A General Procedure for Localising Strictly Proper Scoring Rules

Ramon de Punder* Cees Diks* Roger Laeven* Dick van Dijk†

July 10, 2022

Abstract

Forecasters are typically not equally interested in all possible realisations of a random variable under scrutiny. Financial risk managers, for instance, usually put relatively more weight on regions of extreme losses. In density forecast comparison, it is common practice to use strictly proper scoring rules to rank a collection of candidate predictive distributions. When focusing on a region of interest, however, weighted scoring rules obtained via conditioning are no longer strictly proper. We develop a general procedure for focusing, i.e., localising, scoring rules in a way that preserves their strict propriety.

A critical insight of our paper is that censoring observations outside the region of interest, as opposed to conditioning, retains just enough information about the original distribution to maintain strict propriety. Our procedure provides a myriad of strictly locally proper scoring rules beyond the censored likelihood score. We obtain a localised Neyman-Pearson result based on this scoring rule. Using a collection of popular scoring rules, including the Logarithmic, Spherical, Quadratic and Continuously Ranked Probability Score (CRPS), Monte Carlo simulations align with the intuition that censoring is power-enhancing, especially if the number of expected tail observations is small.

*University of Amsterdam, Tinbergen Institute
†Erasmus University Rotterdam, Tinbergen Institute

Contents

1 Introduction

Any forecasting application necessitates quantifying the relative performance of different forecasting methods. Gneiting and Raftery (2007) motivated the use of strictly proper scoring rules for this job, which has become the industry standard (Brehmer and Gneiting, 2020; Patton, 2020). The reason for this is that (strictly) proper scoring rules assign a score to the actual distribution that is (strictly) larger than the score of any other predictive distribution. Although strictly proper scoring rules admit point forecasts (e.g. mean squared error), we concentrate on their use in combination with predictive distributions and densities. Forecasts in the form of predictive distributions have gained interest in many different forecasting fields because they give a complete picture of the stochastic nature of the variable of interest (Dawid, 1984). At the same time, the specific characteristics of such applications encourage us to zoom in on certain parts of this picture, i.e. to localise the original scoring rule. In this paper, we present a general censoring-based procedure for localising scoring rules that preserves strict propriety. Our framework nests the censored likelihood (csl) scoring rule proposed by Diks et al. (2011) as a special case. We show that the uniformly most powerful test for a localised hypothesis test is based on this strictly locally proper scoring rule.

Motivating examples for local scoring rules can be found in different application areas. In risk management, for example, one is particularly interested in the left tail of the loss distribution, largely driven by regulatory capital requirements, formulated in terms of risk measures such as the Value-at-Risk (VaR) and Expected Shortfall (ES). See e.g. Diks et al. (2014), Kole et al. (2017), Opschoor et al. (2017) and Diks and Fang (2020) for applications. In macroeconomics, policymakers set – whether regulated by law or not – targets for central variables like inflation, nominal GDP and unemployment rates. For such clear targets, it is logical to zoom in on the part of the distribution around the target value. We refer to Gneiting and Ranjan (2011) and Iacopini et al. (2022) (and references therein) for interesting examples in macroeconomics.

The literature on focused scoring rules starts with the weighted likelihood score of

Amisano and Giacomini (2007), which simply multiplies the unweighted logarithmic scoring rule by a weight function. As independently observed by Diks et al. (2011) and Gneiting and Ranjan (2011), this procedure produces improper scoring rules because it favours distributions with more mass assigned to regions with higher weights, independent of the underlying distribution. As proper alternatives, Gneiting and Ranjan (2011) develop the weighted continuously ranked probability scoring (wCRPS) rule, while Diks et al. (2011) propose the conditional (cl) and csl scoring rule. Holzmann and Klar (2017a, Theorem 1) observe that the procedure of the cl scoring rule can be generalised to other scoring rules than the logarithmic scoring rule. They propose a general procedure for focusing regular scoring rules that applies the regular scoring rule to a weighted transformation of the original distribution. Their approach differs from ours by the suggested transformation of the original distribution: a conditional vis-à-vis censored distribution. The impact of this difference is that our censoring-based mechanism is the only one guaranteed to deliver strictly locally proper scoring rules. Interestingly, another route leading to the conditioning mechanism of Holzmann and Klar (2017a, Theorem 1) is to first generalise the weighted log-likelihood scoring rule proposed by Amisano and Giacomini (2007) and then apply a transformation coined *properisation* by Brehmer and Gneiting (2020, Theorem 1).

