$, such that $\mathrm{F}(r)=0,\forall\mathrm{F}\in\mathcal{P}_r$, where $r\in\mathcal{Y}$, including all continuous distributions on $\mathcal{Y}$. Consider the kernel score family $S_\rho(\mathrm{F},y)=\frac{1}{2}\mathbb{E}_\mathrm{F}\rho(X,X')-$

$\mathbb{E}_{\mathrm{F}}\rho(X,y)+\frac{1}{2}\rho(y,y)$ with divergence $\mathbb{D}_{S_\rho}(\mathrm{F}\|\mathrm{G}) = \mathbb{E}_{\mathrm{F},\mathrm{G}}\rho(X,Y)-\frac{1}{2}\mathbb{E}_{\mathrm{F}}\rho(X,X')-\frac{1}{2}\mathbb{E}_{\mathrm{G}}\rho(Y,Y')$, where $X, X' \sim \mathrm{F}$ and $Y, Y' \sim \mathrm{G}$ are independent. Here, we use $\mathbb{E}_{\mathrm{F},\mathrm{G}}\varrho(X,Y)$ as short-hand notation for the expectation with respect to the product measure $\mathrm{F}\otimes\mathrm{G}$, and $\mathbb{E}_{\mathrm{F}}\varrho(X,X')$ for the expectation with respect to $\mathrm{F}\otimes\mathrm{F}$, for any measurable function ϱ. The associated generalized censored scoring rule for $\mathrm{H} = \delta_r$ reads

$$\begin{aligned}
S^\flat_{\rho,w}(\mathrm{F},y) &= \frac{1}{2}\mathbb{E}_{\mathrm{F}^\flat_w}\rho(X,X') - w(y)\left(\mathbb{E}_{\mathrm{F}^\flat_w}\rho(X,y) - \frac{1}{2}\rho(y,y)\right) \\
&\quad - \big(1-w(y)\big)\left(\mathbb{E}_{\mathrm{F}^\flat_w}\rho(X,r) - \frac{1}{2}\rho(r,r)\right) \\
&= \frac{1}{2}\left(\mathbb{J}_{\mathrm{F}_w}\rho(X,X') + 2\bar{F}_w\mathbb{J}_{\mathrm{F}_w}\rho(X,r) + \bar{F}_w^2\rho(r,r)\right) \\
&\quad - w(y)\left(\mathbb{J}_{\mathrm{F}_w}\rho(X,y) + \bar{F}_w\rho(r,y) - \frac{1}{2}\rho(y,y)\right) \\
&\quad - \big(1-w(y)\big)\left(\mathbb{J}_{\mathrm{F}_w}\rho(X,r) + \bar{F}_w\rho(r,r) - \frac{1}{2}\rho(r,r)\right),
\end{aligned}$$

where $\mathbb{J}_{\mathrm{F}_w}q(Y) := \int_{\mathcal{Y}} q(y)\mathrm{F}_w(\mathrm{d}y)$, for any measurable function q, is the integral generalizing the expectation operator to allow for integration with respect to any weighted kernel F_w, for which we adopt similar short-hand notation as for the expectations $\mathbb{E}_{\mathrm{F},\mathrm{G}}\varrho(X,Y)$ and

$\mathbb{E}_{\mathrm{F}}\varrho(X, X')$ introduced above. The corresponding score divergence reads

$$
\begin{aligned}
\mathbb{D}_{S^\flat_{\rho,w}}(\mathrm{F}\|\mathrm{G}) &\equiv \mathbb{D}_{S_\rho}(\mathrm{F}^\flat_w\|\mathrm{G}^\flat_w) \\
&= \mathbb{E}_{\mathrm{F}^\flat_w,\mathrm{G}^\flat_w}\rho(X,Y) - \frac{1}{2}\mathbb{E}_{\mathrm{F}^\flat_w}\rho(X,X') - \frac{1}{2}\mathbb{E}_{\mathrm{G}^\flat_w}\rho(Y,Y') \\
&= \mathbb{J}_{\mathrm{F}_w,\mathrm{G}_w}\rho(X,Y) + \bar{G}_w\mathbb{J}_{\mathrm{F}_w}\rho(X,r) + \bar{F}_w\mathbb{J}_{\mathrm{G}_w}\rho(r,Y) + \bar{F}_w\bar{G}_w\rho(r,r) \\
&\quad - \frac{1}{2}\mathbb{J}_{\mathrm{F}_w}\rho(X,X') - \bar{F}_w\mathbb{J}_{\mathrm{F}_w}\rho(X,r) - \frac{1}{2}\bar{F}^2_w\rho(r,r) \\
&\quad - \frac{1}{2}\mathbb{J}_{\mathrm{G}_w}\rho(Y,Y') - \bar{G}_w\mathbb{J}_{\mathrm{G}_w}\rho(r,Y) - \frac{1}{2}\bar{G}^2_w\rho(r,r) \\
&= \mathbb{J}_{\mathrm{F}_w,\mathrm{G}_w}\rho(X,Y) - \frac{1}{2}\mathbb{J}_{F_w}\rho(X,X') - \frac{1}{2}\mathbb{J}_{G_w}\rho(Y,Y') \\
&\quad - (\bar{F}_w - \bar{G}_w)\left(\mathbb{J}_{\mathrm{F}_w}\rho(X,r) - \mathbb{J}_{\mathrm{G}_w}\rho(r,Y)\right) - \frac{1}{2}(\bar{F}_w - \bar{G}_w)^2\rho(r,r).
\end{aligned}
$$

Assumption 1 is satisfied for all weight functions and distributions $\mathrm{F} \in \mathcal{P}_r$ since $\mathrm{F}_w(r) = 0$. Therefore, the score divergence is a localized divergence if S_ρ is strictly proper relative to $\mathcal{P}^\flat_r$, which follows from the conditions under which S_ρ is strictly proper with respect to $\mathcal{P}_r$.

We verify this condition for a popular subclass of kernel scores known as the Energy Score (ES) family. The ES family of scoring rules is defined as

$$
\mathrm{ES}_\beta(\mathrm{F}, y) := \frac{1}{2}\mathbb{E}_{\mathrm{F}}\|\mathbf{Y} - \mathbf{Y}'\|^\beta_2 - \mathbb{E}_{\mathrm{F}}\|\mathbf{Y} - \mathbf{y}\|^\beta_2,
$$

where $\beta \in (0, 2)$, $\|\cdot\|_2$ denotes the Euclidean norm, and $\mathbf{Y}$ and $\mathbf{Y}'$ denote independent copies of a random vector with distribution $\mathrm{F} \in \mathcal{P}_\beta$, the class of Borel probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that $\mathbb{E}_{\mathrm{F}}\|\mathbf{Y}\|^\beta_2 < \infty$, relative to which $\mathrm{ES}_\beta(\mathrm{F}, y)$ is strictly proper (Gneiting and Raftery 2007, Section 4.3). We consider the subclass $\tilde{\mathcal{P}}_\beta \subseteq \mathcal{P}_\beta$ of continuous distributions, in which case the use of any collection of pivotal points is allowed by Assumption 1. Take k arbitrary pivotal points by choosing $\mathrm{H} = \sum_{i=1}^k \gamma_i\delta_{\mathbf{r}_i}$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)' \in \Delta(k)$

and $\mathbf{r}_i \in \mathbb{R}^d, \forall i$. The verification of $\tilde{\mathcal{P}}_\beta^\flat \subseteq \tilde{\mathcal{P}}_\beta$ follows from

$$\mathbb{E}_{\mathrm{F}^\flat_{w,\mathcal{R}_k}} \|\mathbf{Y}\|_2^\beta = \int_{\mathbb{R}^d} \|\mathbf{y}\|_2^\beta \mathrm{F}_w(\mathrm{d}\mathbf{y}) + \bar{F}_w \sum_{i=1}^k \gamma_i \|\mathbf{r}_i\|_2^\beta < \mathbb{E}_\mathrm{F}\|\mathbf{Y}\|_2^\beta + \sum_{i=1}^k \|\mathbf{r}_i\|_2^\beta < \infty,$$

$\forall \mathrm{F} \in \mathcal{P}_\beta$, where $\mathcal{R}_k = \{\mathbf{r}_i\}_{i=1}^k$. Consequently,

$$\mathrm{ES}^\flat_{\beta,w}(\mathrm{F}, y) = \frac{1}{2}\mathbb{E}_{\mathrm{F}^\flat_{w,\mathcal{R}_k}} \|\mathbf{Y} - \mathbf{Y}'\|_2^\beta - w(\mathbf{y})\mathbb{E}_{\mathrm{F}^\flat_{w,\mathcal{R}_k}} \|\mathbf{Y} - \mathbf{y}\|_2^\beta - \big(1 - w(\mathbf{y})\big) \sum_{i=1}^k \mathbb{E}_{\mathrm{F}^\flat_{w,\mathcal{R}_k}} \|\mathbf{Y} - \mathbf{r}_i\|_2^\beta.$$

is strictly locally proper with respect to $\tilde{\mathcal{P}}_\beta$, for all weight functions.

To facilitate comparisons with the threshold-weighted kernel score framework proposed by Allen et al. (2023), we provide similar calculations for their approach. By Propositions 4.4 and 4.5 in Allen et al. (2023) the threshold-weighted kernel score $\mathrm{twS}_\rho(\mathrm{F}, y; v) = \mathbb{E}_\mathrm{F}\left[\rho\big(v(X), v(y)\big)\right] - \frac{1}{2}\mathbb{E}_\mathrm{F}\left[\rho(v(X), v(X'))\right] - \frac{1}{2}\rho(v(y), v(y))$, is strictly locally proper with respect to w if and only if $\rho\big(v(z), v(\cdot)\big) = \rho\big(v(z'), v(\cdot)\big), \forall z, z' \in A_w^\mathrm{c}$ and the restriction of v to A_w is injective. Therefore, assume that v is injective on A_w. Moreover, as suggested by Allen et al. (2023), restrict $v(z') = v(z) = y_0, \forall z, z' \in A_w^\mathrm{c}$, for some arbitrary $y_0 \in \mathcal{Y}$.

For the scoring rule, first consider the case where $y \in A_w$, in which case it reduces to

$$\begin{aligned}\mathrm{twS}_\rho(\mathrm{F}, y; v, y_0)|y \in A_w = {} & \mathbb{J}_{\mathrm{F}_w}\rho\big(v(X), v(y)\big) + \bar{F}_w\rho\big(y_0, v(y)\big) - \bar{F}_w\mathbb{J}_{\mathrm{F}_w}\rho\big(v(X), y_0\big) \\ & - \frac{1}{2}\bar{F}_w^2\rho(y_0, y_0) - \frac{1}{2}\rho\big(v(y), v(y)\big) - \frac{1}{2}\mathbb{J}_{\mathrm{F}_w}\rho\big(v(X), v(X')\big).\end{aligned}$$

Second, if $y \in A_w^c$, the scoring rule reads

$$\begin{aligned}\mathrm{twS}_\rho(\mathrm{F}, y; v, y_0)|y \in A_w^c = {} & (1 - \bar{F}_w)\mathbb{J}_{\mathrm{F}_w}\rho\big(v(X), y_0\big) - \frac{1}{2}(\bar{F}_w - 1)^2\rho\big(y_0, y_0\big) \\ & - \frac{1}{2}\mathbb{J}_{\mathrm{F}_w}\rho\big(v(X), v(X')\big).\end{aligned}$$

Put together, this yields the following class of weighted kernel scores

$$\begin{aligned}\mathrm{twS}_\rho(\mathrm{F}, y; v, y_0) = {} & \frac{1}{2}\mathbb{J}_{\mathrm{F}_{A_w}}\rho\big(v(X), v(X')\big) - \Big(\mathbb{J}_{\mathrm{F}_{A_w}}\rho\big(v(X), v(y)\big) + \bar{F}_w\rho\big(y_0, v(y)\big) \\ & - \bar{F}_w\mathbb{J}_{\mathrm{F}_{A_w}}\rho\big(v(X), y_0\big) - \frac{1}{2}\bar{F}_w^2\rho\big(y_0, y_0\big) - \frac{1}{2}\rho\big(v(y), v(y)\big)\Big)\mathbb{1}_{A_w}(y) \\ & - \Big((1 - \bar{F}_w)\mathbb{J}_{\mathrm{F}_{A_w}}\rho\big(v(X), y_0\big) - \frac{1}{2}(\bar{F}_w - 1)^2\rho\big(y_0, y_0\big)\Big)\mathbb{1}_{A_w^c}(y).\end{aligned}$$

Accordingly, the corresponding score divergence is given by

$$\begin{aligned}\mathbb{D}_{\mathrm{twS}_\rho}(\mathrm{F}\|\mathrm{G}) &= \mathbb{E}_{\mathrm{F}}\mathrm{twS}_\rho(\mathrm{G}, Y; v, y_0) - \mathbb{E}_{\mathrm{F}}\mathrm{twS}_\rho(\mathrm{F}, Y; v, y_0) \\ &= \mathbb{J}_{\mathrm{F}_w,\mathrm{G}_w}\rho\big(v(X), v(Y)\big) - \frac{1}{2}\mathbb{J}_{\mathrm{F}_w}\rho\big(v(X), v(X')\big) - \frac{1}{2}\mathbb{J}_{\mathrm{G}_w}\rho\big(v(Y), v(Y')\big) \\ &\quad - (\bar{F}_w - \bar{G}_w)\mathbb{J}_{\mathrm{F}_w}\rho\big(y_0, v(Y)\big) + (\bar{F}_w - \bar{G}_w)\mathbb{J}_{\mathrm{G}_w}\rho\big(v(X), y_0\big) \\ &\quad - \frac{1}{2}(\bar{F}_w - \bar{G}_w)^2\rho\big(y_0, y_0\big) \\ &= \mathbb{J}_{\mathrm{F}_{A_w},\mathrm{G}_{A_w}}\rho\big(v(X), v(Y)\big) - \frac{1}{2}\mathbb{J}_{\mathrm{F}_{A_w}}\rho\big(v(X), v(X')\big) - \frac{1}{2}\mathbb{J}_{\mathrm{G}_{A_w}}\rho\big(v(Y), v(Y')\big) \\ &\quad - (\bar{F}_w - \bar{G}_w)\Big(\mathbb{J}_{\mathrm{F}_{A_w}}\rho\big(y_0, v(Y)\big) - \mathbb{J}_{\mathrm{G}_{A_w}}\rho\big(v(X), y_0\big)\Big) - \frac{1}{2}(\bar{F}_w - \bar{G}_w)^2\rho\big(y_0, y_0\big).\end{aligned}$$

If, additionally, $v(z) = z$ on A_w, the threshold-weighted kernel score framework coincides with generalized censoring based on $\mathrm{H} = \delta_{y_0}$.

We next consider the construction of the chaining function for the multivariate logistic weight function $w(\mathbf{y}) = \Lambda^d_{a,\mathrm{L}}(\mathbf{y}; \mathbf{r}) := \Lambda_{a,\mathrm{L}}(y_1; r_1) \times \cdots \times \Lambda_{a,\mathrm{L}}(y_d; r_d)$ where $\Lambda_{a,\mathrm{L}}(y_i; r_i) = \frac{1}{1+\exp(a(y_i - r_i))}$, $i = 1, \ldots, d$, $a > 0$. Since this is a product of marginal weight functions, it is natural to define its chaining function in direct analogy with the procedure suggested by Allen et al. (2023), as a vector of integrals of the marginal weight functions $\Lambda_{a,\mathrm{L}}(y_i; r_i)$ with respect to y_i. The marginal weight functions have respective (indefinite) integrals

$\int \Lambda_{a,\mathrm{L}}(y_i; r_i)\,\mathrm{d}y_i = y_i - \frac{1}{a}\log\left(1 + \exp(a(y_i - r_i))\right) + C$, leading to the chaining function

$$\mathbf{v}(\mathbf{y}) = \left(y_1 - \frac{1}{a}\log\left(1 + \mathrm{e}^{a(y_1 - r_1)}\right), \ldots, y_d - \frac{1}{a}\log\left(1 + \mathrm{e}^{a(y_d - r_d)}\right) \right).$$

F Additional Example

Example F.1. *Reconsider the setting of Example D.3, where the region of interest is* B*, with corresponding weight function* $w(y) = \mathbb{1}_B(y)$*, and* $\mathrm{P} \overset{B}{=} \mathrm{F}$*, while* P *and* G *are not coinciding on* B*. A well-known example of a localizing but improper scoring rule is the weighted likelihood score* $WLS(f, y) := \log f(y)\mathbb{1}_B(y)$ *proposed by Amisano and Giacomini (2007). In the current setting, this impropriety is revealed by the observation that* $\log g(y) > \log p(y), \forall y \in B$*, implies* $\mathbb{D}_{WLS}(\mathrm{P}\|\mathrm{P}) > \mathbb{D}_{WLS}(\mathrm{P}\|\mathrm{G})$*. That is, the fact that the likelihood of* G *restricted to* B *exceeds that of* P *is sufficient for WLS to favor* G*, the candidate distribution distinct from true distribution* P*. The conditional scoring rule proposed by Holzmann and Klar (2017a), also referred to as observation-weighted scoring rule in the context of kernel scores (Allen et al. 2023), is localizing and proper but not strictly locally proper. Specifically,* $S^{\sharp}_{w}(\mathrm{F}, y) = w(y)S(\mathrm{F}^{\sharp}_{w}, y)$*, where* $\mathrm{dF}^{\sharp}_{w} = \frac{1}{1-\bar{F}_w}\mathrm{dF}_w$*, assuming* $\bar{F}_w = \int_{\mathcal{Y}}(1-w)\mathrm{dF} < 1$*. This scoring rule is localizing and proper for weight functions for which it remains a scoring rule (see Section 2.1). However,* $S^{\sharp}_{B}(\mathrm{F}, y) = S^{\sharp}_{B}(\mathrm{G}, y) = S^{\sharp}_{B}(\mathrm{P}, y), \forall y \in B$*, since* $S^{\sharp}_{w}$ *cannot discriminate between distributions that are proportional to each other on* A_w*. Accordingly,* $\mathbb{D}_{S^{\sharp}_{B}}(\mathrm{P}\|\mathrm{F}) = \mathbb{D}_{S^{\sharp}_{B}}(\mathrm{P}\|\mathrm{G}) = 0$*, while only* F *coincides with* P *on* B*. In other words, the score divergence* $\mathbb{D}_{S^{\sharp}_{B}}$ *of a candidate distribution and* P *is zero if, but not only if, the candidate coincides with* P *on* B*, as is the case for* F*.*

G Monte Carlo Study

We use Monte Carlo simulations to evaluate the usefulness of our censoring approach for discriminating between competing density forecasts on a specific region of interest. We do so by analyzing the size and power properties of the Diebold and Mariano (2002) (DM) test based on different censored scoring rules.

As discussed in Section 4, the null hypothesis under consideration is given by

$$\mathbb{H}_0 : \mathbb{E}_{p_t} S_w(f_t, Y_{t+1}) = \mathbb{E}_{p_t} S_w(g_t, Y_{t+1}),$$

where f_t and g_t denote two (conditional) density forecasts of Y_{t+1}. The DM test statistic is computed as

$$t_T := \frac{\frac{1}{T}\sum_{t=0}^{T-1}\big(S_w(f_t, Y_{t+1}) - S_w(g_t, Y_{t+1})\big)}{\sqrt{\hat{\sigma}_T^2/T}},$$

where T is the number of observations used for evaluation and $\hat{\sigma}_T^2$ is the sample variance of the score difference $S_w(f_t, Y_{t+1}) - S_w(g_t, Y_{t+1})$. Note that in the empirical applications in Section 4 the DM test was used in the context of the Giacomini and White (2006) framework to accommodate parameter estimation uncertainty involved in producing the competing density forecasts. Here we abstain from this issue and assume that f and g are given and specified completely.