Our research also builds on the existing work on strictly proper scoring rules and their associated divergence measures. Although Gneiting and Raftery (2007) are responsible for the formal definition of strict propriety, scoring rules satisfying this property date back to at least the quadratic scoring (QS) rule proposed by Brier (1950). It is useful to know that this research area is dichotomous in the sense that much of the research prior to the rigorous treatment of general probability measures by Gneiting and Raftery (2007) has been conducted relative to discrete distributions on a finite outcome space, while more recent work more often follows the generality of Gneiting and Raftery (2007). For instance, the introduction of the LogS (Good, 1952; Toda, 1963) and spherical scoring (SphS) rule (Roby, 1964; Good, 1971), the initial generalisations of QS and SphS to the PowS_α and PsSphS_α families, and the axiomatic characterisations of the LogS, PowS_α and PsSphS_α rules provided by Shuford et al. (1966), Savage (1971), Selten (1998) and Jose (2009), are all presented in a discrete context. In our analysis, we work with the generalisations of the PowS_α and PsSphS_α families advocated by Gneiting and Raftery (2007) and Ovcharov (2018).

Moreover, the expected score differences of many scoring rules are recognised as well-known divergence measures, which reduce all together to the class of Bregman divergences (Bregman, 1967) when solely considering strictly proper scoring rules (Dawid, 2007; Gneiting and Raftery, 2007; Ovcharov, 2018; Painsky and Wornell, 2019). Consequently, concentrating the score divergences of strictly proper scoring rules excludes all f-divergences except the Kullback Leibler divergence (Kullback and Leibler, 1951), which is the unique intersection of the Bregman and f-divergence families. Due to its favourable properties (Liese and Vajda, 2006) the Kullback Leibler divergence has become the cornerstone in measuring the discrepancy between densities. For example, it is the divergence that is minimised in the maximum likelihood framework (Fisher, 1922), which bears optimal properties in the context of testing and estimation. Specifically, the likelihood ratio test is the most powerful test (Neyman and Pearson, 1933) and maximum likelihood estimators are unbiased estimators reaching the Cramér–Rao lower bound.

Pivotised sample equivalents of the expected score differences are fundamental in hypothesis tests about the relative performance of two candidate predictive distributions. In line with the weighted applications we have in mind, we localise the simple versus simple hypothesis of the Neyman-Pearson lemma into statements about the underlying distribution on the region of interest. By doing so, the hypothesis about the underlying distribution becomes a multiple versus multiple hypothesis, equivalent to the hypothesis studied by Holzmann and Klar (2016). Unlike them, we are still able to derive the uniformly most powerful test for this hypothesis. The test statistic of this test is given by a localised likelihood ratio, where the localisation is performed by censoring, and necessarily not by conditioning.

Power analyses based on localised scoring rules have more frequently been studied for the Giacomini and White (2006) test (Diks et al., 2011, 2014; Holzmann and Klar, 2016; Lerch et al., 2017). The null hypothesis of this test entails that the expected score difference between one candidate to the actual distribution is equivalent to the expected score difference between the other candidate and the actual distribution. A great advantage of this test is that all choices underlying the predictor, such as parameter uncertainty, can be seen as an integral part of the candidate, therefore also called prediction methods. For a strictly proper scoring rule, the null implies that both candidates are necessarily misspecified under the null, namely 'equally misspecified'. Yet, since which distributions are equally off from both candidates is determined by the scoring rule, this means that the null set of the GW test is a function of the selected candidates and the selected scoring rule, complicating comparisons between GW tests based on different scoring rules. To illustrate this interplay, we include a parametric example for which the conditional GW null set coincides with the full parameter space, whereas the censored GW null is a lower-dimensional subspace of the parameter space. We also compare the power properties of the GW test of the censored scoring rules with their conditional counterparts and other commonly used localised scoring rules like the wCRPS of Gneiting and Raftery (2007). In line with Diks et al. (2011), we find that censoring often leads to higher power.

The remainder of this paper is organised as follows. Section describes the fundamental concepts on which the subsequent chapters rely. Chapter 3 defines the generalised censored scoring rule and includes the assumption under which it shown to be strictly locally proper. This chapter also includes a randomisation procedure, called Z-Q-randomisation, equivalent to the generalised censored scoring rule. Chapter 4 is devoted to the Localised Neyman-Pearson lemma, for which Chapter 5 includes a variety of simulation examples. Chapter 6 concludes.