To put our findings for the censoring approach into perspective, we also evaluate DM test statistics based on focused scoring rules resulting from alternative localization procedures. In particular, we include conditional scoring rules and composite rules based on the proposal of Holzmann and Klar (2017a) to augment a conditional scoring rule with the auxiliary rule sbar or slog, as discussed in Section 3.5. For these composite scoring rules, we also assess to what extent their discriminative ability is attributable to the original and auxiliary scoring rules, by decomposing the associated standardized score divergences, as follows. We denote the score differences of the original conditional rule and of the auxiliary rule by $D_S^\sharp \equiv D_S^\sharp(Y;F,G) := S_w^\sharp(F,Y) - S_w^\sharp(G,Y)$ and $D_s \equiv D_s(Y;F,G) := s(b_{\bar{F}_w},Y) - s(b_{\bar{G}_w},Y)$, respectively. The score divergence for the composite rule can then be written as $\mathbb{D}_{S^\sharp+s}(\mathrm{F}\|\mathrm{G}) \equiv \mathbb{E}_F(D_S^\sharp + D_s) = \mathbb{E}_F(D_S^\sharp) + \mathbb{E}_F(D_s)$, with associated variance $\mathbb{V}_F(D_S^\sharp + D_s) = \mathbb{V}_F(D_S^\sharp) + \mathbb{V}_F(D_s) + 2\mathbb{C}\mathrm{ov}_F(D_S^\sharp, D_s)$. Define the standardized divergence as $\tilde{\mathbb{D}}_S(\mathrm{F}\|\mathrm{G}) := \mathbb{D}_S(\mathrm{F}\|\mathrm{G})/\sqrt{\mathbb{V}_F(D_S)}$. To evaluate the contribution of the auxiliary scoring rule s to $\tilde{\mathbb{D}}_{S^\sharp+s}$, we may compare $\tilde{\mathbb{D}}_{S^\sharp}(\mathrm{F}\|\mathrm{G}) = \mathbb{E}_F(D_{S^\sharp})/\sqrt{\mathbb{V}_F(D_{S^\sharp})}$ versus $\tilde{\mathbb{D}}_{S^\sharp+s}(\mathrm{F}\|\mathrm{G}) = \mathbb{E}_F(D_{S^\sharp} + D_s)/\sqrt{\mathbb{V}_F(D_{S^\sharp} + D_s)}$. In the power experiments below, we therefore consider their ratio $\xi_{S,s} := \tilde{\mathbb{D}}_{S^\sharp}(\mathrm{F}\|\mathrm{G})/\tilde{\mathbb{D}}_{S^\sharp+s}(\mathrm{F}\|\mathrm{G})$. Note that this ratio is nonnegative by construction, but not necessarily less than unity. Since $\mathbb{E}_F(D_S) \geq 0$ for any proper scoring rule S, it follows that $\mathbb{E}_F(D_{S^\sharp} + D_s) \geq \mathbb{E}_F(D_{S^\sharp})$. For the variance, however, it is possible that $\mathbb{V}_F(D_{S^\sharp}) \geq \mathbb{V}_F(D_{S^\sharp} + D_s)$, namely whenever $\mathbb{V}_F(s) \leq -2\mathbb{C}\mathrm{ov}_F(D_{S^\sharp}, D_s)$. Hence, it might be that the standardized score divergence $\tilde{\mathbb{D}}_{S^\sharp+s}(\mathrm{F}\|\mathrm{G})$ based on the composite scoring rule actually is smaller than the corresponding $\tilde{\mathbb{D}}_{S^\sharp}(\mathrm{F}\|\mathrm{G})$ based on the original conditional rule.

We employ a simulation design similar to Diks et al. (2011), Holzmann and Klar (2017b) and Lerch et al. (2017). For the size experiment, the null hypothesis of the DM test demands a particularly symmetric design, as explained by Diks et al. (2011). We thus compare two density forecasts f and g that are equally far from the true density p. For the power experiments, we take either f or g as the data generating process (DGP), and evaluate rejection frequencies against one-sided alternative hypotheses, in order to assess both 'true' and 'spurious' power of the test statistics. For the composite scoring rules, we thus also consider two sets of standardized score divergences $\tilde{\mathbb{D}}_{S^\sharp+s}$ and ratios $\xi_{S,s}$, namely with either

f or g being the DGP. In all experiments, we make use of indicator weight functions; either for the left tail $I_{\mathrm{L}}(y;r) = \mathbb{1}_{(-\infty,r)}(y)$ or for the center $I_{\mathrm{C}}(y;r) := I_{\mathrm{C}}(y;0,r) = \mathbb{1}_{(-r,r)}(y)$. We consider a range of values for the threshold r to vary the region of interest. The number of observations T varies, in such a way that the expected number of observations that falls within the region of interest is the same across the different values considered for r. All experiments are based on 10,000 replications.

G.1 Size

Following Diks et al. (2011), we assess the size properties of the DM test statistic using an i.i.d. standard normal density as DGP. The two candidate density forecasts f and g also are normal with unit variance, but with means $\mu_f = -0.2$ and $\mu_g = 0.2$. The evaluation is focused on the central region around the true mean 0 by using $I_{\mathrm{C}}(y;r)$ as weight function. Due to the symmetric design, the norms and $\bar{F}_w$-probabilities of the candidates f and g are equivalent, such that the test statistics based on QS and SphS scoring rules coincide. The equal norms and discrete probabilities also imply that the censoring and conditioning rules are equivalent within a semi-local scoring rule family, because in this case observations outside the region of interest obtain the same scores under both candidates. By the same token, the auxiliary scoring rules sbar and slog of the composite scoring rule of Holzmann and Klar (2017a) become equivalent to the conditional rules of both semi-local rules and the CRPS. Given these observations, the 16 weighted scoring rules arising from the combination of the unweighted scoring rules $\{\mathrm{LogS}, \mathrm{SphS}, \mathrm{QS}, \mathrm{CRPS}\}$ with the conditioning, censoring, conditioning appended with sbar and conditioning appended with slog localization techniques, can be represented by the censored LogS, SphS, CRPS and conditional CRPS. The twCRPS is added because it will be included as a benchmark in the power studies based on weight functions for which the censored CRPS variants do not reduce to the twCRPS.

Figure G.1: Size properties of the DM test

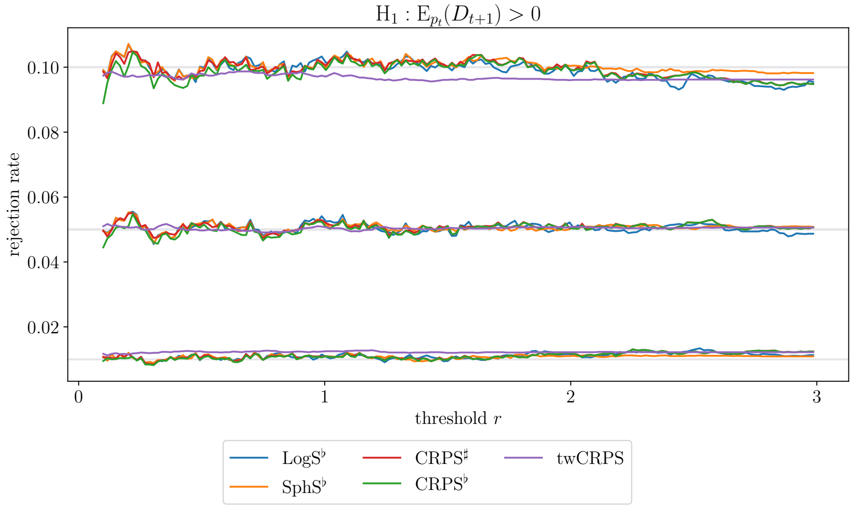


Figure G.1 displays the rejection rates of the null of equal predictive ability against the one-sided alternative that candidate f is statistically closer to p than g as a function of the threshold r for a fixed sample size $T = 500$. The rejection rates are given at nominal significance levels 0.01, 0.05 and 0.10, for focused versions of the LogS, SphS and CRPS scoring rules. In line with Diks et al. (2011), we find that none of the rejection rates displayed in Figure G.1 give reason to doubt that the tests are correctly sized.

G.2 Power

Normal versus Student-t: Left-tail. In our first power experiment, the competing density forecasts f and g are standard normal and Student-t_5, respectively. We focus the evaluation on the left tail by using $I_{\mathrm{L}}(y; r) = \mathbb{1}_{(-\infty,r)}(y)$ as weight function. The number of observations T is varied such that the expected number of observations below the threshold r is kept constant at $c = 20$. The combination of the selected candidates f and g and the left-tail region of interest make the current setting particularly interesting for financial risk management applications.

Figure G.2 shows the rejection rates of the DM test based on focused versions of the LogS, QS, SphS and CRPS rules in panels (a)-(d). In each panel, rejection rates are shown if the DGP is f (left-hand side) or g (right-hand side), and in favor of f (top) or g (bottom). Hence, the graphs in the top-left and bottom-right of each panel display true power, while the bottom-left and top-right plots show spurious power, i.e., the fraction of rejections of the null hypothesis of equal predictive ability in favor of the wrong density forecast. Concerning the selection of scoring rules, recall that the censored CRPS coincides with the twCRPS for the selected weight function. Finally, in the graphs in panels (b)-(d) we also include the LogS rule, motivated by the fact that the auxiliary rule in the Holzmann and Klar (2017a) approach is closely related to the censored log score, so that this can be considered as a relative benchmark comparison for their composite rules as well. Since the censored log score and the conditional log score with logs correction coincide, we omit the latter from panel a), in all figures below. Moreover, we highlight that sbar and slog overlap (visually) in cases that only one is visible, and, the same holds for the twCRPS and the censored CRPS.

Two main findings stand out from Figure G.2. First, the test statistics based on censored scoring rules do not suffer from spurious power, in the sense that rejection rates in the top-right and bottom-left graphs generally are close to zero. The same applies to tests based on the composite rules of Holzmann and Klar (2017a). The test based on the conditional scoring rule displays some spurious power, especially for threshold values around -1; however, as discussed in Diks et al. (2011), this might be a small sample issue, in the sense that the rejection rates decline towards zero for larger values of c quite rapidly.

Second, the tests based on the censored scoring rules and Holzmann and Klar (2017a) composite rules generally show comparable true power, but with some notable exceptions.

Specifically, in case the standard normal distribution is the DGP, we observe higher power for the censored rules for threshold values below -1.5 (except for a small region around $r = -2$) for the CRPS family, while the opposite is the case for the SphS scoring rules. In case the Student-t_5 distribution is the DGP, except for the LogS family, the composite scoring rules achieve higher rejection rates than the censored rules for threshold values r below -2.2, approximately. Interestingly, the higher power of the tests based on the composite scoring rules can be attributed to a large extent to the auxiliary scoring rule. This conclusion follows from observing that the tests based on the corresponding conditional scoring rules invariably have lower (and often substantially lower) power for regions quite far into the left-tail, i.e. for small values of r. This is corroborated by the (standardized) score divergences $\tilde{\mathbb{D}}_S(f_t\|g_t)$ shown in panels (a)-(d) of Figure G.3. These decline towards zero as r becomes smaller for the conditional rules but not for the composite rules. Hence, the ratios $\xi_{S,s}$ in panel (e) also decay when concentrating on the far left-tail as the region of interest; i.e. the auxiliary rule completely dominates the original conditional scoring rule in terms of contributing to the (standardized) score divergence $\tilde{\mathbb{D}}_{S^\sharp+s}(f_t\|g_t)$ of the composite scoring rule. Interestingly, the standardized divergence increases quite substantially for the censored scoring rule as r becomes smaller; hence; its discriminative ability improves when focusing more.

A final observation concerning Figure G.2 is that the test based on the censored logarithmic rule generally achieves the highest true power across all scoring rules considered; that is, higher than the censored rules based on QS, SphS and CRPS but also higher than the composite scoring rules.

Normal versus Student-t: Center. In our second power experiment, we again consider standard normal and Student-t_5 densities as the competing forecasts, but now focus on the central region by using $I_{\mathrm{C}}(y;r) = \mathbb{1}_{(-r,r)}(y)$ as weight function. We also increase T, so that the expected number of observations in the region of interest is equal to $c = 200$. Figure G.4 displays the rejection rates for the same selection of regular scoring rules and focusing procedures as in the first power experiment.

We make three observations, which are in line with the findings from the previous experiment. First, for a given scoring rule family, the DM test statistics based on censored and composite rules generally perform on par. The composite rules may achieve higher rejection rates, but this is rather sensitive to both the DGP and the auxiliary rule used. For example, panels (b) and (c) show that using slog as auxiliary rule improves performance for thresholds r exceeding 2, approximately, if the DGP is the standard normal distribution. By contrast, the discriminative ability of the test based on the same composite rules deteriorates for larger thresholds in case the DGP is Student-t_5. Also, using sbar as auxiliary rule generally leads to lower power, especially for relatively small values of the threshold r, i.e. when focusing on a fairly narrow region around the mean.

Second, the seemingly good performance of the composite scoring rules is again mostly due to the auxiliary scoring rule. This follows by noting that the rejection rates of the tests based on the conditional scoring rule alone rapidly decline towards zero when we lower the threshold value, especially when $r < 1.5$. This holds independent of the scoring rule family and the DGP. Confirmation is provided by Figure G.5. The standardized score divergences of the conditional scoring rules are quite comparable to those of the other localization procedures for $r > 1.5$. Focusing on narrower regions around the mean, however, both $\tilde{\mathbb{D}}_S(f_t \| g_t)$ and $\tilde{\mathbb{D}}_S(g_t \| f_t)$ display similar declines as the rejection rates in Figure G.4. Although the standardized score divergences also decay for the composite scoring rules, this kicks in only for substantially lower values of r. As a result, the ratios

$\xi_{S,s}$ in panel (e) accordingly slide down towards zero, indicating that the original conditional rule does not really contribute to the discriminative ability of the composite scoring rule.

Third, Figure G.4 shows that across scoring rule families and localization procedures, the censored likelihood score achieves highest power. In particular, while the composite scoring rule with slog as auxiliary rule may outperform its censored counterpart for a given scoring rule family, it is still dominated by the censored likelihood scoring rule.

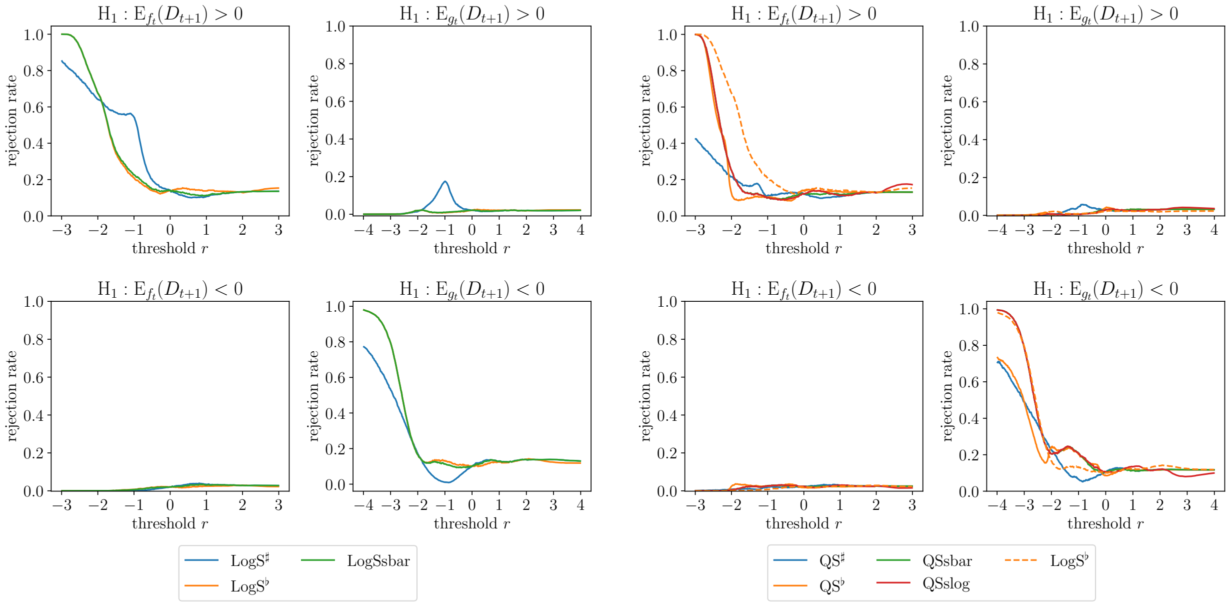
Finally, we note that the $\mathrm{CRPS}_w^\flat$ displayed in the Figure G.4 is the generalized censored scoring rule based on the generalized censored measure in Equation (4). Due to the symmetry of the setup, there is visually no difference between using the suggested value $\gamma = \frac{1}{2}$ (included in Figure G.4) or the fraction of observations smaller than $-r$. As the censored CRPS variants do not reduce to the twCRPS in this case, we include this separately in the graphs in panel (d) of Figure G.4. The rejection rates of the twCRPS in the top-left and bottom-right graphs are somewhat lower than those of the censored and composite rules; except for a modest range of threshold values around $r = 1$ if the DGP is standard normal.

Laplace tails In our third and final power experiment we study the consequences of the inability of the conditional scoring rule to distinguish two proportional tails when using the left-tail indicator $I_{\mathrm{L}}(y;r) = \mathbb{1}_{(-\infty,r)}(y)$ as weight function; see Example F.1. In particular, we consider two Laplace candidates with different location $\mu_f = -1$ and $\mu_g = 1$. Interestingly, if we were to set equivalent scale $\theta_f = \theta_g = 1$, even if $\mu_p \to \mu_f$ the conditional scoring rule does not have any power against the null of the candidates being statistically equally far away from p, that is, for thresholds $r < \mu_f$ for which the conditional distributions on $(-\infty, r)$ coincide. Since movements of p in terms of μ_p are invisible through the lens of a conditional score divergence, this is essentially not a lack of power against $\mathbb{H}_0$, which is based on the conditional scoring rule. Yet, it is a lack of power against the distributions being statistically equally far away from the actual density on A_w through the lens of the regular score divergence and, therefore, still a lack of local discriminative ability. More fundamentally, the test statistic degenerates in this case, as the score differences are exactly zero.

Because of these issues, we analyze what happens if the scale parameters are not exactly the same, but still fairly close. Specifically, we set $\theta_f = 1$ and $\theta_g = 1.1$. Figure G.6 shows the rejection rates of the DM test based on focused versions of the LogS, QS, SphS and CRPS rules.

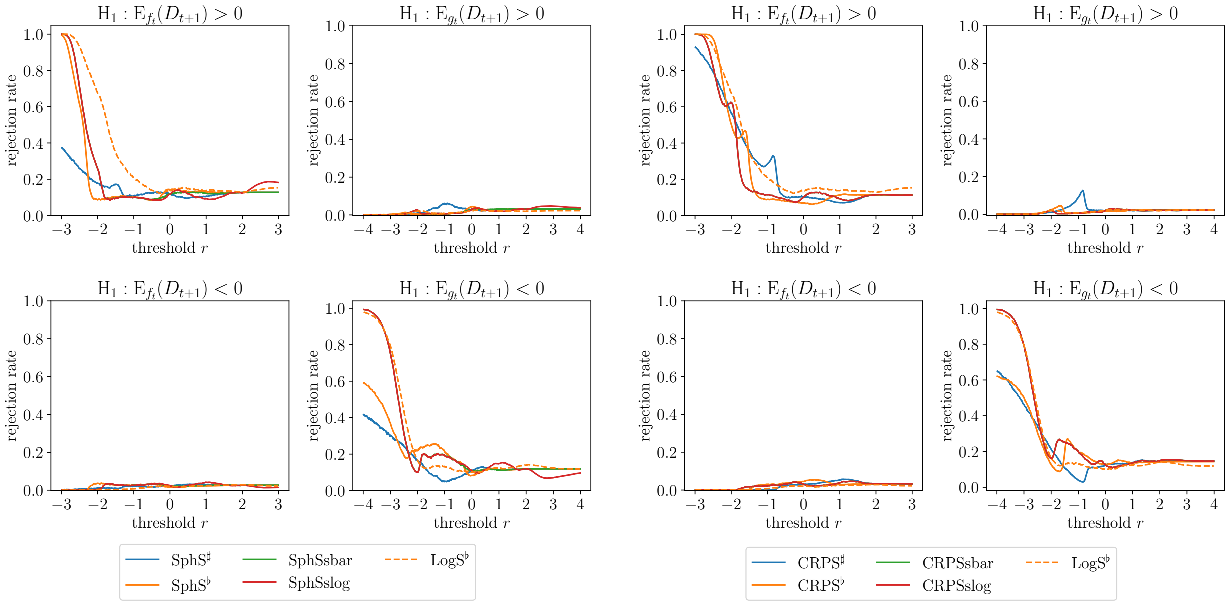
Four observations are clearly apparent. First, the increase in true power when moving from the conditional operator to the censoring operator is immense for all four scoring rules and thresholds $r < \mu_f$. The difference decreases over the interval $r \in (\mu_f, \mu_g)$, after which both conditioning and censoring have close to unit power. This observation is in line with the lack of discriminative ability of proportional and apparently close to proportional tails. Second, there is a clear difference in spurious power between the focusing operators: The censoring operator does not seem to suffer from spurious power at all, whereas the conditional rules have spurious power up to 0.10 for thresholds smaller than $\mu_f = -1$. Third, as in the previous experiments, we find that the censored likelihood score dominates the other scoring rules in terms of power, here in particular in case f is the DGP and $r < \mu_f$. Fourth, the Holzmann and Klar (2017a) augmentation of the conditional scoring rules renders great benefits also in this case, in the sense that it lifts power up to the level of the censored likelihood for all four scoring rules. Given that the power of the 'raw' conditional rules is rather poor, especially for thresholds $r < \mu_f$, this is likely due to the discriminative ability of the correction terms sbar or slog. This is confirmed by the standardized score divergences shown in Figure G.7.