2 Theoretical framework

2.1 Regular scoring rules

Consider a random variable $Y : \Omega \to \mathcal{Y}$ from a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the measurable space $(\mathcal{Y}, \mathcal{G})$ and denote the actual distribution of Y on $(\mathcal{Y}, \mathcal{G})$ by P. The goal of a forecaster is to choose a distribution from a convex class of candidate distributions $\mathrm{F} \in \mathscr{P}$ on $(\mathcal{Y}, \mathcal{G})$, that minimises the *score divergence*

$$\mathbb{D}_{\mathrm{P}}(\mathrm{P}, \mathrm{F}; S) := \int_{\mathcal{Y}} S(\mathrm{P}, \cdot)\mathrm{dP} - \int_{\mathcal{Y}} S(\mathrm{F}, \cdot)\mathrm{dP},$$

based on a *strictly proper scoring rule* S, over $\mathscr{P}$ (Gneiting and Raftery, 2007). To explicitly rule out cases in which the score divergence between two distributions from $\mathscr{P}$ is undefined, e.g. $\mathbb{D}_{\mathrm{P}}(\mathrm{P}, \mathrm{F}; S) = -\infty + \infty$, we adopt the definition of a scoring rule proposed by Holzmann and Klar (2017a), included as Definition 1. Furthermore, if there exists a σ-finite measure μ on $(\mathcal{Y}, \mathcal{G})$ such that $\mathrm{F} \ll \mu, \forall \mathrm{F} \in \mathscr{P}$, we can alternatively define the scoring rule S relative to the induced class of μ-densities $\mathscr{p}$ by replacing the distributions by their corresponding μ-densities $f \in \mathscr{p}$.

Definition 1 (Scoring rule)**.** *A scoring rule is any extended real-valued* $(\bar{\mathbb{R}} \equiv [-\infty, \infty])$ *function* $S : \mathscr{P} \times \mathcal{Y} \to \bar{\mathbb{R}}$ *such that* $S(\mathrm{F}, \cdot)$ *is measurable with respect to* $\mathcal{G}$ *and quasi-integrable with respect to all* $\mathrm{P} \in \mathscr{P}$*, for all* $\mathrm{F} \in \mathscr{P}$*, and for which*

$$\int_{\mathcal{Y}} S(\mathrm{F}, \cdot)\mathrm{dP} < \infty \ \textit{and} \ \int_{\mathcal{Y}} S(\mathrm{P}, \cdot)\mathrm{dP} \in \mathbb{R}, \qquad \forall \mathrm{P}, \mathrm{F} \in \mathscr{P}.$$

Gneiting and Raftery (2007) stress that it is natural to exclusively consider strictly proper scoring rules to compare distributions from $\mathscr{P}$ since score divergences based on strictly proper scoring rules are P-uniquely minimised by the actual distribution. Here P-uniquely refers to the fact that any measure $\tilde{\mathrm{P}}$ that is P-equivalent to P, that is, $\tilde{\mathrm{P}} = \mathrm{P}$, P-a.s., leads to the same P-expected score as P. Hence, the uniqueness of the minimiser of the score divergence $\mathbb{D}_{\mathrm{P}}(\mathrm{P}, \mathrm{F}; S)$ is defined up to a class of measures that is P-equivalent to P. Bearing this in mind, we suppress this technicality in all that follows. This includes the formal definition of a strictly proper scoring rule adopted from Gneiting and Raftery (2007), given by Definition 2.

Definition 2 ((Strictly) proper scoring rule)**.** *A scoring rule $S : \mathscr{P} \times \mathcal{Y} \to \bar{\mathbb{R}}$ is proper relative to $\mathscr{P}$ if $\mathbb{D}_{\mathrm{P}}(\mathrm{P}, \mathrm{F}; S) \geq 0$, $\forall \mathrm{P}, \mathrm{F} \in \mathscr{P}$, and strictly proper relative to $\mathscr{P}$ if, additionally, $\mathbb{D}_{\mathrm{P}}(\mathrm{P}, \mathrm{F}; S) = 0$ iff $\mathrm{P} = \mathrm{F}$, $\forall \mathrm{P}, \mathrm{F} \in \mathscr{P}$.*

We are now able to make the relation between Bregman divergences and strictly proper scoring rules, mentioned in Section 1, more precise: A scoring rule is strictly proper if and only if its score divergence is a Bregman divergence.