Figure G.2: Rejection rates $\mathcal{N}(0,1)$ versus Student-t_5: Left-tail ($c = 20$)



(a) LogS

(b) QS

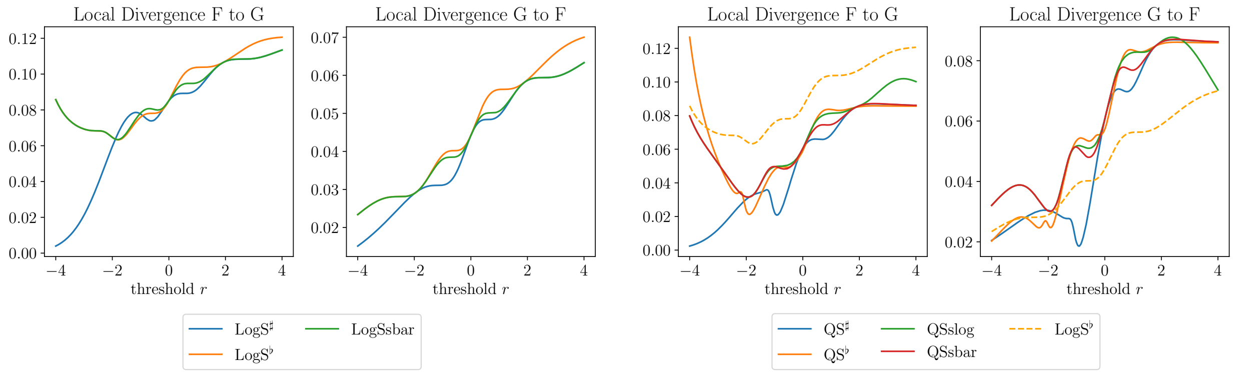


(c) SphS

(d) CRPS

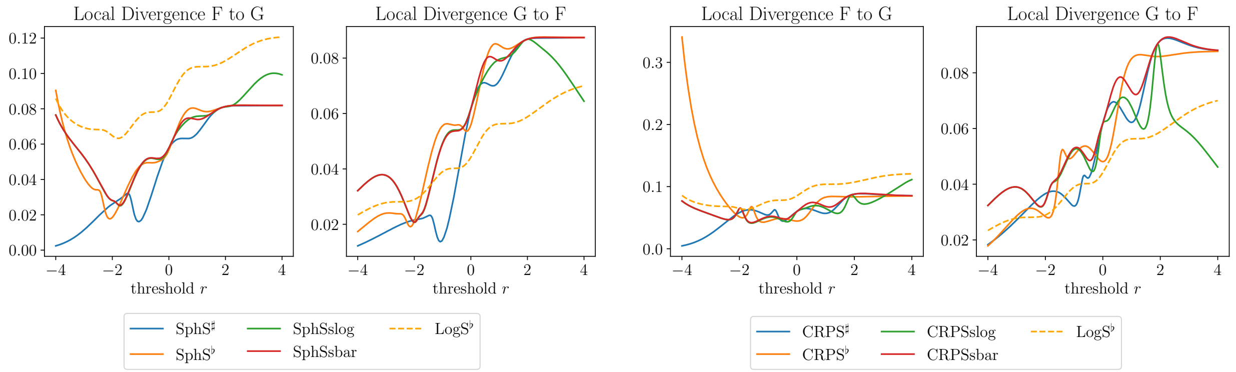
One-sided rejection rates of the DM test of equal predictive ability of the candidates f_t (standard normal) and g_t (Student-t_5) at a nominal significance level of 0.05 based on 10,000 simulations. The DGP is either f_t (left-hand side) or g_t (right-hand side). Moreover, rejections in the top panels are in favor of f_t, while rejections in the bottom panels are in favor of g_t. The incorporated weight function is $w(y) = 1_{(-\infty,r)}(y)$ and the expected number of observations in the region of interest is kept constant at $c = 20$.

Figure G.3: Standardized local divergences $\mathcal{N}(0,1)$ - Student-t_5



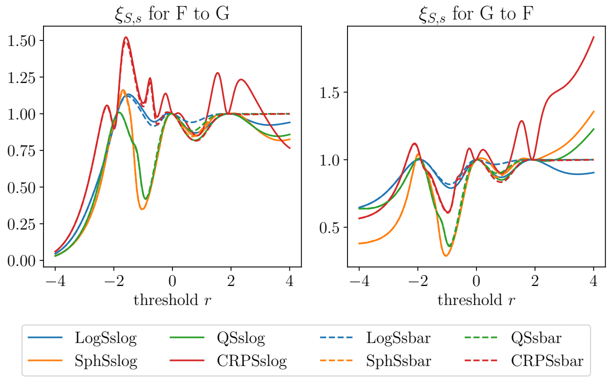
(a) LogS

(b) QS



(c) SphS

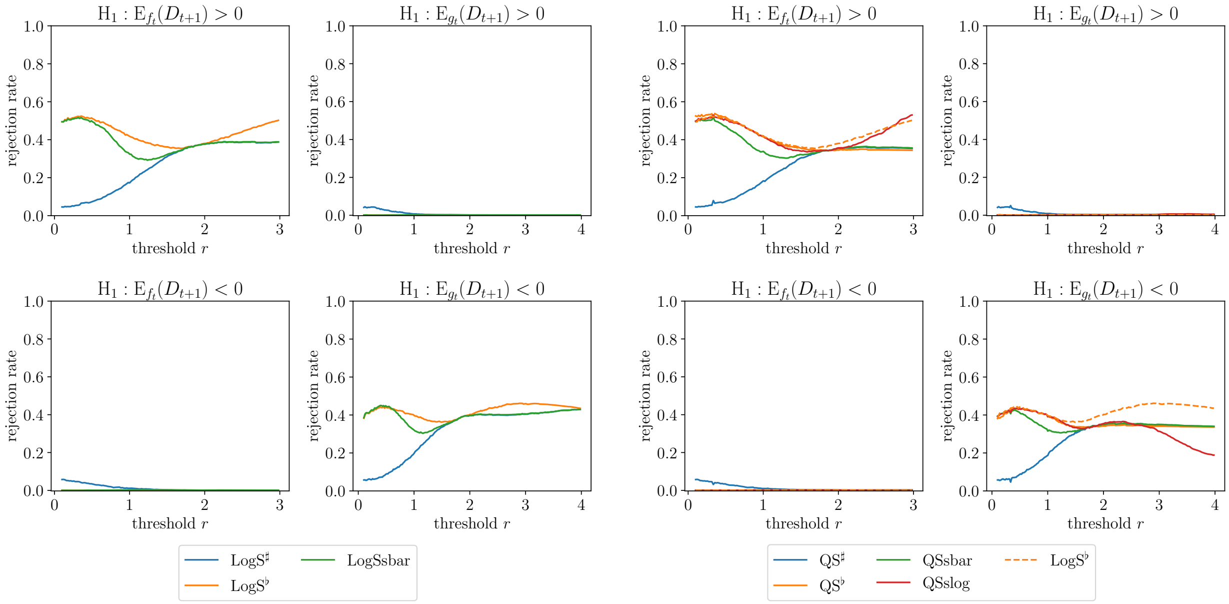
(d) CRPS



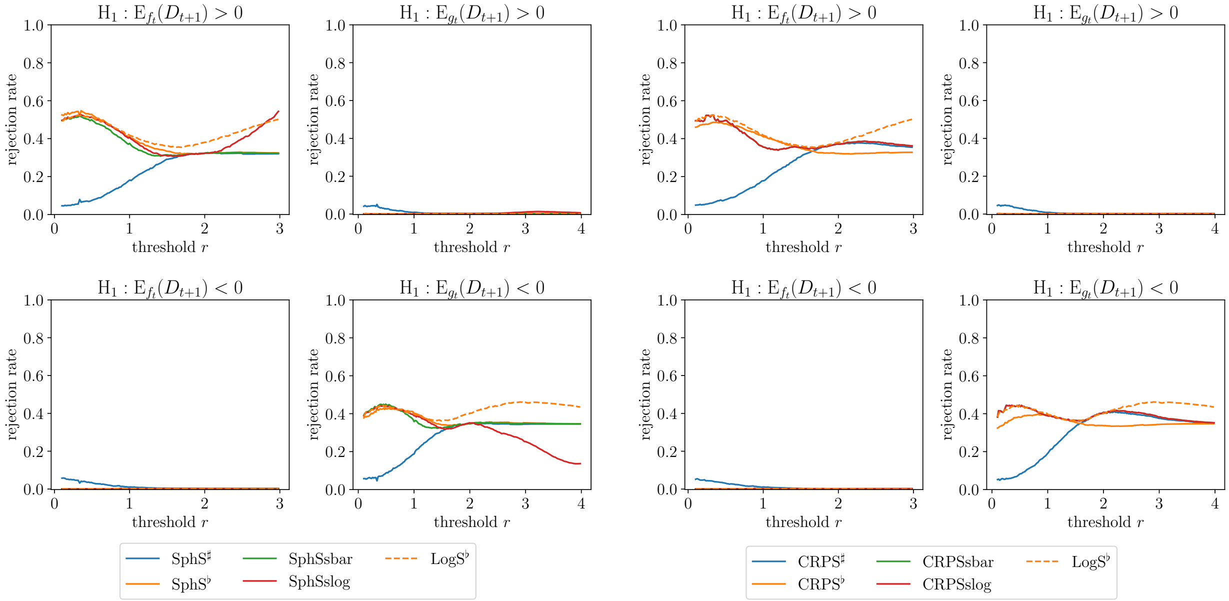
(e) $\xi_{S,s}$

Standardized divergences $(\tilde{\mathbb{D}}_S(f_t\|g_t))$ and ratios $(\xi_{S,s})$ of the candidates f_t $(\mathcal{N}(0,1))$ and g_t (Student-t_5), for weighted versions of unweighted scoring rules $S \in \{\text{LogS}, \text{QS}, \text{SphS}, \text{CRPS}\}$. Each plot includes the conditional and censored rules of S and two composite scoring rules $S+s$, where $s \in \{\text{sbar}, \text{slog}\}$. The incorporated weight function is $w(y) = 1_{(-\infty,r)}(y)$.

Figure G.4: $\mathcal{N}(0,1)$ versus Student-t_5: Center ($c = 200$)



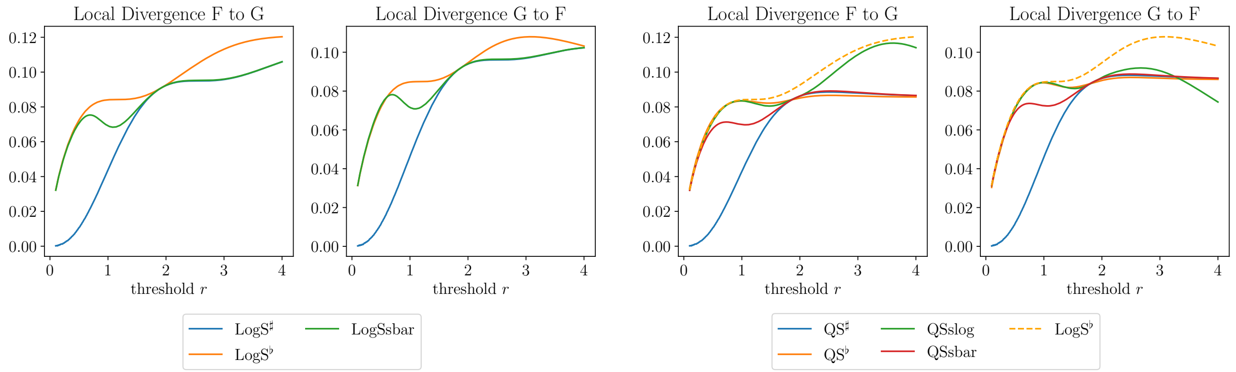
(a) LogS (b) QS



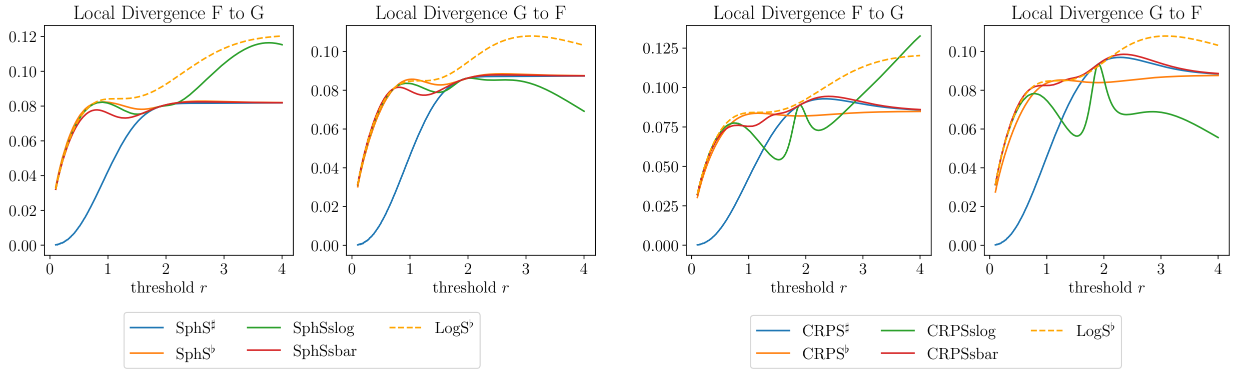
(c) SphS (d) CRPS

One-sided rejection rates of the DM test of equal predictive ability of the candidates f_t (standard normal) and g_t (Student-t_5) at a nominal significance level of 0.05 based on 10,000 simulations. The DGP is either f_t (left-hand side) or g_t (right-hand side). Moreover, rejections in the top panels are in favor of f_t, while rejections in the bottom panels are in favor of g_t. The incorporated weight function is $w(y) = 1_{(-r,r)}(y)$ and the expected number of observations in the region of interest is kept constant at $c = 200$.

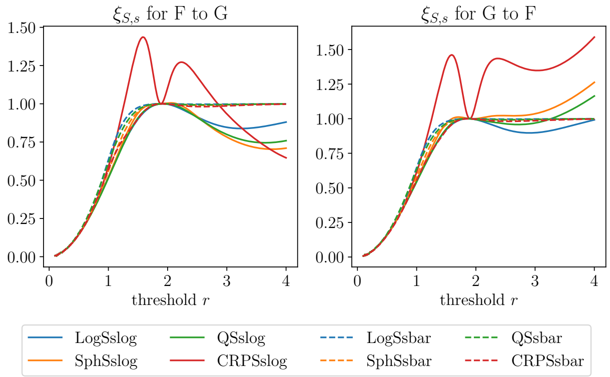
Figure G.5: Standardized local divergences $\mathcal{N}(0,1)$ - Student-t_5



(a) LogS (b) QS



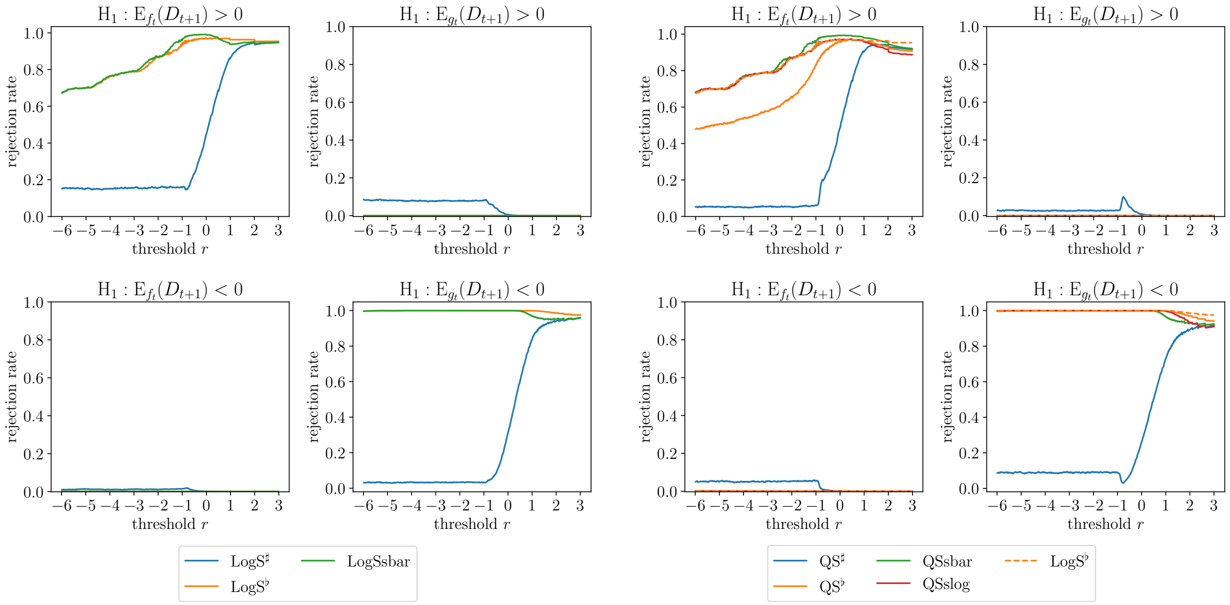
(c) SphS (d) CRPS



(e) $\xi_{S,s}$

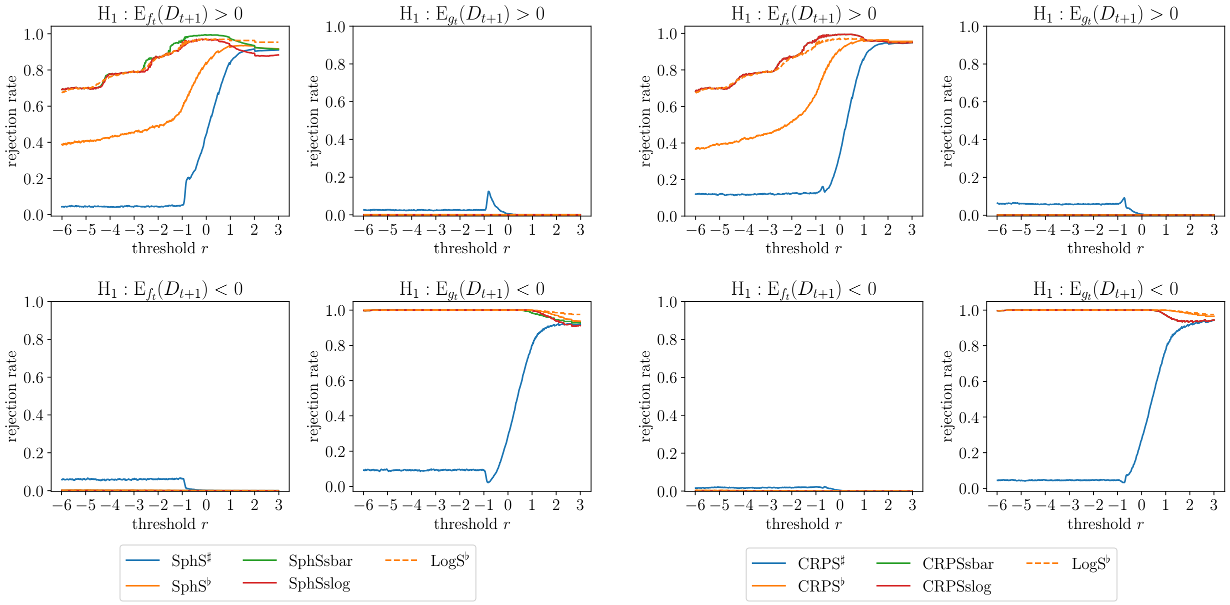
Standardized divergences ($\tilde{\mathbb{D}}_S(f_t\|g_t)$) and ratios ($\xi_{S,s}$) of the candidates f_t ($\mathcal{N}(0,1)$) and g_t (Student-t_5), for weighted versions of unweighted scoring rules $S \in \{\text{LogS}, \text{QS}, \text{SphS}, \text{CRPS}\}$. Each plot includes the conditional and censored rules of S and two composite scoring rules $S+s$, where $s \in \{\text{sbar}, \text{slog}\}$. The incorporated weight function is $w(y) = 1_{(-\infty,r)}(y)$.

Figure G.6: Rejection rates Laplace experiment ($c = 20$)



(a) LogS

(b) QS

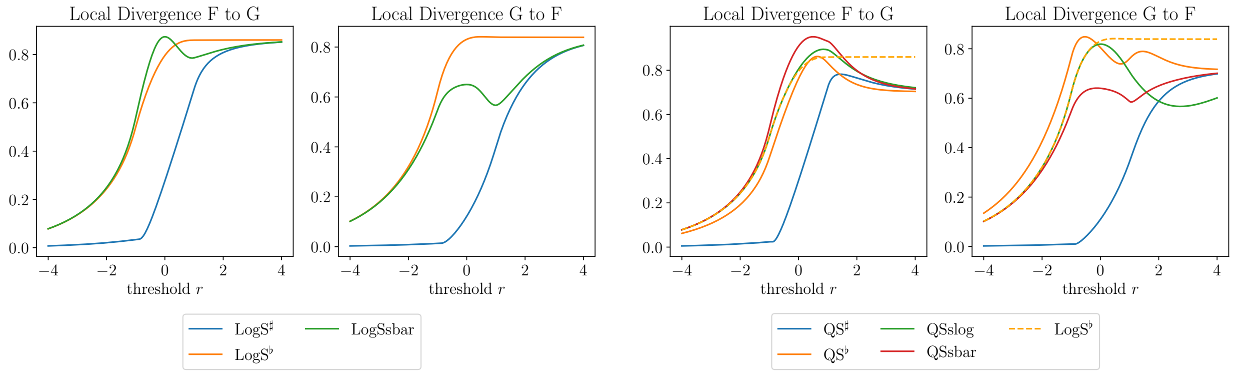


(c) SphS

(d) CRPS

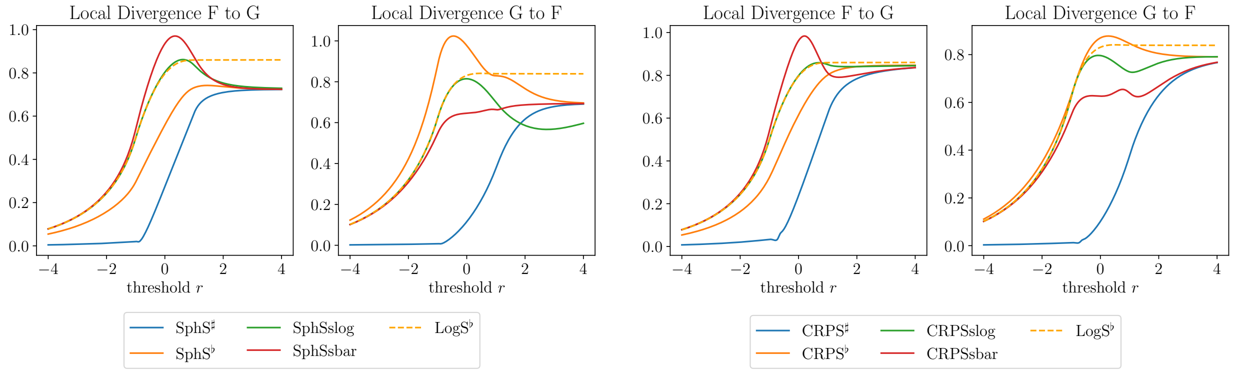
One-sided rejection rates of the DM test of equal predictive ability of the candidates f_t (Laplace$(-1, 1)$ and g_t (Laplace$(1, 1.1)$) at a nominal significance level of 0.05 based on 10,000 simulations. The DGP is either f_t (left-hand side) or g_t (right-hand side). Moreover, rejections in the top panels are in favor of f_t, while rejections in the bottom panels are in favor of g_t. The incorporated weight function is $w(y) = 1_{(-\infty,r)}(y)$ and the expected number of observations in the region of interest is kept constant at $c = 20$.