2.2 Weighted scoring rules

In this research, we are particularly interested in *weighted scoring rules* (see Definition 3, adopted from Holzmann and Klar (2017a)), allowing forecasters to emphasise particular regions of the outcome space $\mathcal{Y}$ via a weight function $w \in \mathcal{W}$, defined as a map $w : \mathcal{Y} \to [0, 1]$. As mentioned in Section 1, emphasising parts of the outcome space, say the left tail, is reasonable if the objective of a forecaster is to produce reliable forecasts of a risk measure that is entirely based on the left tail. Indeed, the differences in scores between two candidates F and G in the right tail are completely irrelevant in such applications. This example motivates the use of scoring rules that give the same score to candidate distributions that are equivalent on $\{w > 0\}$, that is, the region of the outcome space on which the weight function is positive. Following Holzmann and Klar (2017a), we refer to weighted scoring rules satisfying this property, formalised in Definition 4, as *localising weighted scoring rules.*

Definition 3 (Weighted scoring rule)**.** *A weighted scoring rule is a map $S : \mathscr{P} \times \mathcal{Y} \times \mathcal{W} \to \bar{\mathbb{R}}$ such that $S(\cdot, \cdot; w)$ is a scoring rule for each $w \in \mathcal{W}$.*

Definition 4 (Localising weighted scoring rule)**.** *A weighted scoring rule $S : \mathscr{P} \times \mathcal{Y} \times$*

$\mathcal{W} \to \bar{\mathbb{R}}$ *is localising if for any* P, F $\in \mathscr{P}$, $w \in \mathcal{W}$,

$$\forall E \in \mathcal{G} : \mathrm{P}(\{w > 0\} \cap E) = \mathrm{F}(\{w > 0\} \cap E) \implies S(\mathrm{P}, y; w) = S(\mathrm{F}, y; w),\ \forall y \in \mathcal{Y}.$$

Since a localising weighted scoring rule is, by definition, a scoring rule for any given weight function $w \in \mathcal{W}$, it is again natural to look at the subset of localising weighted scoring rules for which the P-expected score is maximised by the actual distribution P. In other words, to only consider the class of localising proper weighted scoring rules defined in Definition 5.

Suppose that a weight function w is not strictly positive on $\mathcal{Y}$. In that case, a localising proper weighted scoring rule based on this weight function can never be strictly proper, since any distribution that is different from the actual distribution on $\{w = 0\}$ but equivalent to the actual distribution on $\{w > 0\}$ implies the same expected score. Consequently, the uniqueness of the maximiser of the P-expected score can only be achieved on $\{w > 0\}$. Similar to Holzmann and Klar (2017a), we call a localising proper weighted scoring rule that additionally satisfies this local uniqueness property *strictly locally proper* (see Definition 5).

Definition 5 ((Strictly) locally proper scoring rule)**.** *A weighted scoring rule* $S : \mathscr{P} \times \mathcal{Y} \times \mathcal{W} \to \bar{\mathbb{R}}$ *is locally proper relative to*$(\mathscr{P}, \mathcal{W})$ *if it is localising and* $S(\cdot, \cdot, w)$ *is proper for each* $w \in \mathcal{W}$*. Furthermore, it is strictly locally proper relative to* $(\mathscr{P}, \mathcal{W})$ *if additionally and* $\forall w \in \mathcal{W}$,

$$\mathrm{P}(\{w > 0\} \cap E) = \mathrm{F}(\{w > 0\} \cap E), \forall E \in \mathcal{G} \iff \int_{\mathcal{Y}} S(\mathrm{P}, y; w)\mathrm{P}(\mathrm{d}y) = \int_{\mathcal{Y}} S(\mathrm{F}, y; w)\mathrm{P}(\mathrm{d}y).$$

We aim to develop a general procedure that maps an unweighted scoring rule S that is strictly proper relative to $\mathscr{P}$ and a class of weight functions $\mathcal{W}$ into a class of scoring rules that is strictly locally proper relative to $(\mathscr{P}, \mathcal{W})$. To illustrate how one could develop such a procedure, Example 1 summarises the procedure leading to the conditional scoring rule $S^{\sharp}_{w}$ proposed by Holzmann and Klar (2017a), nesting the conditional likelihood score of Diks et al. (2011) as a special case. Unfortunately, however, this procedure delivers proportionally locally proper instead of strictly locally proper scoring rules (Holzmann and Klar, 2017a).