Figure G.7: Standardized divergences Laplace experiment ($c = 20$)



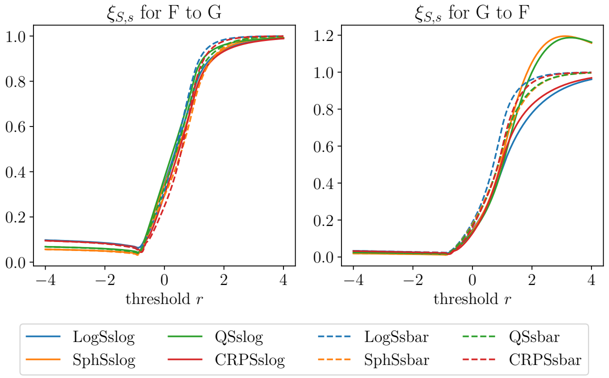
(a) LogS

(b) QS



(c) SphS

(d) CRPS



(e) $\xi_{S,s}$

Standardized divergences ($\tilde{\mathbb{D}}_S(f_t \| g_t)$) and ratios ($\xi_{S,s}$) of the candidates f_t (Laplace$(-1,1)$ and g_t (Laplace$(1,1.1)$), for weighted versions of unweighted scoring rules $S \in \{\text{LogS}, \text{QS}, \text{SphS}, \text{CRPS}\}$. Each plot includes the conditional and censored rules of S and two composite scoring rules $S+s$, where $s \in \{\text{sbar}, \text{slog}\}$. The incorporated weight function is $w(y) = 1_{(-\infty,r)}(y)$.

H Details on Model Specifications

H.1 Financial risk management

In the univariate application to daily returns on the S&P500 index in Section 4.1, we use forecast methods that conform to $Y_t|\mathcal{F}_{t-1} \sim \mathcal{D}(\mu, \sigma_t^2, \boldsymbol{\vartheta})$, denoting a parametric family of distributions with constant mean μ, time-varying variance σ_t^2, and any additional parameters collected in $\boldsymbol{\vartheta}$. We consider three specifications for the conditional variance σ_t^2. Specifically, we include (i) the GARCH(1,1) model proposed by Bollerslev (1986), with

$$\sigma_t^2 = \omega + \alpha(y_{t-1} - \mu)^2 + \beta\sigma_{t-1}^2,$$

(ii) the threshold GARCH(1,1) (TGARCH) model put forward by Glosten et al. (1993), with

$$\sigma_t^2 = \omega + \alpha(y_{t-1} - \mu)^2 + \gamma(y_{t-1} - \mu)^2 \mathbb{1}_{(-\infty,\mu]}(y_{t-1}) + \beta\sigma_{t-1}^2,$$

and (iii) the realized GARCH(1,1) (RGARCH) model of Hansen et al. (2012), with

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha x_{t-1} + \beta\sigma_{t-1}^2, \\ x_t &= \xi + \phi\sigma_t^2 + \tau z_t + \kappa(z_t^2 - 1) + u_t, \end{aligned}$$

where x_t represents a realized measure[1], $z_t = (y_t - \mu)/\sigma_t$, and u_t denotes a white noise process with variance σ_u^2. We combine each of the volatility specifications with standard normal and Student-t_ν distributions. Parameter estimates are obtained via maximum-likelihood estimation (MLE).

In the bivariate application we construct density forecasts for the vector of log-returns

---

[1] Obtained from `https://dachxiu.chicagobooth.edu/#risklab`

$\mathbf{y}_t = (y_{1,t}, y_{2,t}) \in \mathbb{R}^2$ for the Energy Select Sector SPDR Fund (XLE) and Financial Select Sector SPDR Fund (XLF). For the individual means μ_i $(i = 1, 2)$ and volatilities $\sigma^2_{i,t}$ $(i = 1, 2)$ we adopt the same specifications as used in the univariate application for the S&P500 returns. We model the dependence between the two sector returns by means of the Dynamic Conditional Correlation (DCC) methodology of Engle (2002), exploiting the decomposition of the conditional covariance matrix $\mathbf{\Sigma}_t$ as $\mathbf{\Sigma}_t = \mathbf{D}_t^{\frac{1}{2}}\mathbf{R}_t\mathbf{D}_t^{\frac{1}{2}}$, where $\mathbf{D}_t = \text{diag}(\sigma^2_{1,t}, \sigma^2_{2,t})$ contains the conditional variances and $\mathbf{R}_t$ is the conditional correlation matrix. In the DCC approach, the latter is computed as $\mathbf{R}_t = (\mathbf{Q}_t \odot \mathbf{I}_2)^{-\frac{1}{2}}\mathbf{Q}_t(\mathbf{Q}_t \odot \mathbf{I}_2)^{-\frac{1}{2}}$, with

$$\mathbf{Q}_t = (1 - \alpha_3 - \beta_3)\bar{\mathbf{Q}} + \alpha_3\mathbf{z}_{t-1}\mathbf{z}'_{t-1} + \beta_3\mathbf{Q}_{t-1},$$

where $z_{i,t} = (y_{i,t} - \mu_i)/\sigma_{i,t}$ $(i = 1, 2)$ and $\bar{\mathbf{Q}}$ denotes the unconditional covariance matrix of $\mathbf{z}_t$. For estimation we adopt the 2-step approach of Engle (2002), where we first estimate the parameters in the individual volatility specifications using MLE. In the second step we then estimate the parameters in the conditional correlation dynamics, again using MLE, where we use 'correlation targeting' by replacing $\bar{\mathbf{Q}}$ with the sample covariance matrix of the (first-stage) residuals $\hat{\mathbf{z}}_t$.

H.2 Macroeconomics

Following Stock and Watson (2002), among many others, we construct direct forecasts for annualized τ-month inflation rates $y^\tau_{t+\tau} = (1,200/\tau)\log\big(P_{t+\tau}/P_t\big)$, where P_t denotes the U.S. consumer price index (CPI) in month t, for horizons $\tau = 6$ and 24. Each of the forecast methods we consider can be represented as $y^\tau_{t+\tau} = \mu^\tau_{t+\tau}(\mathbf{x}_t) + u^\tau_{t+\tau}$, where $\mathbf{x}_t$ denotes a subset of $n = 122$ variables from the FRED-MD[2] database, cf. Medeiros et al. (2021). We use

---

[2] See `https://research.stlouisfed.org/econ/mccracken/fred-databases/`

six of the forecast methods listed by Medeiros et al. (2021) for the conditional mean $\mu^{\tau}_{t+\tau}$. First, we include a random walk forecast $\hat{\mu}^{\tau}_{t+\tau} = y^{\tau}_{t}$, equal to the most recent τ-month inflation. Second, we obtain a forecast from an autoregressive model (AR) as

$$\hat{\mu}^{\tau}_{t+\tau} = \hat{\phi}_{0,\tau} + \hat{\phi}_{1,\tau} y_t + \ldots + \hat{\phi}_{p,\tau} y_{t-p+1},$$

where $y_t \equiv y^{\tau}_{t}$ with $\tau = 1$ denotes the one-month inflation rate, the order p is selected using the Bayesian Information Criterion (BIC) and the parameters are estimated by OLS.

The third forecast method involves bagging, which combines forecasts obtained with different bootstrap subsamples. We adopt the implementation of Medeiros et al. (2021) for linear regression models, with the following steps:

1. For each bootstrap sample b, estimate a linear regression with all candidate predictors with OLS, and select those variables with an absolute t-statistic above a certain threshold c.
2. Estimate a linear regression only with the variables selected in the previous step, and use this second regression to obtain a forecast $\hat{\mu}^{\tau}_{t+\tau,b}$.
3. Repeat the first two steps for B bootstrap samples and average the B forecasts to obtain the bagging forecast: $\hat{\mu}^{\tau}_{t+\tau} = \frac{1}{B}\sum_{b=1}^{B} \hat{\mu}^{\tau}_{t+\tau,b}$.

We set the number of bootstrap samples $B = 100$ and set $c = 1.96$, such that the pre-test selection in step 1 is done at the 5% level.

As a fourth forecast method, we include Complete Subset Regression (CSR) developed by Elliott et al. (2013, 2015). The key idea of this approach is to average the forecasts obtained from linear regressions for all possible combinations of q predictors selected out of n available candidates, where typically q is set (much) smaller than n. With $n = 122$ as in our case, however, the number of regressions to be estimated is impractically large. We therefore adopt a pre-testing approach: First estimate a linear regression of $y^{\tau}_{t+\tau}$ on each of the candidate predictors individually, rank the t-statistics of their coefficients by absolute value, and select the $\tilde{n}$ highest-ranked variables. Then, apply the CSR approach to this subset of variables. We set $\tilde{n} = 20$ and $q = 4$.

The fifth forecast method again assumes a linear specification for the conditional mean, $\mu^{\tau}_{t+\tau}(\mathbf{x}_t) = \beta'_{\tau}\mathbf{x}_t$, with coefficients estimated using the Least Absolute Shrinkage and Selection Operator (LASSO), augmenting the sum of squared errors with the penalty term $\lambda \sum_{i=1}^{n} |\beta_{\tau,i}|$. We determine the regularization parameter λ by means of the BIC.

The sixth and final forecast is obtained from a Random Forest (RF) as proposed by Breiman (2001). Using B regression trees with K_b denoting the number of terminal nodes (or 'leaves') in the b-th tree (which is grown such that each terminal node has (at least) a pre-specified minimum number of observations), the RF forecast is given by

$$\mu^{\tau}_{t+\tau}(\mathbf{x}_t) = \frac{1}{B}\sum_{b=1}^{B}\sum_{k=1}^{K_b} \hat{c}_{k,b}\mathbb{1}_{k,b}(\mathbf{x}_t; \hat{\theta}_{k,b}),$$

where $\mathbb{1}_{k,b}(\mathbf{x}_t; \cdot)$ is an indicator function that is equal to 1 in case $\mathbf{x}_t$ is such that the observation ends up in terminal node k in tree b and 0 otherwise, and $\hat{c}_{k,b}$ is the average τ-month inflation rate for the observations in the training sample that correspond with this leaf. We use RFs with $B = 500$ trees. Each tree is grown until there are no less than five observations in each leaf. The fraction of randomly selected predictors at each split is set to 1/3.

Finally, for each of the six forecast methods considered for $\mu^{\tau}_{t+\tau}$, we obtain a density forecast for $y^{\tau}_{t+\tau} = \mu^{\tau}_{t+\tau} + u^{\tau}_{t+\tau}$ by assuming that $u^{\tau}_{t+\tau}$ follows a two-piece normal distribution, with density given by

$$f(y;\mu,\sigma_1,\sigma_2) = \frac{2}{\sigma_1+\sigma_2}\left(\phi\left(\frac{y-\mu}{\sigma_1}\right)\mathbb{1}_{y<\mu} + \phi\left(\frac{y-\mu}{\sigma_2}\right)\mathbb{1}_{y\geq\mu}\right), \qquad \sigma_1,\sigma_2>0,$$

where $\phi(z)$ denotes the density of the standard normal distribution. This distributional choice is congruent with the underlying statistical model employed in the fan charts published by the Monetary Policy Committee of the Bank of England (Clements 2004; Mitchell and Hall 2005; Gneiting and Ranjan 2011).

H.3 Climate

The forecast methods we use for predicting daily temperatures closely follow the GARCH, QGARCH-I, and QGARCH-II specifications as in Franses et al. (2001), but with alterations in seasonal trend estimation. Specifically, we use local day averages for the mean and a sine function for volatility, as opposed to a quadratic function. The models can be formalized as: $Y_t|\mathcal{F}_{t-1} \sim \mathcal{D}(\mu_t, \sigma_t^2, \boldsymbol{\vartheta})$, where $\mu_t = m_{t|t-1} + \phi y_{t-1}$ and

$$\sigma_t^2 = \varphi(t;\omega_0,\omega_1) + \alpha\big(y_{t-1} - \mu_{t-1} - \varphi(t;\gamma_0,\gamma_1)\big)^2 + \beta\sigma_{t-1}^2.$$

Here, $m_{t|t-1}$ is the average temperature of days with the same day number in the estimation window, that is, all $s \in [t-m, t-1]$ such that $\tilde{T}_s = \tilde{T}_t$, where $\tilde{T}_t = \min(T_t, 365)$, in which T_t is the day number, with $T_t = 1$ on the first of February. The latter choice exploits the periodic pattern revealed by Figure 1 in Franses et al. (2001), which we model by $\varphi(t;\theta_0,\theta_1) = \theta_0 + \theta_1|\sin(\pi/365 \cdot \tilde{T}_t)|$. These specifications are combined with both Normal and Student-t_ν distributions to produce six forecast methods in total.

I Additional Tables

I.1 MCS table at confidence level 0.75

Table I.1: MCS cardinality of censored and (un)corrected conditional scoring rules

			no correction			sbar			slog		
Sec.	w_t	τ	$\leq$	$<$	$\sharp/\flat$	$\leq$	$<$	$\sharp/\flat$	$\leq$	$<$	$\sharp/\flat$
4.1	$I_{\mathrm{L}}(y_t;\hat{r}_t^q)$	1	92%	62%	2.04	75%	38%	1.28	75%	29%	1.16
		5	50%	46%	1.50	67%	29%	1.16	67%	29%	1.07
	$I_{\mathrm{L}}^2(\mathbf{y}_t;\tilde{\mathbf{r}}_t^q)$	1	62%	38%	1.28	62%	17%	1.07	62%	21%	1.08
		5	92%	33%	1.39	92%	33%	1.39	96%	33%	1.45
	$\Lambda_{3,\mathrm{L}}^2(\mathbf{y}_t;\tilde{\mathbf{r}}_t^q)$	1	62%	42%	1.13	67%	25%	0.99	67%	25%	0.99
		5	79%	42%	1.28	92%	17%	1.15	92%	17%	1.15
4.2	$I_{\mathrm{C}}(y_t;2,r_1)$	6	100%	100%	3.72	100%	92%	2.35	92%	33%	1.31
		24	100%	100%	3.31	75%	33%	1.17	75%	42%	1.29
	$I_{\mathrm{C}}^c(y_t;2,r_1)$	6	100%	83%	2.84	83%	25%	1.36	75%	33%	1.26
		24	92%	58%	2.71	50%	25%	1.25	75%	25%	1.09
4.3	$I_{\mathrm{R}}(y_t;\hat{r}_t^q)$	1	71%	54%	1.82	83%	54%	1.50	75%	21%	1.06
		3	79%	38%	1.33	83%	38%	1.29	79%	4%	0.94
	$I_{\mathrm{C}}(y_t;18,r_2)$	1	100%	58%	2.04	100%	42%	1.42	92%	25%	1.21
		3	100%	25%	1.25	100%	0%	1.00	100%	0%	1.00
Total average			84%	56%	1.97	81%	33%	1.31	80%	24%	1.15

NOTE: This table presents changes in cardinality of the MCS in absolute and relative terms, at confidence level 0.75, across different forecast horizons τ, corresponding to the forecasting applications in risk management (Section 4.1), inflation (Section 4.2) and temperature (Section 4.3). sbar and slog refer to the correction terms for conditional scoring rules proposed by Holzmann and Klar (2017a). Columns labeled $\leq$ ($<$) display the percentage of cases where $\mathrm{MCS}^\flat$ contains (strictly) fewer forecast methods than $\mathrm{MCS}^\sharp$ and the column labeled $\sharp/\flat$ reports the ratio $|\mathrm{MCS}^\sharp|/|\mathrm{MCS}^\flat|$. Each result represents an average over a set of scoring rules $S \in \{\mathrm{LogS}, \mathrm{QS}, \mathrm{SphS}, \mathrm{CRPS}/S_{\rho_1}\}$ and quantile levels $q \in \{0.01, 0.05, 0.10, 0.15, 0.20, 0.25\}$ or levels $r_1 \in \{1, 1.5, 2\}$ and $r_2 \in \{1, 2, 4\}$. The empirical q-th quantiles $\hat{r}_t^q$ of y_t are based on the forecast method estimation window, and $\tilde{\mathbf{r}}_t^q := (\hat{r}_{1,t}^{q_2}, \hat{r}_{2,t}^{q_2})$, with $q_2 = \sqrt{q}$, approximates a bivariate empirical q-th quantile of $\mathbf{y}_t$. The p-values are obtained via a block bootstrap of $B = 10{,}000$ replications, with block length $b = 5$, or $b = 200$ for the climate data.

I.2 Financial risk management

Table I.2 shows the complete set of MCS p-values for the individual forecast methods in the univariate risk management application in Section 4.1, for the four focused variants of the LogS, QS, SphS and CRPS scoring rules, for horizons $\tau = 1$ and 5, and quantiles $q = 0.01$, 0.05, 0.1, 0.15,. 0.2 and 0.25.

At a 0.90 confidence level and $\tau = 1$, $\text{MCS}^\sharp$ is smaller only in a single case across all examined quantiles and scoring rules, namely for $q = 0.25$ and $S = \text{QS}$. Equality in MCS size occurs mainly for larger quantiles, where information scarcity with respect to the distributions on $(-\infty, \hat{r}_t^q)$ is less critical. Examining the composition of the MCSs reveals that, in case $|\text{MCS}^\flat| \leq |\text{MCS}^\sharp|$, the censored MCSs are often a subset of the conditional MCSs. The significance of reductions in the number of methods retained in the MCS due to censoring is further emphasized by the fact that the resulting MCSs encompass more complex model specifications, which would be the optimal choices in the absence of parameter estimation uncertainty.

Robustness checks, pertaining to b and m, confirm the stability with respect to these parameters, see Table I.3. There we also report results based on the alternative statistic proposed by Hansen et al. (2011), namely $\text{Tmax} := \max_{i \in \mathcal{M}_k} |t_{m,n}^{(i,\cdot)}|$, where $t_{m,n}^{(i,\cdot)}$ corresponds to the $t_{m,n}$-statistic based on the average loss between forecast method i and all other methods in the MCS. Comparing results across Tmax and TR statistics, we observe that the use of TR tends to expedite model elimination, yielding smaller MCS p-values compared to Tmax; this acceleration, however, is consistent across both censoring and conditioning.

Table I.4 displays MCS p-values and model confidence sets obtained with the indicator product weight function, while Table I.5 reports the corresponding MCS p-values for the logistic product weight function. The columns labeled 'tw' in the latter table contain the results for the twS_{ρ_1} of Allen et al. (2023), for the chaining function derived in Section E.4. Overall, it can be observed that the MCSs obtained with the twS_{ρ_1} resemble those of $\text{S}_{\rho_1}^\flat$. Note that this may be related to the fact that the twS_{ρ_1} tends to the limiting $\text{S}_{\rho_1}^\flat$ score for large values of the logistic transition speed parameter a, which has a fairly large value of 3 in the applications here.

Table I.2: MCS p-values for univariate risk management application.

			LogS				QS			
q	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog
0.01	1	RGARCH-t	1.00	0.63	1.00	1.00	0.46	0.95	1.00	1.00
		TGARCH-t	0.98	1.00	0.98	0.96	0.64	1.00	0.61	0.57
		GARCH-t	0.52	0.70	0.52	0.52	0.34	0.83	0.37	0.33
		RGARCH-$\mathcal{N}$	***0.09***	**0.20**	***0.09***	***0.09***	1.00	0.95	***0.07***	***0.05***
		TGARCH-$\mathcal{N}$	***0.03***	***0.10***	***0.03***	***0.03***	0.64	0.95	***0.01***	***0.01***
		GARCH-$\mathcal{N}$	***0.01***	***0.08***	***0.01***	***0.01***	0.46	0.83	***0.00***	***0.00***
	5	RGARCH-t	0.37	0.87	0.37	0.40	**0.12**	1.00	0.43	0.46
		TGARCH-t	0.82	1.00	0.82	0.78	0.85	0.44	0.83	0.78
		GARCH-t	1.00	0.97	1.00	1.00	**0.17**	0.37	1.00	1.00
		RGARCH-$\mathcal{N}$	***0.01***	***0.05***	***0.01***	***0.01***	**0.12**	0.82	***0.01***	***0.01***
		TGARCH-$\mathcal{N}$	***0.01***	***0.05***	***0.01***	***0.01***	1.00	0.74	***0.01***	***0.01***
		GARCH-$\mathcal{N}$	***0.01***	***0.04***	***0.01***	***0.01***	**0.17**	0.74	***0.00***	***0.00***
0.05	1	RGARCH-t	1.00	0.78	1.00	1.00	***0.04***	1.00	0.27	0.36
		TGARCH-t	***0.09***	1.00	***0.09***	***0.10***	***0.02***	0.82	***0.00***	***0.00***
		GARCH-t	***0.01***	0.78	***0.01***	***0.01***	***0.00***	0.82	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	***0.08***	***0.06***	***0.08***	***0.07***	1.00	1.00	1.00	1.00
		TGARCH-$\mathcal{N}$	***0.00***	***0.03***	***0.00***	***0.00***	***0.04***	0.82	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	0.82	***0.00***	***0.00***
	5	RGARCH-t	0.76	**0.24**	0.76	0.79	0.29	**0.25**	0.26	0.28
		TGARCH-t	0.99	0.75	0.99	0.92	0.98	1.00	0.34	0.33
		GARCH-t	1.00	1.00	1.00	1.00	0.30	0.36	0.34	0.33
		RGARCH-$\mathcal{N}$	***0.01***	***0.01***	***0.01***	***0.01***	0.98	**0.12**	1.00	1.00
		TGARCH-$\mathcal{N}$	***0.01***	***0.01***	***0.01***	***0.01***	1.00	0.36	***0.01***	***0.01***
		GARCH-$\mathcal{N}$	***0.01***	***0.01***	***0.01***	***0.01***	0.30	**0.25**	***0.02***	***0.02***
0.1	1	RGARCH-t	1.00	0.74	1.00	1.00	1.00	0.70	***0.04***	***0.06***
		TGARCH-t	***0.04***	1.00	***0.04***	***0.05***	**0.17**	1.00	***0.00***	***0.00***
		GARCH-t	***0.00***	0.40	***0.00***	***0.00***	***0.01***	0.42	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	***0.04***	***0.03***	***0.04***	***0.04***	0.48	0.70	1.00	1.00
		TGARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.06***	0.39	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.06***	***0.00***	***0.00***
	5	RGARCH-t	0.45	**0.15**	0.45	0.45	0.34	**0.14**	0.91	0.94
		TGARCH-t	1.00	0.38	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	0.55	1.00	0.55	0.58	0.34	0.66	0.74	0.71
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	**0.14**	0.96	0.97
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	**0.14**	0.96	0.97
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	**0.14**	0.74	0.71

NOTE: This table continues on the next page.

Table I.2 (continued): MCS p-values for univariate risk management application.

			LogS				QS			
q	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog
0.15	1	RGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	0.31	0.42
		TGARCH-t	***0.03***	0.78	***0.03***	***0.03***	***0.07***	0.63	***0.00***	***0.00***
		GARCH-t	***0.00***	**0.19**	***0.00***	***0.00***	***0.00***	**0.19**	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	***0.03***	***0.02***	***0.03***	***0.02***	***0.07***	0.63	1.00	1.00
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	0.37	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.04***	***0.00***	***0.00***
	5	RGARCH-t	0.39	***0.09***	0.39	0.28	**0.12**	0.48	0.61	0.63
		TGARCH-t	1.00	**0.23**	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	0.76	1.00	0.76	0.81	0.80	0.85	0.61	0.60
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	0.48	0.39	0.30
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	0.51	0.61	0.63
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	0.48	0.43	0.43
0.2	1	RGARCH-t	1.00	1.00	1.00	1.00	1.00	0.40	1.00	1.00
		TGARCH-t	***0.02***	**0.22**	***0.02***	***0.03***	***0.09***	***0.06***	***0.01***	***0.01***
		GARCH-t	***0.00***	***0.06***	***0.00***	***0.00***	***0.06***	***0.00***	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	***0.01***	***0.04***	***0.01***	***0.01***	***0.00***	1.00	0.40	0.47
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	**0.15**	0.35	**0.15**	**0.23**	***0.00***	0.89	0.27	0.28
		TGARCH-t	1.00	0.38	1.00	1.00	0.76	1.00	1.00	1.00
		GARCH-t	0.77	1.00	0.77	0.88	1.00	0.89	0.27	0.28
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	0.89	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	0.80	***0.06***	***0.09***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	0.80	***0.06***	***0.06***
0.25	1	RGARCH-t	1.00	1.00	1.00	1.00	1.00	***0.04***	1.00	1.00
		TGARCH-t	***0.03***	0.27	***0.03***	***0.05***	0.74	***0.08***	***0.01***	***0.02***
		GARCH-t	***0.00***	***0.04***	***0.00***	***0.00***	0.74	***0.00***	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	***0.01***	***0.04***	***0.01***	***0.01***	***0.00***	1.00	***0.02***	***0.06***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.08***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	**0.12**	0.30	**0.12**	**0.14**	***0.00***	0.77	***0.02***	***0.05***
		TGARCH-t	1.00	0.50	1.00	0.97	**0.15**	1.00	1.00	1.00
		GARCH-t	0.98	1.00	0.98	1.00	1.00	0.77	**0.15**	**0.10**
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	0.77	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	0.77	***0.00***	***0.01***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	0.64	***0.00***	***0.01***

NOTE: This table continues on the next page.

Table I.2 (continued): MCS p-values for univariate risk management application.

			SphS				CRPS			
q	τ	Method	♭	♯	sbar	slog	♭	♯	sbar	slog
0.01	1	RGARCH-t	0.64	0.88	1.00	1.00	0.72	0.97	1.00	1.00
		TGARCH-t	0.88	1.00	0.67	0.63	0.81	1.00	0.69	0.65
		GARCH-t	0.64	0.88	0.47	0.45	0.81	0.91	0.50	0.47
		RGARCH-$\mathcal{N}$	1.00	0.67	***0.05***	***0.04***	0.81	0.97	**0.14**	**0.12**
		TGARCH-$\mathcal{N}$	0.88	0.38	***0.01***	***0.01***	1.00	0.97	***0.01***	***0.01***
		GARCH-$\mathcal{N}$	0.64	0.45	***0.00***	***0.00***	0.81	0.63	***0.00***	***0.00***
	5	RGARCH-t	**0.23**	1.00	0.61	0.65	0.56	1.00	0.54	0.58
		TGARCH-t	1.00	0.40	0.75	0.73	0.66	0.64	1.00	1.00
		GARCH-t	0.49	0.40	1.00	1.00	0.66	0.46	0.83	0.84
		RGARCH-$\mathcal{N}$	**0.17**	0.40	***0.02***	***0.02***	0.56	**0.17**	***0.01***	***0.01***
		TGARCH-$\mathcal{N}$	1.00	**0.24**	***0.01***	***0.01***	1.00	**0.16**	***0.01***	***0.01***
		GARCH-$\mathcal{N}$	0.41	0.28	***0.01***	***0.01***	0.66	***0.08***	***0.00***	***0.00***
0.05	1	RGARCH-t	***0.01***	1.00	1.00	1.00	0.40	0.80	0.32	0.41
		TGARCH-t	***0.01***	0.73	***0.01***	***0.01***	0.40	1.00	***0.01***	***0.02***
		GARCH-t	***0.00***	0.73	***0.00***	***0.00***	**0.13**	0.79	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	1.00	0.58	0.60	0.48	1.00	0.80	1.00	1.00
		TGARCH-$\mathcal{N}$	***0.01***	**0.25**	***0.00***	***0.00***	0.48	0.64	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	**0.25**	***0.00***	***0.00***	0.40	0.38	***0.00***	***0.00***
	5	RGARCH-t	**0.24**	0.30	0.92	0.95	0.46	0.56	0.78	0.91
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	0.80	0.91
		GARCH-t	0.40	0.55	0.92	0.95	0.57	0.56	0.80	0.91
		RGARCH-$\mathcal{N}$	0.93	***0.02***	0.83	0.78	0.46	***0.03***	1.00	1.00
		TGARCH-$\mathcal{N}$	0.80	***0.02***	***0.00***	***0.00***	0.77	***0.05***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	**0.24**	***0.01***	***0.00***	***0.00***	0.47	***0.05***	***0.01***	***0.01***
0.1	1	RGARCH-t	0.34	0.89	1.00	1.00	**0.11**	0.74	**0.13**	**0.10**
		TGARCH-t	***0.01***	1.00	***0.02***	***0.02***	***0.04***	1.00	***0.00***	***0.00***
		GARCH-t	***0.00***	0.37	***0.00***	***0.00***	***0.01***	**0.19**	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	1.00	**0.21**	1.00	0.82	1.00	0.74	1.00	1.00
		TGARCH-$\mathcal{N}$	***0.00***	***0.03***	***0.00***	***0.00***	**0.11**	***0.04***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.03***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	0.55	**0.14**	0.73	0.67	0.49	0.43	0.79	0.73
		TGARCH-t	1.00	0.53	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	0.55	1.00	0.73	0.67	0.63	0.43	0.47	0.42
		RGARCH-$\mathcal{N}$	0.27	***0.03***	0.73	0.67	0.52	***0.00***	0.79	0.73
		TGARCH-$\mathcal{N}$	**0.21**	***0.00***	***0.01***	***0.01***	0.63	***0.00***	0.47	0.41
		GARCH-$\mathcal{N}$	**0.19**	***0.00***	***0.02***	***0.02***	0.49	***0.00***	**0.12**	***0.09***

NOTE: This table continues on the next page.

Table I.2 (continued): MCS p-values for univariate risk management application.

			SphS				CRPS			
q	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog
0.15	1	RGARCH-t	1.00	1.00	1.00	1.00	***0.08***	0.84	**0.23**	0.29
		TGARCH-t	***0.01***	**0.19**	***0.00***	***0.00***	***0.01***	0.88	***0.00***	***0.00***
		GARCH-t	***0.00***	**0.10**	***0.00***	***0.00***	***0.00***	**0.11**	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	0.50	**0.10**	0.64	0.42	1.00	1.00	1.00	1.00
		TGARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.03***	**0.11**	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	0.33	0.49	0.71	0.70	0.52	0.43	0.33	0.37
		TGARCH-t	1.00	0.59	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	0.42	1.00	0.71	0.70	0.57	0.65	0.33	0.36
		RGARCH-$\mathcal{N}$	***0.00***	0.49	0.40	0.28	0.52	**0.24**	**0.22**	**0.15**
		TGARCH-$\mathcal{N}$	***0.02***	***0.06***	**0.17**	**0.13**	0.57	***0.00***	**0.22**	**0.19**
		GARCH-$\mathcal{N}$	***0.02***	***0.01***	**0.13**	***0.07***	0.50	***0.00***	**0.11**	***0.09***
0.2	1	RGARCH-t	1.00	0.26	1.00	1.00	***0.10***	0.41	1.00	0.98
		TGARCH-t	***0.01***	***0.02***	***0.00***	***0.00***	***0.01***	**0.13**	***0.09***	***0.07***
		GARCH-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	***0.01***	1.00	0.71	0.83	1.00	1.00	0.88	1.00
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.09***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.00***	***0.00***
	5	RGARCH-t	***0.03***	0.97	0.60	0.68	0.38	0.69	**0.21**	**0.21**
		TGARCH-t	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	0.51	0.97	0.60	0.68	0.38	0.69	**0.21**	**0.21**
		RGARCH-$\mathcal{N}$	***0.00***	1.00	***0.07***	***0.09***	0.32	0.69	***0.01***	***0.01***
		TGARCH-$\mathcal{N}$	***0.00***	0.91	**0.19**	0.28	0.38	***0.02***	***0.01***	***0.02***
		GARCH-$\mathcal{N}$	***0.00***	0.76	**0.20**	0.28	0.25	***0.00***	***0.01***	***0.01***
0.25	1	RGARCH-t	1.00	***0.02***	1.00	1.00	**0.17**	**0.11**	1.00	1.00
		TGARCH-t	***0.03***	***0.02***	***0.01***	***0.02***	***0.01***	***0.10***	**0.19**	0.31
		GARCH-t	***0.01***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.01***
		RGARCH-$\mathcal{N}$	***0.00***	1.00	0.43	0.82	1.00	1.00	**0.19**	0.34
		TGARCH-$\mathcal{N}$	***0.00***	***0.02***	***0.00***	***0.00***	***0.01***	**0.11**	***0.01***	***0.01***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.00***	***0.00***
	5	RGARCH-t	***0.00***	0.96	0.41	0.54	0.27	0.50	***0.03***	***0.05***
		TGARCH-t	0.83	0.96	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	1.00	0.93	0.41	0.54	0.27	0.52	**0.25**	**0.20**
		RGARCH-$\mathcal{N}$	***0.00***	0.96	***0.02***	***0.08***	**0.12**	0.50	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	1.00	**0.13**	0.31	**0.14**	0.44	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	0.72	**0.14**	0.30	**0.12**	***0.08***	***0.00***	***0.00***

NOTE: This table presents the MCS p-values implied by censored ($\flat$) and (augmented) conditional ($\sharp$) scoring rules underlying the results presented in Table 2, for different forecast horizons τ. Calculations are conducted by the `R` package `MCS` by Bernardi and Catania (2018), using $B = 10,000$ simulations and block length $b = 5$. Bold (and italic) p-values signify a forecast method's elimination from $\text{MCS}_{0.75}$ (and $\text{MCS}_{0.90}$).

Table I.3: Robustness analysis MCS cardinality

		no correction			sbar			slog		
m	Statistic	$\leq$	$<$	$\sharp/\flat$	$\leq$	$<$	$\sharp/\flat$	$\leq$	$<$	$\sharp/\flat$
					$\tau = 1$					
1000	TR_{20}	96	75	2.38	71	38	1.20	71	38	1.20
	Tmax_{20}	92	50	1.61	54	29	1.10	54	17	0.95
750	TR_5	88	62	2.04	62	29	1.17	62	29	1.15
	Tmax_5	71	33	1.15	42	8	0.82	42	4	0.80
1250	TR_5	92	58	2.20	62	29	1.09	62	25	1.05
	Tmax_5	83	58	1.74	58	21	1.01	58	17	0.97
					$\tau = 5$					
1000	TR_{20}	62	29	1.29	54	25	1.08	62	25	1.10
	Tmax_{20}	79	38	1.46	62	29	1.28	67	25	1.22
750	TR_5	50	29	1.23	58	29	1.09	58	29	1.10
	Tmax_5	71	25	1.33	62	25	1.24	67	25	1.19
1250	TR_5	67	33	1.47	58	29	1.17	54	29	1.12
	Tmax_5	71	42	1.53	67	29	1.38	67	29	1.25

NOTE: The table presents changes in cardinality of the MCS in absolute and relative terms, at confidence level 0.90, across different forecast horizons $\tau = 1$ and 5, for the univariate forecasting application in risk management (Section 4.1); where we vary the length of the estimation window to $m = 750$ and $1,250$, or the block length to $b = 20$ for the value of $m = 1,000$ as reported in Table 2.

Table I.4: MCS p-values for indicator product risk management application.

			LogS				QS			
q	τ	Method	♭	♯	sbar	slog	♭	♯	sbar	slog
0.01	1	RGARCH-t	1.00	1.00	1.00	1.00	***0.00***	**0.16**	***0.07***	***0.09***
		TGARCH-t	**0.22**	0.95	**0.22**	**0.22**	***0.02***	0.26	***0.09***	***0.09***
		GARCH-t	**0.13**	0.95	**0.13**	**0.13**	***0.03***	0.27	***0.05***	***0.05***
		RGARCH-$\mathcal{N}$	**0.22**	0.38	**0.22**	**0.22**	1.00	0.27	1.00	1.00
		TGARCH-$\mathcal{N}$	***0.06***	**0.16**	***0.06***	***0.06***	0.52	1.00	**0.12**	**0.19**
		GARCH-$\mathcal{N}$	***0.02***	**0.12**	***0.02***	***0.02***	0.52	0.27	***0.06***	***0.05***
	5	RGARCH-t	0.43	0.43	0.43	0.45	**0.16**	0.61	0.41	0.47
		TGARCH-t	1.00	1.00	1.00	1.00	***0.02***	0.43	0.92	0.97
		GARCH-t	**0.22**	**0.25**	**0.22**	**0.22**	***0.01***	0.61	0.77	0.77
		RGARCH-$\mathcal{N}$	***0.07***	***0.08***	***0.07***	***0.07***	0.80	0.61	0.92	0.98
		TGARCH-$\mathcal{N}$	***0.07***	***0.08***	***0.07***	***0.07***	0.80	1.00	1.00	1.00
		GARCH-$\mathcal{N}$	***0.06***	***0.08***	***0.06***	***0.06***	1.00	0.61	0.61	0.61
0.05	1	RGARCH-t	1.00	1.00	1.00	1.00	0.87	0.86	1.00	1.00
		TGARCH-t	**0.14**	0.53	**0.14**	**0.15**	1.00	1.00	**0.17**	0.32
		GARCH-t	***0.05***	0.53	***0.05***	***0.05***	***0.08***	0.86	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	**0.14**	**0.18**	**0.14**	**0.13**	0.67	***0.00***	0.62	0.41
		TGARCH-$\mathcal{N}$	***0.01***	***0.01***	***0.01***	***0.01***	0.66	***0.00***	**0.17**	**0.24**
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.06***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	**0.21**	0.39	**0.21**	**0.22**	***0.06***	***0.02***	0.26	**0.17**
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	***0.10***	0.36	***0.10***	**0.10**	***0.06***	0.65	**0.18**	**0.17**
		RGARCH-$\mathcal{N}$	***0.01***	***0.01***	***0.01***	***0.01***	***0.00***	***0.00***	***0.06***	***0.02***
		TGARCH-$\mathcal{N}$	***0.01***	***0.01***	***0.01***	***0.01***	***0.06***	***0.00***	0.56	0.31
		GARCH-$\mathcal{N}$	***0.01***	***0.01***	***0.01***	***0.01***	***0.02***	***0.00***	0.26	**0.17**
0.1	1	RGARCH-t	1.00	1.00	1.00	1.00	0.80	***0.07***	1.00	1.00
		TGARCH-t	***0.09***	0.33	***0.09***	**0.18**	1.00	1.00	0.54	0.69
		GARCH-t	***0.04***	**0.14**	***0.04***	***0.04***	***0.10***	***0.05***	***0.01***	***0.00***
		RGARCH-$\mathcal{N}$	***0.09***	**0.14**	***0.09***	***0.09***	***0.00***	***0.00***	***0.01***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.01***	***0.00***	***0.00***	***0.01***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	**0.13**	**0.25**	**0.13**	**0.18**	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	**0.13**	**0.25**	**0.13**	**0.18**	***0.04***	***0.04***	***0.02***	***0.01***
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***

NOTE: This table continues on the next page.

Table I.4 (continued): MCS p-values for indicator product risk management application.

			LogS				QS			
q	τ	Method	♭	♯	sbar	slog	♭	♯	sbar	slog
0.15	1	RGARCH-t	1.00	1.00	1.00	1.00	0.39	0.60	1.00	1.00
		TGARCH-t	0.29	0.39	0.29	0.31	1.00	1.00	0.66	0.81
		GARCH-t	***0.06***	**0.12**	***0.06***	***0.07***	**0.16**	**0.15**	***0.02***	***0.02***
		RGARCH-$\mathcal{N}$	***0.07***	**0.12**	***0.07***	***0.07***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	***0.07***	**0.11**	***0.07***	***0.07***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	***0.07***	**0.11**	***0.07***	***0.07***	***0.05***	***0.06***	***0.04***	***0.03***
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
0.2	1	RGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		TGARCH-t	0.34	0.35	0.34	0.35	0.61	0.69	0.68	0.68
		GARCH-t	***0.06***	***0.07***	***0.06***	***0.06***	**0.21**	0.25	0.32	**0.25**
		RGARCH-$\mathcal{N}$	***0.06***	***0.07***	***0.06***	***0.06***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	***0.03***	***0.05***	***0.03***	***0.03***	***0.02***	***0.00***	***0.00***	***0.00***
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	**0.12**	**0.13**	**0.12**	**0.11**	0.26	0.47	0.65	0.59
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
0.25	1	RGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		TGARCH-t	0.37	0.35	0.37	0.36	0.33	0.49	0.72	0.63
		GARCH-t	***0.03***	***0.04***	***0.03***	***0.03***	***0.08***	0.31	0.37	0.26
		RGARCH-$\mathcal{N}$	***0.03***	***0.04***	***0.03***	***0.03***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	**0.13**	**0.12**	**0.13**	**0.13**	***0.05***	0.59	0.60	0.58
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***

NOTE: This table continues on the next page.

Table I.4 (continued): MCS p-values for indicator product risk management application.