Definition 6 (Weighted kernel)**.** *The weighted kernel of a distribution* $\mathrm{F} \in \mathscr{P}$ *is,* $\forall w \in \mathcal{W}$*, defined as*

$$\mathrm{dF}_w = w\mathrm{dF}.$$

Example 1 (Conditional scoring rule)**.** *One way of weighting a regular scoring rule is to apply the regular scoring rule to a weighted distribution. The conditional scoring rule proposed by Holzmann and Klar (2017a) follows this recipe, based on the specific weighted distribution*

$$\mathrm{dF}_w^\sharp = \frac{1}{1 - \bar{\mathrm{F}}_w}\mathrm{dF}_w, \qquad \mathrm{F} \in \mathscr{P}, \; w \in \mathcal{W}, \tag{1}$$

where $\bar{F}_w := \int_{\mathcal{Y}}(1-w)\mathrm{dF}$*,* F_w *is defined in Definition 6 and it is assume that* $\int_{\mathcal{Y}} w\mathrm{dF} > 0, \forall \mathrm{F} \in \mathscr{P} \; w \in \mathcal{W}$*. For the specific choice* $w = \mathbb{1}_A$*,* $A \subseteq \mathcal{Y}$*, the weighted distribution* $\mathrm{F}_w^\sharp$ *reduces to the conditional distribution of* F *on* A*.*

Holzmann and Klar (2017a) use the weighted distribution $\mathrm{F}_w^\sharp$ *to define the conditional scoring rule*

$$S_w^\sharp(\mathrm{F}, y) := w(y)S(\mathrm{F}_w^\sharp, y),$$

where the multiplication with the weight function $w \propto \frac{\mathrm{dF}_w^\sharp}{\mathrm{dF}}$ *is needed to arrive at a localising proper weighted scoring rule relative to* $(\mathscr{P}, \mathcal{W})$*. See Holzmann and Klar (2017b) for a detailed proof of this result.*

From the definition of the weighted distribution $\mathrm{F}_w^\sharp$ *in Equation (1), one can easily see that a weighted scoring rule based on this transformation is not strictly locally proper. After all, any distribution* $\tilde{\mathrm{F}}$ *that is proportional to* F *on* $\{w > 0\}$*, that is,* $\mathrm{d}\tilde{\mathrm{F}} = c\mathrm{dF}$ *for some constant* $c > 0$*, implies the same weighted distribution* $\mathrm{F}_w^\sharp$*. As a* *consequence, there exists a continuum of distributions different from the true measure* P *on* $\{w > 0\}$ *(beyond those mentioned above* P*-.a.s. equivalence class) that leads to the same scoring rule. For example, this means that we cannot discriminate between the actual VaR and the VaR based on a measure proportional to* P *(and hence potentially largely different from the actual VaR) if we use the conditional scoring rule – score divergence.*

2.3 Commonly used scoring rules

Weighted scoring rules are transformations of regular scoring rules. In applications, we focus on the Logarithmic scoring rule (LogS), Power family of scoring rules (PowS_α), the PseudoSpherical family of scoring rules (PsSphS_α) and the Continuously Ranked Probability Score (CRPS), all of which are strictly proper. The inclusion of the PowS_α and PsSphS_α families, which include the LogS scoring rule as a limiting case for $\alpha \downarrow 1$, is partly due to the connection with the expected utility maximisation problems described by Jose et al. (2008). After all, the duality with specific investment problems based on the one-parameter Hyperbolic Absolute Risk Aversion (HARA) utility function family, generated by the absolute risk tolerance function $\tau_\alpha(x) = \beta + \alpha x$, with $\beta = 1$ (see e.g. Merton (1971, p. 389)), gives α its interpretation as a risk tolerance parameter.

It is also interesting to examine the relation between the scoring rules under consideration and their associated Bregman divergences. For this purpose, recall that the so-called *separable Bregman divergences*

$$\mathbb{D}_\phi : \mathcal{P}(\mathcal{Y}, \mathcal{G}, \mu)^2 \to \mathbb{R}^+, \qquad \mathbb{D}_\phi(p, q) = \int_{\mathcal{Y}} d_\phi(p(y), q(y))\mathrm{d}\mu(y), \tag{2}$$

arise from the subclass of score divergences based on strictly proper scoring rules of the form

$$S_\phi : \mathcal{P}(\mathcal{Y}, \mathcal{G}, \mu) \times \mathcal{Y} \to \mathbb{R}, \qquad S_\phi(p, y) = \phi'(p(y)) - \int_{\mathcal{Y}} \phi'(p(y))p(y) - \phi(p(y))\mathrm{d}\mu(y). \tag{3}$$

As revealed by their generator functions $\phi(t)$ in Table 1, the LogS and PowS_α rules are clearly of this form. Consequently, the score divergences based on these two scoring rules, being the Kullback Leibler divergence and the $L^2(\mathcal{Y}, \mathcal{G}, \mu)$-distance for $\alpha = 2$, are both separable Bregman divergences. Of course, we are not limited to these specific choices for $\phi(t)$. Rather, Eq. (3) can be seen as a generator for many other strictly proper scoring rules. However, not every Bregman divergence is separable and hence not every strictly proper scoring rule can be generated by a specific choice for $\phi(t)$ in Eq. (3). For example, there exists no strictly convex function $\phi(t)$ for which Eq. (3) yields the PsSphS_α rule (Good, 1971; Gneiting and Raftery, 2007; Jose, 2009).

3 Localising scoring rules by censoring

3.1 Censoring

As illustrated by Example 1, the lacking strictness in the local propriety of the conditional scoring rule is inherent to the definition of the weighted distribution on which it is built. Inspired by the csl rule of Diks et al. (2011) and the explicit definition of a censored density by Gatarek et al. (2013), we propose to alternatively consider a weighted scoring rule that departs from a *censored distribution*

$$\mathrm{dF}_w^\flat = \mathrm{dF}_w + \bar{\mathrm{F}}_w \mathrm{d}\delta_*, \qquad w \in \mathcal{W},\ \mathrm{F} \in \mathscr{P}, \tag{4}$$

which is defined relative to the extended measurable space $(\mathcal{Y}^*, \mathcal{G}^*)$, where $\mathcal{Y}^* = \mathcal{Y} \cup *$ and $\mathcal{G}^* = \sigma(\{\mathcal{G}, *\})$, that is, the smallest σ-algebra containing the collection $\{\mathcal{G}, *\}$. Furthermore, δ_* denotes the Dirac measure at $*$, i.e. $\delta_*(E) = \mathbb{1}_E(*)$.

The censored distribution in Equation (4) generalises the standard notion of censoring on the real line (Bernoulli, 1760; Tobin, 1958) in which values above (or below) a certain threshold r are known to be unobservable. Indeed, for $w(y) = \mathbb{1}_{y \le r}$, the censored distribution is nothing but the distribution of the random variable Y for which the outcomes above r are made unobserved, that is, replaced by $*$. So, in this specifc

Table 1: Summary main characteristics density-based scoring rules

Name	Logarithmic	Power family	PseudoSpherical family
Definition $S(p,Y)$	$\mathrm{LogS}(p,Y)=\log p(Y)$	$\mathrm{PowS}_\alpha(p,Y)=\alpha p(Y)^{\alpha-1}-(\alpha-1)\|p\|_\alpha^\alpha$	$\mathrm{PsSphS}_\alpha(p,Y)=\frac{p(Y)^{\alpha-1}}{\|p\|_\alpha^{\alpha-1}}$
Special cases		$\mathrm{QS}(p,Y)=\mathrm{PowS}_2(p,Y)$	$\mathrm{SphS}(p,Y)=\mathrm{PsSphS}_2(p,Y)$
		$\mathrm{LogS}(p,Y)=\lim_{\alpha\downarrow 1}\mathrm{PowS}_\alpha(p,Y)$	$\mathrm{LogS}(p,Y)=\lim_{\alpha\downarrow 1}\mathrm{PsSphS}_\alpha(p,Y)$
Negative entropy $\mathbb{H}(p;S)$	$\mathbb{E}_p\log p$	$\|p\|_\alpha^\alpha$	$\|p\|_\alpha$
Divergence $\mathbb{D}_p(p,f;S)$	$\mathbb{E}_p\log\left(\frac{p}{f}\right)$	$\|p\|_\alpha^\alpha-\alpha\int f^{\alpha-1}(p-f)\mathrm{d}\mu-\|f\|_\alpha^\alpha$	$\|p\|_\alpha-\frac{\int pf^{\alpha-1}\mathrm{d}\mu}{\|f\|_\alpha^{\alpha-1}}$
Divergence for $\alpha=2$	–	$\|p-f\|_2^2$	$\|p\|_2\left(1-C(p,f)\right)$
Strictly proper class	$\mathcal{L}^1$	$\mathcal{L}^\alpha$	$\mathcal{L}^\alpha$
Bregman generator $\phi(t)$	$t\log t$	t^α	–