			SphS				S_{ρ_1}			
q	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog
0.01	1	RGARCH-t	***0.01***	**0.12**	***0.09***	**0.11**	0.38	0.30	0.29	0.29
		TGARCH-t	***0.03***	**0.15**	***0.10***	**0.11**	0.38	0.30	0.29	0.29
		GARCH-t	***0.03***	**0.15**	***0.05***	***0.05***	0.38	0.30	**0.23**	**0.24**
		RGARCH-$\mathcal{N}$	1.00	**0.15**	0.45	0.44	0.38	0.30	0.53	0.53
		TGARCH-$\mathcal{N}$	0.72	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-$\mathcal{N}$	0.72	**0.15**	***0.10***	**0.11**	0.38	0.30	**0.20**	**0.20**
	5	RGARCH-t	***0.04***	0.35	0.29	0.30	**0.22**	0.41	0.64	0.63
		TGARCH-t	***0.03***	0.35	0.29	0.31	0.37	0.89	0.93	0.94
		GARCH-t	***0.03***	0.35	0.29	0.30	0.46	0.89	0.93	0.94
		RGARCH-$\mathcal{N}$	**0.21**	0.44	0.61	0.65	**0.22**	0.64	0.81	0.82
		TGARCH-$\mathcal{N}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-$\mathcal{N}$	0.68	0.35	0.29	0.30	0.55	0.89	0.81	0.82
0.05	1	RGARCH-t	***0.05***	0.89	0.78	1.00	0.36	0.71	1.00	1.00
		TGARCH-t	***0.05***	1.00	0.28	0.29	***0.07***	1.00	0.98	0.98
		GARCH-t	***0.00***	***0.09***	***0.00***	***0.00***	***0.07***	***0.08***	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	1.00	***0.00***	1.00	0.88	0.67	***0.01***	**0.19**	**0.13**
		TGARCH-$\mathcal{N}$	**0.25**	***0.00***	0.28	0.29	1.00	***0.08***	0.98	0.98
		GARCH-$\mathcal{N}$	***0.01***	***0.00***	***0.00***	***0.00***	***0.07***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	***0.00***	***0.01***	**0.13**	***0.08***	**0.18**	0.48	0.36	0.37
		TGARCH-t	**0.11**	1.00	1.00	1.00	**0.18**	1.00	1.00	1.00
		GARCH-t	***0.02***	**0.12**	***0.06***	***0.05***	**0.18**	0.48	**0.25**	0.27
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.05***	***0.05***	**0.18**	***0.04***	**0.17**	**0.12**
		TGARCH-$\mathcal{N}$	1.00	***0.00***	0.55	0.31	1.00	***0.04***	0.58	0.43
		GARCH-$\mathcal{N}$	**0.18**	***0.00***	**0.13**	***0.08***	**0.18**	***0.00***	***0.00***	***0.00***
0.1	1	RGARCH-t	1.00	0.32	1.00	1.00	**0.18**	0.31	0.34	0.33
		TGARCH-t	***0.09***	1.00	0.61	0.74	**0.21**	0.83	0.67	0.72
		GARCH-t	***0.02***	***0.01***	***0.00***	***0.00***	***0.08***	***0.01***	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	***0.05***	***0.00***	***0.00***	***0.00***	0.80	0.59	0.67	0.72
		TGARCH-$\mathcal{N}$	***0.02***	***0.00***	***0.00***	***0.00***	1.00	1.00	1.00	1.00
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	**0.18**	***0.01***	***0.01***	***0.01***
	5	RGARCH-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.10***	**0.16**	***0.10***	**0.10**
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	***0.03***	***0.00***	***0.00***	***0.00***	0.40	**0.18**	***0.10***	**0.10**
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.10***	**0.18**	***0.10***	**0.10**
		TGARCH-$\mathcal{N}$	***0.03***	***0.00***	***0.00***	***0.00***	0.26	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.01***	***0.00***	***0.00***	***0.00***	0.40	***0.00***	***0.00***	***0.00***

NOTE: This table continues on the next page.

Table I.4 (continued): MCS p-values for indicator product risk management application.

			SphS				S_{ρ_1}			
q	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog
0.15	1	RGARCH-t	1.00	1.00	1.00	1.00	**0.18**	***0.02***	***0.03***	***0.03***
		TGARCH-t	0.63	0.77	0.36	0.45	0.88	1.00	1.00	1.00
		GARCH-t	***0.03***	***0.05***	***0.00***	***0.01***	***0.02***	***0.02***	***0.01***	***0.01***
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	0.83	1.00	0.95
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.43	***0.02***	***0.03***	***0.02***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	**0.19**	***0.02***	***0.02***	***0.02***
	5	RGARCH-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.05***	***0.06***	***0.06***	***0.06***
		TGARCH-t	1.00	1.00	1.00	1.00	0.33	0.47	0.49	0.49
		GARCH-t	***0.06***	***0.01***	***0.01***	***0.01***	0.37	0.62	0.50	0.52
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.02***	***0.03***	***0.03***	***0.02***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	1.00	1.00	1.00
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.05***	***0.00***	***0.00***	***0.00***
0.2	1	RGARCH-t	1.00	1.00	1.00	1.00	**0.14**	***0.01***	***0.02***	***0.02***
		TGARCH-t	0.36	**0.19**	**0.19**	**0.18**	1.00	0.99	1.00	1.00
		GARCH-t	0.27	***0.05***	***0.06***	***0.05***	0.93	***0.09***	**0.11**	***0.09***
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	1.00	1.00	1.00
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	0.99	1.00	1.00
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.74	***0.02***	***0.03***	***0.02***
	5	RGARCH-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.04***	***0.02***	***0.03***
		TGARCH-t	1.00	1.00	1.00	1.00	0.93	1.00	1.00	1.00
		GARCH-t	0.41	**0.13**	0.27	**0.20**	0.93	0.40	0.46	0.44
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.02***	***0.09***	***0.05***	***0.06***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.93	0.43	0.46	0.44
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	0.36	0.44	0.42
0.25	1	RGARCH-t	1.00	1.00	1.00	1.00	0.95	0.99	0.97	0.98
		TGARCH-t	0.29	**0.19**	**0.24**	**0.21**	0.91	0.80	0.83	0.82
		GARCH-t	**0.13**	***0.04***	***0.09***	***0.03***	0.35	***0.01***	***0.01***	***0.01***
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	**0.11**	***0.05***	***0.06***	***0.06***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	1.00	1.00	1.00
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.31	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.08***	0.30	0.32	0.31
		TGARCH-t	1.00	1.00	1.00	1.00	**0.25**	0.57	0.57	0.57
		GARCH-t	**0.14**	**0.12**	**0.18**	**0.14**	0.84	0.67	0.65	0.66
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.04***	***0.02***	***0.02***	***0.02***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	1.00	1.00	1.00
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.08***	***0.00***	***0.00***	***0.00***

NOTE: This table presents the MCS p-values implied by censored ($\flat$) and (corrected) conditional ($\sharp$) scoring rules underlying the results presented in Table 2, for different forecast horizons τ. Calculations are conducted by the `R` package `MCS` by Bernardi and Catania (2018), using $B = 10,000$ simulations and block length $b = 5$. Bold (and italic) p-values signify a forecast method's elimination from $\mathrm{MCS}_{0.75}$ (and $\mathrm{MCS}_{0.90}$).

Table I.5: MCS p-values for logistic product risk management application.

			LogS				QS			
q	τ	Method	♭	♯	sbar	slog	♭	♯	sbar	slog
0.01	1	RGARCH-t	1.00	1.00	1.00	1.00	0.94	0.99	***0.06***	***0.06***
		TGARCH-t	0.28	0.70	0.28	0.29	0.97	1.00	***0.06***	***0.06***
		GARCH-t	**0.13**	0.68	**0.13**	**0.14**	0.65	0.99	***0.01***	***0.01***
		RGARCH-$\mathcal{N}$	**0.16**	**0.24**	**0.16**	**0.16**	1.00	0.99	1.00	1.00
		TGARCH-$\mathcal{N}$	***0.04***	***0.04***	***0.04***	***0.04***	0.97	0.99	***0.06***	***0.06***
		GARCH-$\mathcal{N}$	***0.01***	***0.02***	***0.01***	***0.01***	0.48	0.64	***0.01***	***0.01***
	5	RGARCH-t	**0.20**	0.40	**0.20**	**0.21**	0.41	1.00	0.45	0.50
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	0.93	0.52	0.52
		GARCH-t	***0.07***	**0.11**	***0.07***	***0.07***	0.41	0.91	0.45	0.48
		RGARCH-$\mathcal{N}$	***0.04***	***0.05***	***0.04***	***0.04***	0.26	0.93	0.52	0.52
		TGARCH-$\mathcal{N}$	***0.04***	***0.05***	***0.04***	***0.04***	0.74	0.60	1.00	1.00
		GARCH-$\mathcal{N}$	***0.04***	***0.05***	***0.04***	***0.04***	0.32	0.41	0.52	0.52
0.05	1	RGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		TGARCH-t	**0.18**	0.31	**0.18**	**0.19**	0.84	0.71	**0.18**	**0.18**
		GARCH-t	***0.05***	**0.18**	***0.05***	***0.06***	0.33	0.39	***0.02***	***0.01***
		RGARCH-$\mathcal{N}$	**0.11**	**0.18**	**0.11**	**0.11**	***0.04***	0.88	0.49	0.44
		TGARCH-$\mathcal{N}$	***0.01***	***0.01***	***0.01***	***0.01***	***0.01***	0.63	**0.15**	**0.14**
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	**0.14**	***0.01***	***0.01***
	5	RGARCH-t	**0.16**	0.50	**0.16**	**0.16**	***0.01***	0.78	***0.01***	***0.05***
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	**0.16**	**0.25**	**0.16**	**0.16**	0.63	0.78	0.37	0.29
		RGARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.05***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	**0.13**	0.37	0.29
		GARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	**0.13**	0.33	0.29
0.1	1	RGARCH-t	1.00	1.00	1.00	1.00	0.82	1.00	1.00	1.00
		TGARCH-t	**0.23**	0.26	**0.23**	**0.23**	1.00	0.71	0.47	0.50
		GARCH-t	***0.07***	***0.10***	***0.07***	***0.08***	0.69	***0.06***	***0.06***	***0.06***
		RGARCH-$\mathcal{N}$	***0.09***	***0.10***	***0.09***	***0.09***	***0.01***	0.45	***0.06***	***0.06***
		TGARCH-$\mathcal{N}$	***0.01***	***0.01***	***0.01***	***0.01***	***0.00***	**0.17**	***0.06***	***0.06***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	**0.11**	0.27	**0.11**	**0.13**	***0.00***	***0.03***	***0.00***	***0.00***
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	**0.13**	0.27	**0.13**	**0.13**	0.90	***0.08***	0.41	0.36
		RGARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.01***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***

NOTE: This table continues on the next page.

Table I.5 (continued): MCS p-values for logistic product risk management application.

			LogS				QS			
q	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog
0.15	1	RGARCH-t	1.00	1.00	1.00	1.00	0.91	1.00	1.00	1.00
		TGARCH-t	0.25	0.25	0.25	0.25	1.00	0.74	0.75	0.76
		GARCH-t	***0.08***	***0.07***	***0.08***	***0.08***	0.78	***0.03***	**0.19**	**0.11**
		RGARCH-$\mathcal{N}$	***0.08***	***0.08***	***0.08***	***0.08***	***0.00***	***0.03***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	***0.02***	**0.18**	***0.02***	***0.09***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	**0.13**	**0.18**	**0.13**	**0.12**	0.82	***0.03***	0.62	0.48
		RGARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
0.2	1	RGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		TGARCH-t	0.27	**0.25**	0.27	0.26	0.87	0.68	0.91	0.87
		GARCH-t	***0.08***	***0.06***	***0.08***	***0.07***	0.70	***0.01***	0.47	0.26
		RGARCH-$\mathcal{N}$	***0.08***	***0.07***	***0.08***	***0.07***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	***0.03***	***0.05***	***0.03***	***0.04***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	**0.16**	**0.14**	**0.16**	**0.15**	0.63	***0.03***	0.82	0.59
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
0.25	1	RGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		TGARCH-t	0.27	0.26	0.27	0.27	0.62	0.63	0.95	0.87
		GARCH-t	***0.07***	***0.05***	***0.07***	***0.07***	0.57	***0.02***	0.72	0.42
		RGARCH-$\mathcal{N}$	***0.07***	***0.06***	***0.07***	***0.07***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	***0.01***	***0.01***	***0.01***	***0.01***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	**0.18**	**0.12**	**0.18**	**0.16**	0.44	***0.07***	0.90	0.64
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***

NOTE: This table continues on the next page.

Table I.5 (continued): MCS p-values for logistic product risk management application.

			SphS				S_{ρ_1}				
q	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog	tw
0.01	1	RGARCH-t	***0.03***	***0.00***	***0.00***	***0.00***	**0.22**	0.51	0.83	0.77	***0.02***
		TGARCH-t	***0.03***	***0.00***	***0.00***	***0.00***	**0.22**	0.51	**0.21**	**0.24**	***0.01***
		GARCH-t	***0.03***	***0.00***	***0.00***	***0.00***	**0.22**	***0.10***	***0.02***	***0.02***	***0.01***
		RGARCH-$\mathcal{N}$	1.00	**0.21**	1.00	1.00	0.47	0.40	0.83	0.77	0.28
		TGARCH-$\mathcal{N}$	0.34	1.00	0.47	0.48	1.00	1.00	1.00	1.00	1.00
		GARCH-$\mathcal{N}$	**0.19**	**0.21**	***0.01***	***0.01***	**0.22**	***0.01***	***0.00***	***0.00***	***0.01***
	5	RGARCH-t	***0.02***	***0.00***	***0.00***	***0.00***	**0.20**	0.26	**0.17**	**0.17**	***0.00***
		TGARCH-t	***0.03***	***0.00***	***0.00***	***0.00***	0.27	1.00	0.95	1.00	0.37
		GARCH-t	***0.03***	***0.00***	***0.00***	***0.00***	0.34	0.64	0.39	0.40	***0.05***
		RGARCH-$\mathcal{N}$	***0.06***	***0.00***	***0.00***	***0.00***	0.27	0.64	0.39	0.40	0.37
		TGARCH-$\mathcal{N}$	1.00	1.00	1.00	1.00	1.00	0.64	1.00	0.96	1.00
		GARCH-$\mathcal{N}$	**0.23**	0.44	**0.10**	***0.10***	0.35	***0.02***	***0.03***	***0.03***	***0.04***
0.05	1	RGARCH-t	1.00	0.94	1.00	1.00	0.81	0.75	0.95	0.93	1.00
		TGARCH-t	***0.07***	1.00	0.35	0.38	**0.24**	0.29	0.48	0.46	0.93
		GARCH-t	***0.00***	***0.05***	***0.00***	***0.00***	**0.16**	***0.01***	***0.00***	***0.00***	0.44
		RGARCH-$\mathcal{N}$	***0.04***	***0.00***	***0.00***	***0.00***	0.81	0.75	0.95	0.93	**0.21**
		TGARCH-$\mathcal{N}$	***0.02***	***0.00***	***0.00***	***0.00***	1.00	1.00	1.00	1.00	0.93
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	**0.18**	***0.01***	***0.00***	***0.00***	0.81
	5	RGARCH-t	***0.00***	***0.00***	***0.00***	***0.00***	**0.16**	0.67	**0.18**	**0.20**	***0.04***
		TGARCH-t	1.00	1.00	1.00	1.00	0.27	0.67	0.74	0.74	0.45
		GARCH-t	***0.00***	***0.06***	***0.04***	***0.04***	0.27	0.67	0.40	0.41	1.00
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	**0.16**	0.67	**0.14**	**0.17**	***0.01***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	1.00	1.00	1.00	***0.09***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.27	***0.00***	***0.00***	***0.00***	***0.01***
0.1	1	RGARCH-t	1.00	0.73	1.00	0.99	0.27	***0.02***	**0.16**	**0.12**	0.32
		TGARCH-t	0.30	1.00	0.96	1.00	0.48	**0.18**	0.57	0.49	1.00
		GARCH-t	***0.00***	**0.15**	***0.05***	***0.05***	**0.10**	***0.00***	***0.00***	***0.00***	***0.01***
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.80	0.50	0.59	0.57	0.32
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	1.00	1.00	1.00	0.32
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.26	***0.01***	***0.01***	***0.01***	**0.13**
	5	RGARCH-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.03***	***0.04***	***0.02***	***0.03***	***0.00***
		TGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		GARCH-t	***0.00***	**0.14**	**0.24**	**0.20**	**0.20**	***0.09***	***0.07***	***0.06***	**0.10**
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.02***	***0.09***	***0.02***	***0.03***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.07***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	**0.18**	***0.01***	***0.01***	***0.01***	***0.00***

NOTE: This table presents the MCS p-values implied by censored ($\flat$) and (corrected) conditional ($\sharp$) scoring rules underlying the results presented in Table 2, for different forecast horizons τ. Calculations are conducted by the `R` package `MCS`, developed by Bernardi and Catania (2018), using $B = 10,000$ simulations and block length $b = 5$. Bold (and italic) p-values signify a forecast method's elimination from $\mathrm{MCS}_{0.75}$ (and $\mathrm{MCS}_{0.90}$).

Table I.5 (continued): MCS p-values for logistic product risk management application.

			SphS				S_{ρ_1}				
q	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog	tw
0.15	1	RGARCH-t	1.00	0.74	0.81	0.79	0.28	***0.00***	***0.01***	***0.00***	***0.00***
		TGARCH-t	0.56	1.00	1.00	1.00	1.00	0.74	0.87	0.84	0.96
		GARCH-t	***0.00***	0.41	0.43	0.39	***0.06***	***0.02***	***0.03***	***0.02***	***0.00***
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.96	1.00	1.00	1.00	1.00
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.28	**0.12**	***0.08***	***0.09***	***0.00***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.28	**0.12**	***0.08***	***0.09***	0.96
	5	RGARCH-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.04***	***0.01***	***0.02***	***0.00***
		TGARCH-t	1.00	1.00	1.00	1.00	0.62	**0.17**	0.44	0.36	0.50
		GARCH-t	***0.00***	0.41	0.80	0.73	0.63	0.25	0.54	0.47	1.00
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.03***	***0.01***	***0.01***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	1.00	1.00	1.00	0.50
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.06***	***0.00***	***0.00***	***0.00***	***0.00***
0.2	1	RGARCH-t	1.00	0.76	0.89	0.83	**0.10**	***0.00***	***0.00***	***0.00***	***0.00***
		TGARCH-t	0.77	1.00	1.00	1.00	1.00	0.95	1.00	1.00	1.00
		GARCH-t	***0.00***	0.71	0.89	0.83	0.28	***0.00***	***0.01***	***0.00***	***0.00***
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.84	1.00	0.94	0.97	0.87
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.54	0.95	0.74	0.81	***0.03***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.28	***0.01***	***0.02***	***0.02***	***0.00***
	5	RGARCH-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.02***	***0.01***	***0.01***	***0.00***
		TGARCH-t	1.00	1.00	0.73	0.83	1.00	1.00	1.00	1.00	1.00
		GARCH-t	***0.00***	0.77	1.00	1.00	0.91	0.63	0.78	0.75	0.71
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.07***	***0.02***	***0.03***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.77	0.63	0.78	0.75	***0.01***
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	0.63	0.78	0.75	0.71
0.25	1	RGARCH-t	1.00	0.88	0.89	0.89	1.00	1.00	1.00	1.00	1.00
		TGARCH-t	0.85	1.00	0.90	1.00	0.96	0.75	0.84	0.81	0.74
		GARCH-t	***0.00***	0.88	1.00	0.97	0.46	***0.01***	***0.02***	***0.01***	0.74
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.01***	***0.01***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.96	0.88	0.88	0.89	0.74
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	**0.21**	***0.00***	***0.00***	***0.00***	***0.00***
	5	RGARCH-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.04***	0.27	**0.21**	**0.24**	***0.06***
		TGARCH-t	1.00	0.85	0.45	0.54	0.76	0.43	0.54	0.51	0.32
		GARCH-t	***0.01***	1.00	1.00	1.00	0.94	0.44	0.62	0.56	1.00
		RGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.02***	***0.02***	***0.02***	***0.00***
		TGARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	1.00	1.00	1.00	0.46
		GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.10***	***0.00***	***0.00***	***0.00***	***0.03***

NOTE: This table presents the MCS p-values implied by censored ($\flat$) and (corrected) conditional ($\sharp$) scoring rules underlying the results presented in Table 2, for different forecast horizons τ. Calculations are conducted by the `R` package `MCS` by Bernardi and Catania (2018), using $B = 10,000$ simulations and block length $b = 5$. Bold (and italic) p-values signify a forecast method's elimination from $\text{MCS}_{0.75}$ (and $\text{MCS}_{0.90}$).

I.3 Macroeconomics

The differences between the MCS variants are clearly highlighted by the p-values presented in Table I.7 (center application) and Table I.8 (tails application), which also offers more detailed insights. For $r = 1$ the cardinality of $\mathrm{MCS}^{\sharp}_{0.90}$ consistently exceeds or equals that of $\mathrm{MCS}^{\flat}_{0.90}$ with the sole exceptions occurring in tail cases predicated on the CRPS for $\tau = 24$. These exceptions feature a marginal difference of one. Finally, a closer look at the differences between the twCRPS and $\mathrm{CRPS}^{\flat}$ is in place. In Table I.7 (center application), we observe that the $\mathrm{CRPS}^{\flat}$ is preferred to the twCRPS for $\tau = 6$ and $\tau = 24$, with $r = 1$.

Table I.7: MCS p-values for center inflation application.