Note: $C(p,f)=\frac{\int pf\mathrm{d}\mu}{\sqrt{\int p\mathrm{d}\mu\int f\mathrm{d}\mu}}$ and $\mathbb{D}_p(p,f;S)$ denote the cosine similarity and the score diverence between p and f, respectively. $\mathbb{H}(p)$ denotes the negative entropy and and the PowS_α and PsSphS_α families is restricted to $\alpha>1$. Furthermore, $\mathcal{L}^\alpha(\mathcal{Y},\mathcal{G},\mu)$ denotes the space for which the L^α-norm is finite, where μ is measure relative to which the densities p and f are defined, i.e. $\frac{\mathrm{dP}}{\mathrm{d}\mu}$, with $\mathrm{P}\ll\mu$. An axiomatic characterisation of the LogS, QS and SphS rule is provided by Shuford et al. (1966), Savage (1971) and Selten (1998), and Jose (2009), respectively.

case, $\mathrm{F}^\flat_{\mathbb{1}_{y\le r}}$ is just the distribution of

$$Y^\flat_{\mathbb{1}_{y\le r}} = \begin{cases} Y, & Y \le r, \\ *, & \text{otherwise.} \end{cases}$$

The general case of the censored distribution displayed in (4) also bears a sound interpretation, but this is deferred to Section 3.3.

From a conceptual point of view, this example illustrates that the outcome space is extended along the censoring process, substituting right tail observations by $*$. Mathematically, however, it is then preferred to also already define the non-focused distribution F with respect to the extended measurable space in order to have a legitimately defined change of measure, described by Equation (4). Therefore, we are more formally considering the change of measure $\mathrm{F}^* \mapsto \mathrm{F}^\flat_w$ on $(\mathcal{Y}^*, \mathcal{G}^*)$, where $\mathrm{F}^*(E) = \mathrm{F}(E\backslash\{*\})$, if $* \in E$ and $\mathrm{F}^*(E) = \mathrm{F}(E)$, otherwise, $\forall E \in \mathcal{G}^*$. Combining both perspectives, we view the censored measure as a result of two steps. In the first step, the original measure F on $(\mathcal{Y}, \mathcal{G})$ is extended to F^* on $(\mathcal{Y}^*, \mathcal{G}^*)$. Relative to the extended measurable space $(\mathcal{Y}^*, \mathcal{G}^*)$, we subsequently apply the change of measure given by Equation (4), with, strictly speaking, F and w replaced by F^* and w^*, respectively. For the extended weight function w^*, one can take any value for $w^*(*)$ since $\mathrm{F}^*(*) = 0, \forall \mathrm{F}^* \in \mathscr{P}^*$. To keep the notation uncluttered, we will not be explicit about the differences between the extended and non-extended measures and weight function (classes).

A closer look at the definition of the censored distribution in Equation (4) affirms why a scoring rule based on the censored distribution in Equation (4) intuitively solves the lack of strictness faced by scoring rules departing from the conditional distribution in Equation (1). Put simply, the censored distribution $\mathrm{F}^\flat_w$ preserves the one-to-one connection with the original distribution F relative to $\mathcal{G}$ by normalising the weighted kernel F_w through addition rather than division. The relevant consequence of this is that the *censored scoring rule* defined in Definition 7, which applies the regular scoring rule S to the censored distribution $\mathrm{F}^\flat_w$, is strictly locally proper relative to $(\mathscr{P}, \mathcal{W})$.

This is the content of Theorem 1, for which Appendix A.1 details a proof.