			LogS				QS			
r	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog
1	6	Random Walk	***0.09***	0.53	***0.09***	**0.13**	***0.05***	0.63	***0.08***	**0.11**
		AR	***0.09***	0.88	***0.09***	**0.13**	***0.02***	0.63	***0.05***	***0.09***
		Bagging	***0.00***	***0.02***	***0.00***	***0.00***	***0.01***	***0.07***	***0.00***	***0.00***
		CSR	0.26	0.88	0.26	0.48	***0.05***	0.63	**0.17**	0.34
		LASSO	***0.09***	0.25	***0.09***	**0.13**	***0.05***	0.54	***0.08***	**0.11**
		Random Forest	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	24	Random Walk	1.00	0.64	1.00	1.00	**0.10**	0.96	1.00	1.00
		AR	**0.22**	1.00	**0.22**	0.26	***0.09***	1.00	***0.09***	***0.10***
		Bagging	**0.14**	***0.05***	**0.14**	**0.13**	**0.11**	0.83	***0.02***	***0.01***
		CSR	**0.20**	0.51	**0.20**	**0.18**	***0.09***	0.89	***0.05***	***0.05***
		LASSO	**0.22**	0.48	**0.22**	0.26	**0.11**	0.96	***0.09***	***0.10***
		Random Forest	0.80	0.64	0.80	0.48	1.00	0.96	0.86	0.49
1.5	6	Random Walk	0.28	0.69	0.28	0.39	***0.08***	0.33	**0.18**	0.25
		AR	**0.21**	0.97	**0.21**	0.39	***0.05***	0.32	**0.11**	**0.21**
		Bagging	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		CSR	0.70	1.00	0.70	0.97	**0.22**	0.49	0.36	0.59
		LASSO	**0.21**	0.57	**0.21**	0.39	***0.08***	0.49	**0.12**	**0.21**
		Random Forest	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	24	Random Walk	1.00	0.91	1.00	1.00	***0.00***	0.99	1.00	1.00
		AR	**0.16**	0.94	**0.16**	0.36	***0.00***	0.99	***0.09***	***0.08***
		Bagging	***0.04***	***0.05***	***0.04***	***0.03***	***0.05***	***0.03***	***0.09***	***0.07***
		CSR	***0.04***	0.91	***0.04***	***0.07***	***0.00***	0.86	***0.09***	***0.07***
		LASSO	***0.04***	0.71	***0.04***	**0.14**	***0.02***	0.99	***0.09***	***0.07***
		Random Forest	0.37	1.00	0.37	0.36	1.00	1.00	0.28	**0.15**
2	6	Random Walk	0.38	0.67	0.38	0.43	***0.02***	0.51	**0.14**	0.38
		AR	**0.12**	0.82	**0.12**	0.40	***0.04***	0.53	***0.03***	**0.13**
		Bagging	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.02***	***0.00***	***0.00***
		CSR	1.00	1.00	1.00	1.00	**0.23**	0.68	0.55	0.99
		LASSO	0.38	0.63	0.38	0.43	***0.07***	0.68	**0.14**	0.38
		Random Forest	0.72	0.69	0.72	0.45	1.00	1.00	1.00	1.00
	24	Random Walk	1.00	0.75	1.00	1.00	***0.00***	1.00	1.00	1.00
		AR	0.37	1.00	0.37	0.41	***0.00***	0.74	***0.09***	***0.08***
		Bagging	***0.07***	***0.00***	***0.07***	***0.05***	***0.03***	***0.04***	***0.09***	***0.06***
		CSR	**0.20**	0.57	**0.20**	**0.24**	***0.00***	0.74	***0.08***	***0.06***
		LASSO	0.37	0.41	0.37	0.38	***0.01***	0.74	***0.09***	***0.08***
		Random Forest	0.57	0.57	0.57	0.41	1.00	0.74	0.60	**0.14**

NOTE: This table continues on the next page.

Table I.7 (continued): MCS p-values for center inflation application.

			SphS				CRPS				
r	τ	Method	♭	♯	sbar	slog	♭	♯	sbar	slog	tw
1	6	Random Walk	***0.01***	0.67	***0.09***	**0.15**	***0.07***	0.67	***0.08***	**0.13**	**0.18**
		AR	***0.00***	0.73	***0.07***	**0.15**	***0.03***	0.71	***0.06***	***0.08***	**0.25**
		Bagging	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	**0.18**
		CSR	***0.03***	0.73	**0.23**	0.43	***0.07***	0.71	**0.19**	0.40	0.89
		LASSO	***0.02***	0.44	***0.09***	**0.15**	***0.07***	0.32	***0.08***	***0.09***	1.00
		Random Forest	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89
	24	Random Walk	***0.08***	0.98	1.00	1.00	0.35	0.83	1.00	1.00	0.93
		AR	***0.04***	1.00	**0.13**	**0.13**	***0.06***	1.00	***0.08***	***0.07***	***0.05***
		Bagging	***0.08***	***0.08***	***0.02***	***0.02***	0.35	**0.21**	***0.03***	***0.01***	0.93
		CSR	***0.04***	0.61	***0.06***	***0.06***	**0.10**	0.74	***0.06***	***0.05***	***0.03***
		LASSO	***0.08***	0.67	**0.13**	**0.13**	0.35	0.76	***0.08***	***0.07***	0.93
		Random Forest	1.00	0.98	0.83	0.47	1.00	0.83	0.87	0.49	1.00
1.5	6	Random Walk	***0.05***	0.61	**0.20**	0.27	**0.12**	0.61	**0.21**	0.28	**0.12**
		AR	***0.05***	0.61	**0.13**	**0.24**	***0.05***	0.62	**0.13**	0.25	**0.19**
		Bagging	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	**0.19**
		CSR	**0.12**	0.61	0.41	0.67	**0.12**	0.62	0.44	0.70	0.99
		LASSO	***0.02***	0.61	**0.14**	**0.24**	**0.10**	0.61	**0.13**	0.25	0.99
		Random Forest	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	24	Random Walk	0.41	1.00	1.00	1.00	**0.24**	0.93	1.00	1.00	0.61
		AR	**0.23**	0.87	***0.04***	***0.06***	***0.01***	0.93	***0.09***	**0.12**	***0.04***
		Bagging	**0.23**	***0.00***	***0.00***	***0.00***	**0.24**	***0.04***	***0.09***	***0.06***	0.27
		CSR	***0.03***	0.75	***0.01***	***0.02***	***0.02***	0.93	***0.09***	***0.06***	***0.04***
		LASSO	**0.23**	0.75	***0.04***	***0.04***	**0.24**	0.93	***0.09***	***0.07***	0.29
		Random Forest	1.00	0.75	**0.22**	***0.06***	1.00	1.00	0.29	**0.15**	1.00
2	6	Random Walk	***0.05***	0.48	**0.18**	0.39	**0.17**	0.61	**0.19**	0.42	***0.06***
		AR	***0.05***	0.65	***0.03***	**0.16**	***0.01***	0.89	***0.03***	**0.18**	**0.16**
		Bagging	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.00***	***0.00***	**0.16**
		CSR	0.36	0.77	0.67	1.00	0.25	0.98	0.73	1.00	0.95
		LASSO	***0.02***	0.54	**0.17**	0.39	0.25	0.67	**0.19**	0.42	0.95
		Random Forest	1.00	1.00	1.00	0.90	1.00	1.00	1.00	0.83	1.00
	24	Random Walk	0.36	1.00	1.00	1.00	0.38	1.00	1.00	1.00	1.00
		AR	***0.07***	0.29	***0.07***	***0.07***	***0.00***	0.93	**0.20**	0.27	***0.09***
		Bagging	***0.05***	***0.01***	***0.07***	***0.04***	0.27	***0.01***	***0.08***	***0.08***	***0.09***
		CSR	***0.01***	0.27	***0.07***	***0.04***	***0.02***	0.93	***0.08***	**0.18**	***0.09***
		LASSO	***0.05***	0.27	***0.07***	***0.07***	0.27	0.93	**0.20**	0.27	***0.09***
		Random Forest	1.00	0.27	0.37	***0.09***	1.00	0.93	0.65	0.27	0.91

NOTE: This table presents the MCS p-values implied by censored (♭) and (corrected) conditional (♯) scoring rules underlying the results presented in Table 2 and twCRPS, for different forecast horizons τ. Calculations are conducted by the `R` package `MCS` by Bernardi and Catania (2018), using $B = 10,000$ simulations and block length $b = 5$. Bold (and italic) p-values signify a forecast method's elimination from $\text{MCS}_{0.75}$ (and $\text{MCS}_{0.90}$).

Table I.8: MCS p-values for tails inflation application.

			LogS				QS			
r_1	h	Method	♭	♯	sbar	slog	♭	♯	sbar	slog
1	6	Random Walk	0.26	0.46	0.26	0.31	***0.02***	**0.21**	***0.10***	***0.03***
		AR	***0.01***	**0.17**	***0.01***	***0.03***	***0.01***	0.30	***0.02***	***0.03***
		Bagging	***0.00***	***0.08***	***0.00***	***0.05***	***0.00***	**0.21**	***0.00***	***0.01***
		CSR	1.00	1.00	1.00	1.00	**0.18**	1.00	0.30	0.37
		LASSO	0.43	0.78	0.43	0.75	***0.04***	0.95	***0.10***	0.33
		Random Forest	0.43	0.46	0.43	0.35	1.00	0.95	1.00	1.00
	24	Random Walk	1.00	0.59	1.00	0.67	1.00	0.74	1.00	***0.06***
		AR	**0.13**	0.59	**0.13**	0.67	***0.00***	0.74	***0.04***	***0.03***
		Bagging	***0.05***	0.29	***0.05***	0.32	***0.00***	0.61	***0.01***	***0.06***
		CSR	***0.05***	0.44	***0.05***	0.37	***0.00***	0.61	***0.04***	***0.00***
		LASSO	**0.13**	1.00	**0.13**	1.00	***0.00***	1.00	***0.04***	***0.06***
		Random Forest	0.32	0.30	0.32	0.67	***0.08***	0.61	0.66	1.00
1.5	6	Random Walk	**0.18**	0.27	**0.18**	**0.11**	***0.03***	0.56	**0.17**	***0.02***
		AR	***0.00***	0.38	***0.00***	**0.11**	***0.02***	0.61	***0.04***	***0.02***
		Bagging	***0.02***	0.27	***0.02***	**0.13**	***0.00***	0.56	***0.00***	***0.02***
		CSR	1.00	0.68	1.00	0.87	**0.25**	0.81	0.48	**0.20**
		LASSO	0.52	1.00	0.52	1.00	***0.03***	1.00	**0.16**	***0.05***
		Random Forest	0.47	0.41	0.47	0.53	1.00	0.81	1.00	1.00
	24	Random Walk	1.00	0.36	1.00	0.98	1.00	0.48	1.00	***0.02***
		AR	***0.08***	0.42	***0.08***	**0.19**	***0.00***	0.50	***0.07***	***0.00***
		Bagging	***0.08***	0.42	***0.08***	0.78	***0.00***	1.00	***0.07***	***0.02***
		CSR	***0.06***	0.36	***0.06***	***0.07***	***0.00***	0.35	***0.07***	***0.00***
		LASSO	***0.08***	1.00	***0.08***	0.98	***0.00***	0.78	***0.07***	***0.00***
		Random Forest	0.28	0.42	0.28	1.00	**0.10**	0.78	0.44	1.00
2	6	Random Walk	**0.17**	0.27	**0.17**	**0.10**	**0.14**	0.65	**0.11**	***0.01***
		AR	***0.00***	**0.20**	***0.00***	***0.00***	***0.05***	0.65	***0.01***	***0.00***
		Bagging	***0.06***	0.27	***0.06***	**0.19**	***0.01***	0.98	***0.00***	***0.01***
		CSR	1.00	0.27	1.00	0.44	0.67	0.98	0.63	**0.12**
		LASSO	0.67	1.00	0.67	1.00	**0.12**	0.98	***0.08***	***0.01***
		Random Forest	0.59	0.27	0.59	0.44	1.00	1.00	1.00	1.00
	24	Random Walk	1.00	**0.19**	1.00	0.99	1.00	**0.18**	1.00	***0.00***
		AR	***0.07***	**0.19**	***0.07***	**0.13**	***0.00***	**0.18**	***0.03***	***0.00***
		Bagging	***0.03***	0.27	***0.03***	0.99	***0.00***	**0.18**	***0.02***	***0.00***
		CSR	***0.02***	0.27	***0.02***	***0.05***	***0.00***	**0.18**	***0.02***	***0.00***
		LASSO	***0.07***	1.00	***0.07***	0.95	***0.00***	1.00	***0.05***	***0.00***
		Random Forest	**0.23**	0.27	**0.23**	1.00	***0.07***	**0.18**	0.70	1.00

NOTE: This table continues on the next page.

Table I.8 (continued): MCS p-values for tails inflation application.

			SphS				CRPS				
r_1	h	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog	tw
1	6	Random Walk	***0.04***	0.44	**0.11**	***0.03***	0.32	**0.18**	***0.02***	***0.10***	***0.06***
		AR	***0.01***	0.51	***0.02***	***0.03***	1.00	**0.13**	***0.01***	***0.08***	***0.03***
		Bagging	***0.00***	**0.13**	***0.00***	***0.00***	0.31	0.90	***0.01***	0.93	***0.04***
		CSR	**0.14**	1.00	0.43	0.65	***0.10***	0.90	0.85	0.93	0.28
		LASSO	***0.05***	0.89	**0.11**	0.57	0.31	1.00	0.85	1.00	**0.20**
		Random Forest	1.00	0.89	1.00	1.00	***0.04***	0.90	1.00	0.93	1.00
	24	Random Walk	0.76	0.61	1.00	**0.23**	0.44	0.31	1.00	0.65	1.00
		AR	***0.00***	0.61	***0.05***	**0.19**	0.95	0.31	***0.01***	***0.07***	***0.00***
		Bagging	***0.00***	0.55	***0.02***	**0.21**	1.00	0.27	***0.01***	0.36	***0.00***
		CSR	***0.00***	0.55	***0.04***	***0.01***	0.95	***0.10***	***0.01***	***0.03***	***0.00***
		LASSO	***0.00***	1.00	***0.05***	**0.23**	0.95	1.00	***0.01***	1.00	***0.00***
		Random Forest	1.00	0.55	0.65	1.00	**0.23**	0.31	0.56	0.89	0.32
1.5	6	Random Walk	***0.02***	0.55	**0.21**	***0.02***	0.39	**0.23**	***0.09***	**0.12**	**0.10**
		AR	***0.02***	0.66	***0.04***	***0.04***	1.00	0.44	***0.02***	**0.22**	***0.04***
		Bagging	***0.00***	**0.15**	***0.00***	***0.02***	0.29	1.00	***0.09***	0.79	***0.04***
		CSR	**0.19**	0.84	0.59	0.42	**0.17**	0.59	0.91	0.79	**0.24**
		LASSO	***0.02***	1.00	**0.21**	**0.18**	0.29	0.80	0.86	1.00	**0.11**
		Random Forest	1.00	0.84	1.00	1.00	***0.08***	0.59	1.00	0.79	1.00
	24	Random Walk	0.90	0.44	1.00	***0.05***	0.50	0.40	1.00	***0.05***	1.00
		AR	***0.01***	0.44	***0.08***	***0.01***	0.73	0.27	**0.10**	***0.00***	***0.00***
		Bagging	***0.01***	1.00	***0.08***	***0.05***	1.00	0.59	**0.10**	0.54	***0.00***
		CSR	***0.00***	0.42	***0.07***	***0.00***	0.71	0.27	***0.07***	***0.04***	***0.00***
		LASSO	***0.02***	0.55	***0.08***	***0.03***	0.71	1.00	**0.10**	0.54	***0.00***
		Random Forest	1.00	0.55	0.44	1.00	0.50	0.59	0.44	1.00	0.29
2	6	Random Walk	***0.02***	0.65	**0.12**	***0.01***	0.34	0.27	***0.08***	***0.03***	**0.19**
		AR	***0.02***	0.65	***0.01***	***0.01***	1.00	0.27	***0.01***	***0.03***	**0.11**
		Bagging	***0.00***	0.65	***0.00***	***0.01***	0.34	1.00	**0.17**	0.89	***0.05***
		CSR	0.37	1.00	0.75	0.28	0.30	0.58	1.00	0.89	**0.24**
		LASSO	***0.02***	0.97	***0.09***	***0.01***	0.34	0.60	0.62	1.00	**0.24**
		Random Forest	1.00	0.97	1.00	1.00	**0.22**	0.58	0.96	0.89	1.00
	24	Random Walk	1.00	**0.19**	1.00	***0.01***	0.52	0.25	1.00	***0.03***	1.00
		AR	***0.00***	**0.17**	***0.03***	***0.00***	0.56	0.25	***0.04***	***0.00***	***0.01***
		Bagging	***0.00***	**0.21**	***0.02***	***0.01***	1.00	0.25	***0.03***	**0.23**	***0.01***
		CSR	***0.00***	**0.21**	***0.02***	***0.00***	0.38	0.25	***0.02***	***0.01***	***0.00***
		LASSO	***0.00***	1.00	***0.06***	***0.00***	0.38	1.00	***0.08***	**0.22**	***0.01***
		Random Forest	0.80	**0.17**	0.66	1.00	0.52	0.25	0.68	1.00	0.46

NOTE: This table presents the MCS p-values implied by censored ($\flat$) and (corrected) conditional ($\sharp$) scoring rules underlying the results presented in Table 2 and twCRPS, for different forecast horizons τ. Calculations are conducted by the R package `MCS` by Bernardi and Catania (2018), using $B = 10,000$ simulations and block length $b = 5$. Bold (and italic) p-values signify a forecast method's elimination from $\text{MCS}_{0.75}$ (and $\text{MCS}_{0.90}$).

I.4 Climate

The MCS p-values for the climate application are reported in Table I.9 (right tail application) and Table I.10 (center application). The MCS p-values in Table I.10 reveal that the MCSs are consistently small in the center application, frequently including one or both of the QGARCH-II methods. This observation suggests that the preference for censoring, as depicted in Table 2, translates into the censored scoring rule's more effective recognition of the QGARCH-II methods' pronounced superiority. Table I.10 further demonstrates that the performance of the $\mathrm{CRPS}^{\flat}$ and twCRPS is closely matched.

Table I.9: MCS p-values for right tail climate application.

			LogS				QS			
q	τ	Method	♭	♯	sbar	slog	♭	♯	sbar	slog
0.99	1	GARCH-$\mathcal{N}$	0.66	1.00	0.66	1.00	***0.05***	1.00	**0.24**	0.95
		GARCH-t	0.78	0.34	0.78	0.85	***0.05***	0.86	0.26	1.00
		QGARCH-I-$\mathcal{N}$	0.47	0.34	0.47	0.85	***0.04***	0.58	**0.14**	0.97
		QGARCH-I-t	0.47	0.34	0.47	0.85	***0.04***	0.58	**0.17**	0.98
		QGARCH-II-$\mathcal{N}$	0.33	***0.05***	0.33	0.35	1.00	***0.06***	0.28	0.84
		QGARCH-II-t	1.00	***0.05***	1.00	0.56	0.74	***0.06***	1.00	0.98
	3	GARCH-$\mathcal{N}$	***0.00***	***0.04***	***0.00***	***0.01***	***0.00***	**0.24**	***0.00***	***0.02***
		GARCH-t	1.00	1.00	1.00	1.00	***0.00***	1.00	**0.17**	1.00
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.04***	***0.00***	***0.01***	***0.00***	***0.06***	***0.00***	***0.01***
		QGARCH-I-t	***0.00***	***0.04***	***0.00***	***0.01***	***0.00***	***0.06***	***0.00***	***0.01***
		QGARCH-II-$\mathcal{N}$	***0.00***	***0.02***	***0.00***	***0.00***	1.00	***0.06***	***0.00***	***0.01***
		QGARCH-II-t	0.94	***0.03***	0.94	***0.06***	0.31	**0.23**	1.00	0.32
0.95	1	GARCH-$\mathcal{N}$	***0.03***	0.80	***0.03***	**0.17**	***0.00***	1.00	***0.00***	**0.13**
		GARCH-t	***0.08***	1.00	***0.08***	1.00	***0.00***	***0.09***	***0.01***	0.54
		QGARCH-I-$\mathcal{N}$	***0.01***	***0.03***	***0.01***	***0.09***	***0.00***	***0.09***	***0.00***	0.40
		QGARCH-I-t	***0.01***	***0.03***	***0.01***	***0.06***	***0.00***	***0.03***	***0.00***	0.30
		QGARCH-II-$\mathcal{N}$	***0.03***	**0.14**	***0.03***	***0.06***	1.00	**0.24**	***0.06***	0.40
		QGARCH-II-t	1.00	**0.14**	1.00	0.53	0.38	**0.24**	1.00	1.00
	3	GARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	0.73	***0.00***	***0.01***
		GARCH-t	***0.00***	1.00	***0.00***	1.00	***0.00***	0.73	***0.00***	0.71
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-II-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	0.91	***0.01***	***0.00***	***0.02***
		QGARCH-II-t	1.00	***0.01***	1.00	***0.08***	1.00	1.00	1.00	1.00
0.9	1	GARCH-$\mathcal{N}$	***0.00***	0.42	***0.00***	***0.03***	***0.00***	0.76	***0.00***	***0.01***
		GARCH-t	***0.01***	1.00	***0.01***	0.97	***0.00***	0.76	***0.00***	0.50
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.02***	***0.00***	***0.00***	***0.00***	***0.10***	***0.00***	***0.04***
		QGARCH-I-t	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.05***	***0.00***	***0.03***
		QGARCH-II-$\mathcal{N}$	***0.01***	***0.07***	***0.01***	***0.02***	1.00	0.76	**0.14**	**0.14**
		QGARCH-II-t	1.00	0.70	1.00	1.00	***0.01***	1.00	1.00	1.00
	3	GARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	**0.15**	***0.00***	***0.01***
		GARCH-t	***0.00***	1.00	***0.00***	1.00	***0.00***	**0.15**	***0.00***	0.69
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-II-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	0.41	***0.00***	***0.02***
		QGARCH-II-t	1.00	***0.02***	1.00	**0.10**	***0.03***	1.00	1.00	1.00

NOTE: This table continues on the next page.

Table I.9 (continued): MCS p-values for right tail climate application (Continued).

			LogS				QS			
q	τ	Method	♭	♯	sbar	slog	♭	♯	sbar	slog
0.85	1	GARCH-$\mathcal{N}$	***0.00***	**0.20**	***0.00***	***0.05***	***0.00***	***0.07***	***0.00***	***0.03***
		GARCH-t	***0.00***	0.33	***0.00***	**0.16**	***0.00***	***0.07***	***0.00***	**0.11**
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.02***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***
		QGARCH-II-$\mathcal{N}$	***0.00***	***0.03***	***0.00***	***0.03***	1.00	0.27	0.52	0.34
		QGARCH-II-t	1.00	1.00	1.00	1.00	***0.01***	1.00	1.00	1.00
	3	GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		GARCH-t	***0.00***	1.00	***0.00***	1.00	***0.00***	***0.00***	***0.00***	0.25
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-II-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	***0.00***	***0.00***	***0.02***
		QGARCH-II-t	1.00	0.59	1.00	0.73	***0.00***	1.00	1.00	1.00
0.8	1	GARCH-$\mathcal{N}$	***0.00***	***0.07***	***0.00***	***0.03***	***0.00***	***0.02***	***0.00***	***0.05***
		GARCH-t	***0.00***	***0.07***	***0.00***	***0.03***	***0.00***	***0.02***	***0.00***	***0.05***
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.00***	***0.00***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***	***0.00***	***0.00***
		QGARCH-II-$\mathcal{N}$	***0.01***	***0.03***	***0.01***	***0.03***	1.00	1.00	1.00	0.87
		QGARCH-II-t	1.00	1.00	1.00	1.00	***0.02***	0.63	0.87	1.00
	3	GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.01***
		GARCH-t	***0.00***	0.30	***0.00***	0.32	***0.00***	***0.00***	***0.00***	***0.06***
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-II-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	**0.25**	***0.00***	***0.02***
		QGARCH-II-t	1.00	1.00	1.00	1.00	***0.00***	1.00	1.00	1.00
0.75	1	GARCH-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.02***
		GARCH-t	***0.00***	***0.02***	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	***0.02***
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-II-$\mathcal{N}$	***0.00***	***0.02***	***0.00***	***0.01***	1.00	1.00	**0.21**	**0.11**
		QGARCH-II-t	1.00	1.00	1.00	1.00	**0.16**	0.75	1.00	1.00
	3	GARCH-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.02***
		GARCH-t	***0.00***	***0.01***	***0.00***	***0.02***	***0.00***	***0.00***	***0.00***	***0.03***
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***
		QGARCH-II-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	1.00	1.00	***0.00***	***0.01***
		QGARCH-II-t	1.00	1.00	1.00	1.00	***0.00***	0.98	1.00	1.00

NOTE: This table continues on the next page.

Table I.9 (continued): MCS p-values for right tail climate application (Continued).

			SphS				CRPS			
q	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog
0.99	1	GARCH-$\mathcal{N}$	0.25	0.69	0.31	0.92	0.45	0.56	0.56	0.56
		GARCH-t	0.25	1.00	0.39	1.00	0.45	1.00	1.00	1.00
		QGARCH-I-$\mathcal{N}$	***0.01***	0.56	**0.17**	0.93	***0.03***	***0.05***	***0.05***	***0.05***
		QGARCH-I-t	***0.03***	0.56	**0.24**	0.93	***0.05***	***0.05***	***0.05***	***0.05***
		QGARCH-II-$\mathcal{N}$	1.00	***0.10***	0.39	0.80	1.00	***0.05***	***0.05***	***0.05***
		QGARCH-II-t	0.56	***0.10***	1.00	0.93	0.80	***0.05***	***0.05***	***0.05***
	3	GARCH-$\mathcal{N}$	***0.00***	***0.03***	***0.00***	***0.02***	***0.00***	***0.10***	***0.10***	***0.10***
		GARCH-t	***0.00***	1.00	0.33	1.00	***0.00***	1.00	1.00	1.00
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.02***	***0.00***	***0.01***	***0.00***	***0.09***	***0.09***	***0.09***
		QGARCH-I-t	***0.00***	***0.02***	***0.00***	***0.01***	***0.00***	***0.10***	***0.10***	***0.10***
		QGARCH-II-$\mathcal{N}$	0.71	***0.02***	***0.00***	***0.01***	**0.21**	***0.07***	***0.07***	***0.07***
		QGARCH-II-t	1.00	***0.02***	1.00	**0.20**	1.00	***0.07***	***0.07***	***0.07***
0.95	1	GARCH-$\mathcal{N}$	***0.00***	1.00	***0.01***	**0.19**	***0.00***	**0.18**	**0.18**	**0.18**
		GARCH-t	***0.00***	***0.09***	***0.01***	0.82	***0.00***	1.00	1.00	1.00
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.09***	***0.00***	0.33	***0.00***	***0.06***	***0.06***	***0.06***
		QGARCH-I-t	***0.00***	***0.05***	***0.00***	**0.23**	***0.00***	***0.06***	***0.06***	***0.06***
		QGARCH-II-$\mathcal{N}$	0.71	***0.09***	***0.05***	0.33	1.00	**0.18**	**0.18**	**0.18**
		QGARCH-II-t	1.00	***0.09***	1.00	1.00	0.81	**0.18**	**0.18**	**0.18**
	3	GARCH-$\mathcal{N}$	***0.00***	0.36	***0.00***	***0.01***	***0.00***	***0.08***	***0.08***	***0.08***
		GARCH-t	***0.00***	1.00	***0.00***	1.00	***0.00***	1.00	1.00	1.00
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.08***	***0.08***	***0.08***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.08***	***0.08***	***0.08***
		QGARCH-II-$\mathcal{N}$	0.75	***0.01***	***0.00***	***0.02***	1.00	***0.08***	***0.08***	***0.08***
		QGARCH-II-t	1.00	***0.04***	1.00	0.80	0.88	***0.08***	***0.08***	***0.08***
0.9	1	GARCH-$\mathcal{N}$	***0.00***	1.00	***0.00***	***0.03***	***0.00***	***0.07***	***0.07***	***0.07***
		GARCH-t	***0.00***	0.92	***0.00***	0.73	***0.00***	1.00	1.00	1.00
		QGARCH-I-$\mathcal{N}$	***0.00***	0.48	***0.00***	***0.04***	***0.00***	***0.07***	***0.07***	***0.07***
		QGARCH-I-t	***0.00***	0.28	***0.00***	***0.03***	***0.00***	***0.05***	***0.05***	***0.05***
		QGARCH-II-$\mathcal{N}$	1.00	0.52	**0.13**	**0.15**	1.00	***0.05***	***0.05***	***0.05***
		QGARCH-II-t	**0.14**	0.92	1.00	1.00	0.87	***0.05***	***0.05***	***0.05***
	3	GARCH-$\mathcal{N}$	***0.00***	0.46	***0.00***	***0.01***	***0.00***	***0.05***	***0.05***	***0.05***
		GARCH-t	***0.00***	1.00	***0.00***	1.00	***0.00***	1.00	1.00	1.00
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.05***	***0.05***	***0.05***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.05***	***0.05***	***0.05***
		QGARCH-II-$\mathcal{N}$	1.00	***0.04***	***0.00***	***0.01***	1.00	***0.05***	***0.05***	***0.05***
		QGARCH-II-t	***0.06***	0.46	1.00	0.80	0.51	***0.05***	***0.05***	***0.05***

NOTE: This table continues on the next page.

Table I.9 (continued): MCS p-values for right tail climate application (Continued).

			SphS				CRPS			
q	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog
0.85	1	GARCH-$\mathcal{N}$	***0.00***	**0.24**	***0.00***	***0.02***	***0.00***	0.55	0.55	0.55
		GARCH-t	***0.00***	0.74	***0.00***	0.27	***0.00***	1.00	1.00	1.00
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.01***	***0.00***	***0.02***	***0.00***	***0.04***	***0.04***	***0.04***
		QGARCH-I-t	***0.00***	***0.01***	***0.00***	***0.01***	***0.00***	***0.04***	***0.04***	***0.04***
		QGARCH-II-$\mathcal{N}$	1.00	***0.07***	**0.20**	**0.20**	1.00	***0.04***	***0.04***	***0.04***
		QGARCH-II-t	**0.11**	1.00	1.00	1.00	0.93	***0.04***	***0.04***	***0.04***
	3	GARCH-$\mathcal{N}$	***0.00***	**0.10**	***0.00***	***0.01***	***0.00***	***0.10***	***0.10***	***0.10***
		GARCH-t	***0.00***	1.00	***0.00***	0.92	***0.00***	1.00	1.00	1.00
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.10***	***0.10***	***0.10***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.10***	***0.10***	***0.10***
		QGARCH-II-$\mathcal{N}$	1.00	***0.00***	***0.00***	***0.01***	1.00	***0.10***	***0.10***	***0.10***
		QGARCH-II-t	***0.00***	0.95	1.00	1.00	**0.14**	***0.08***	***0.08***	***0.08***
0.8	1	GARCH-$\mathcal{N}$	***0.00***	0.65	***0.00***	**0.21**	***0.00***	**0.21**	**0.21**	**0.21**
		GARCH-t	***0.00***	0.65	***0.00***	**0.21**	***0.00***	1.00	1.00	1.00
		QGARCH-I-$\mathcal{N}$	***0.00***	0.30	***0.00***	***0.01***	***0.00***	***0.04***	***0.04***	***0.04***
		QGARCH-I-t	***0.00***	**0.12**	***0.00***	***0.00***	***0.00***	***0.04***	***0.04***	***0.04***
		QGARCH-II-$\mathcal{N}$	1.00	0.69	0.93	0.71	1.00	***0.04***	***0.04***	***0.04***
		QGARCH-II-t	***0.02***	1.00	1.00	1.00	0.45	***0.04***	***0.04***	***0.04***
	3	GARCH-$\mathcal{N}$	***0.00***	0.40	***0.00***	***0.02***	***0.00***	***0.10***	***0.10***	***0.10***
		GARCH-t	***0.00***	0.40	***0.00***	0.53	***0.00***	1.00	1.00	1.00
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.10***	***0.10***	***0.10***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.09***	***0.09***	***0.09***
		QGARCH-II-$\mathcal{N}$	1.00	***0.02***	***0.00***	***0.02***	1.00	***0.07***	***0.07***	***0.07***
		QGARCH-II-t	***0.00***	1.00	1.00	1.00	***0.01***	***0.07***	***0.07***	***0.07***
0.75	1	GARCH-$\mathcal{N}$	***0.00***	**0.12**	***0.00***	**0.18**	***0.00***	**0.25**	**0.25**	**0.25**
		GARCH-t	***0.00***	***0.10***	***0.00***	**0.18**	***0.00***	0.80	0.80	0.80
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.02***	***0.00***	***0.00***	***0.00***	***0.09***	***0.09***	***0.09***
		QGARCH-I-t	***0.00***	***0.01***	***0.00***	***0.00***	***0.00***	1.00	1.00	1.00
		QGARCH-II-$\mathcal{N}$	1.00	0.61	**0.16**	**0.18**	1.00	**0.19**	**0.19**	**0.19**
		QGARCH-II-t	**0.13**	1.00	1.00	1.00	**0.18**	**0.19**	**0.19**	**0.19**
	3	GARCH-$\mathcal{N}$	***0.00***	***0.03***	***0.00***	**0.18**	***0.00***	***0.10***	***0.10***	***0.10***
		GARCH-t	***0.00***	***0.03***	***0.00***	0.35	***0.00***	1.00	1.00	1.00
		QGARCH-I-$\mathcal{N}$	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	***0.10***	***0.10***	***0.10***
		QGARCH-I-t	***0.00***	***0.00***	***0.00***	***0.00***	***0.00***	0.99	0.99	0.99
		QGARCH-II-$\mathcal{N}$	1.00	***0.03***	***0.00***	***0.01***	1.00	**0.11**	**0.11**	**0.11**
		QGARCH-II-t	***0.00***	1.00	1.00	1.00	***0.00***	***0.10***	***0.10***	***0.10***

NOTE: This table presents the MCS p-values implied by censored ($\flat$) and (corrected) conditional ($\sharp$) scoring rules underlying the results presented in Table 2, for different forecast horizons τ. Calculations are conducted by the `R` package `MCS` by Bernardi and Catania (2018), using $B = 10,000$ simulations and block length $b = 200$. Bold (and italic) p-values signify a forecast method's elimination from $\text{MCS}_{0.75}$ (and $\text{MCS}_{0.90}$).

Table I.10: MCS p-values for center climate application.

			LogS				QS			
q	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog
1	1	GARCH-$\mathcal{N}$	**0.00**	0.27	**0.00**	**0.00**	**0.00**	0.46	**0.00**	**0.00**
		GARCH-t	**0.00**	0.61	**0.00**	**0.00**	**0.00**	0.46	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.05**	**0.00**	**0.00**	**0.00**	**0.03**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.12**	**0.00**	**0.00**	**0.00**	**0.07**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	***1.00***	**0.21**	***1.00***	0.84	***1.00***	0.46	***1.00***	***1.00***
		QGARCH-II-t	0.87	***1.00***	0.87	***1.00***	**0.02**	***1.00***	**0.15**	0.31
	3	GARCH-$\mathcal{N}$	**0.01**	**0.04**	**0.01**	**0.02**	**0.00**	**0.01**	**0.01**	**0.01**
		GARCH-t	**0.01**	0.31	**0.01**	**0.02**	**0.00**	**0.06**	**0.01**	**0.01**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	**0.01**	**0.04**	**0.01**	**0.02**	***1.00***	**0.02**	**0.05**	**0.03**
		QGARCH-II-t	***1.00***	***1.00***	***1.00***	***1.00***	**0.00**	***1.00***	***1.00***	***1.00***
2	1	GARCH-$\mathcal{N}$	**0.00**	**0.12**	**0.00**	**0.00**	**0.00**	**0.03**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.11**	**0.00**	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.03**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.03**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	0.66	0.74	0.66	0.41	***1.00***	***1.00***	0.96	0.52
		QGARCH-II-t	***1.00***	***1.00***	***1.00***	***1.00***	**0.13**	0.40	***1.00***	***1.00***
	3	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.01**	**0.00**	**0.00**	***1.00***	0.60	**0.06**	**0.01**
		QGARCH-II-t	***1.00***	***1.00***	***1.00***	***1.00***	**0.00**	***1.00***	***1.00***	***1.00***
4	1	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	**0.03**	**0.03**	**0.03**	**0.02**	***1.00***	0.53	**0.13**	**0.20**
		QGARCH-II-t	***1.00***	***1.00***	***1.00***	***1.00***	**0.06**	***1.00***	***1.00***	***1.00***
	3	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	***1.00***	**0.24**	**0.00**	**0.00**
		QGARCH-II-t	***1.00***	***1.00***	***1.00***	***1.00***	**0.00**	***1.00***	***1.00***	***1.00***

NOTE: This table continues on the next page.

Table I.10: MCS p-values for center climate application (Continued).

			SphS				CRPS				
r	τ	Method	$\flat$	$\sharp$	sbar	slog	$\flat$	$\sharp$	sbar	slog	tw
1	1	GARCH-$\mathcal{N}$	**0.00**	0.43	**0.00**	**0.00**	**0.00**	0.44	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	0.43	**0.00**	**0.00**	**0.00**	0.47	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.03**	**0.00**	**0.00**	**0.00**	**0.05**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.04**	**0.00**	**0.00**	**0.00**	**0.11**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	***1.00***	0.38	***1.00***	***1.00***	***1.00***	0.34	***1.00***	***1.00***	***1.00***
		QGARCH-II-t	**0.02**	***1.00***	0.47	0.74	**0.01**	***1.00***	**0.16**	0.31	0.36
	3	GARCH-$\mathcal{N}$	**0.00**	**0.01**	**0.01**	**0.01**	**0.00**	**0.03**	**0.01**	**0.01**	**0.00**
		GARCH-t	**0.00**	**0.13**	**0.01**	**0.01**	**0.00**	**0.19**	**0.01**	**0.01**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	***1.00***	**0.01**	**0.02**	**0.01**	***1.00***	**0.03**	**0.05**	**0.02**	***1.00***
		QGARCH-II-t	**0.00**	***1.00***	***1.00***	***1.00***	**0.00**	***1.00***	***1.00***	***1.00***	**0.00**
2	1	GARCH-$\mathcal{N}$	**0.00**	**0.10**	**0.00**	**0.00**	**0.00**	**0.05**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.06**	**0.00**	**0.00**	**0.00**	**0.05**	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	***1.00***	***1.00***	0.81	0.47	***1.00***	***1.00***	0.84	0.46	***1.00***
		QGARCH-II-t	**0.09**	0.92	***1.00***	***1.00***	0.51	0.83	***1.00***	***1.00***	**0.15**
	3	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	***1.00***	**0.00**	**0.01**	**0.00**	***1.00***	**0.16**	**0.05**	**0.01**	***1.00***
		QGARCH-II-t	**0.00**	***1.00***	***1.00***	***1.00***	**0.00**	***1.00***	***1.00***	***1.00***	**0.00**
4	1	GARCH-$\mathcal{N}$	**0.00**	**0.02**	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.02**	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	***1.00***	**0.02**	**0.04**	**0.05**	***1.00***	0.67	**0.19**	0.28	***1.00***
		QGARCH-II-t	**0.05**	***1.00***	***1.00***	***1.00***	0.74	***1.00***	***1.00***	***1.00***	**0.18**
	3	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	***1.00***	**0.00**	**0.00**	**0.00**	***1.00***	0.71	**0.00**	**0.01**	***1.00***
		QGARCH-II-t	**0.00**	***1.00***	***1.00***	***1.00***	**0.00**	***1.00***	***1.00***	***1.00***	**0.00**

NOTE: This table presents the MCS p-values implied by censored ($\flat$) and (corrected) conditional ($\sharp$) scoring rules underlying the results presented in Table 2 and twCRPS, for different forecast horizons τ. Calculations are conducted by the `R` package `MCS` by Bernardi and Catania (2018), using $B = 10,000$ simulations and block length $b = 200$. Bold (and italic) p-values signify a forecast method's elimination from $\text{MCS}_{0.75}$ (and $\text{MCS}_{0.90}$).

References

Allen, S., D. Ginsbourger, and J. Ziegel (2023), "Evaluating Forecasts for High-Impact Events Using Transformed Kernel Scores", *SIAM/ASA Journal on Uncertainty Quantification*, *11*(3), 906–940.

Amisano, G. and R. Giacomini (2007), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests", *Journal of Business & Economic Statistics*, *25*(2), 177–190.

Bernardi, M. and L. Catania (2018), "The Model Confidence Set Package for R", *International Journal of Computational Economics and Econometrics*, *8*(2), 144–158.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, *31*(3), 307–327.

Breiman, L. (2001), "Random forests", *Machine Learning*, *45*(1), 5–32.

Clements, M. P. (2004), "Evaluating the Bank of England Density Forecasts of Inflation", *The Economic Journal*, *114*(498), 844–866.

Diebold, F. X. and R. S. Mariano (2002), "Comparing Predictive Accuracy", *Journal of Business & Economic Statistics*, *20*(1), 134–144.

Diks, C., V. Panchenko, and D. Van Dijk (2011), "Likelihood-based Scoring Rules for Comparing Density Forecasts in Tails", *Journal of Econometrics*, *163*(2), 215–230.

Elliott, G., A. Gargano, and A. Timmermann (2013), "Complete subset regressions", *Journal of Econometrics*, *177*(2), 357–373.

Elliott, G., A. Gargano, and A. Timmermann (2015), "Complete subset regressions with large-dimensional sets of predictors", *Journal of Economic Dynamics and Control*, *54*, 86–110.

Engle, R. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models", *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Franses, P. H., J. Neele, and D. Van Dijk (2001), "Modeling Asymmetric Volatility in Weekly Dutch Temperature Data", *Environmental Modelling & Software*, *16*(2), 131–137.

Giacomini, R. and H. White (2006), "Tests of Conditional Predictive Ability", *Econometrica*, *74*(6), 1545–1578.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *The Journal of Finance*, *48*(5), 1779–1801.

Gneiting, T. and A. E. Raftery (2007), "Strictly Proper Scoring Rules, Prediction, and Estimation", *Journal of the American Statistical Association*, *102*(477), 359–378.

Gneiting, T. and R. Ranjan (2011), "Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules", *Journal of Business & Economic Statistics*, *29*(3), 411–422.

Hansen, P. R., Z. Huang, and H. H. Shek (2012), "Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility", *Journal of Applied Econometrics*, *27*(6), 877–906.

Hansen, P. R., A. Lunde, and J. Nason (2011), "The Model Confidence Set", *Econometrica*, *79*(2), 453–497.

Holzmann, H. and B. Klar (2017a), "Focusing on Regions of Interest in Forecast Evaluation", *The Annals of Applied Statistics*, *11*(4), 2404–2431.

Holzmann, H. and B. Klar (2017b). "Weighted Scoring Rules and Hypothesis Testing". Available at `https://arxiv.org/abs/1611.07345v2`.

Lerch, S., T. L. Thorarinsdottir, F. Ravazzolo, and T. Gneiting (2017), "Forecaster's Dilemma: Extreme Events and Forecast Evaluation", *Statistical Science*, *32*(1), 106–127.

Medeiros, M. C., G. F. R. Vasconcelos, A. Veiga, and E. Zilberman (2021), "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods", *Journal of Business & Economic Statistics*, *39*(1), 98–119.

Mitchell, J. and S. G. Hall (2005), "Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR 'Fan' Charts of Inflation", *Oxford Bulletin of Economics and Statistics*, *67*(s1), 995–1033.

Neyman, J. and E. Pearson (1933), "IX. On the Problem of the Most Efficient Tests of Statistical Hypotheses", *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *231*(694-706), 289–337.

Stock, J. H. and M. W. Watson (2002), "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business & Economic Statistics*, *20*(2), 147–162.