Definition 7 (Censored scoring rule)**.** *Let $S : \mathscr{P}^\flat \times \mathcal{Y} \to \bar{\mathbb{R}}$, $\mathscr{P}^\flat \equiv \{\mathrm{F}^\flat_w, \mathrm{F} \in \mathscr{P}, w \in \mathcal{W}\}$, denote a scoring rule. Then, the corresponding censored scoring rule is given by the map $S^\flat : \mathscr{P} \times \mathcal{Y} \times \mathcal{W} \to \bar{\mathbb{R}}$,*

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

where the censored distribution $\mathrm{F}^\flat_w$ is given by Equation (4).

Theorem 1. *The censored scoring rule $S^\flat$ in Definition 7 is strictly locally proper relative to $(\mathscr{P}, \mathcal{W})$ if the regular scoring rule S in Definition 7 is strictly proper relative to $\mathscr{P}^\flat$.*

The censored scoring rule in Definition 7 is obtained similarly as the conditional scoring rule depicted in Example 1, in the sense that both scoring rules apply the original scoring rule to a focused version of the original distribution. Although the censored measure is undoubtedly preferred when solely comparing the difference in proportional and strict local propriety implied by conditional and censored measure, respectively, this advantage of the censored distribution comes at the price of putting a restriction on the class of regular scoring rules that can be used for this focusing method. In particular, Definition 7 requires regular scoring rules S to be well-defined for distributions of continuous-discrete type, ruling out regular scoring rules like the CRPS. In Section 3.2, we address this issue by considering a further generalisation of the censored distribution.

Moreover, if the regular scoring rule does only depend on the distribution F through y, in which case the scoring rule is called *local*, the censored scoring rule becomes F-equivalent to

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

which is a weighted average of the regular score of the weighted kernel and the discrete probability $\bar{\mathrm{F}}_w$. For the indicator choice $w(y) = \mathbb{1}_{y \leq r}$ on $\mathcal{Y} = \mathbb{R}$, this further implies that observations in the left tail receive the original score $S(\mathrm{F}, y)$, while observations in the right tail contribute with the regular score of the coverage probability of the right tail.

The most celebrated local scoring rule is without a doubt the Logarithmic scoring rule. Similar to other (non-local) scoring rules like the PsSphS_α and PowS_α family of scoring rules, introduced in Section 2.3, this scoring rule is a function of the density. Evidently, the censored scoring rule in Definition 7 also applies to densities, after replacing the class of probability measures $\mathscr{P}$ by the associated class of μ-densities $\mathscr{p} = \{f, f = \frac{\mathrm{dF}}{\mathrm{d}\mu}, \mathrm{F} \in \mathscr{P}\}$, provided that $\mathrm{F} \ll \mu, \forall \mathrm{F} \in \mathscr{P}$. Likewise, the censored distribution given by Equation (4) should be replaced by the censored density

$$f_w^\flat = wf\mathbb{1}_{y \neq *} + \bar{F}_w \mathbb{1}_{y=*}, \qquad w \in \mathcal{W},\ f \in \mathscr{p}, \tag{5}$$

which is the $(\mu + \delta_*)$-density of $\mathrm{F}_w^\flat$ if $\mathrm{F} \ll \mu$ (see Appendix A.3 for a proof).

We conclude this section with some concrete examples of censored scoring rules. In Section 2.3, we included references for the result that the LogS, PsSphS_α and PowS_α family of scoring rules are strictly proper relative to the class of densities $\mathscr{p}_\alpha$ such that $\|f\|_\alpha < \infty$, with $\alpha = 1$ for LogS. As one can easily verify that $\|f_w^\flat\|_\alpha^\alpha \leq 1 + \|f\|_\alpha^\alpha$, $\forall w \in \mathcal{W}$, it additionally follows that these families are also strictly proper relative to $\mathscr{p}_\alpha^\flat$. Hence Theorem 1 applies, from which it follows that their censored counterparts displayed in Table 2 are strictly locally proper relative to $\mathscr{p}_\alpha$. Comparing the censored and conditioned versions of the rules, we notice that the censored variants have an extra $\bar{F}_w$-dependent second term, preserving more information of the original distribution.

For the conditional PsSphS_α, Table 2 also shows that the normalising constant $1 - \bar{F}_w$ cancels, so that, formulawise, $\mathrm{PsSphS}^\sharp_{\alpha,w}(f, y) = w(y)\mathrm{PsSphS}_\alpha(f_w, y)$. This is somewhat alarming, since $\lim_{\alpha \downarrow 1} \frac{1}{\alpha - 1}\mathrm{PsSphS}_\alpha(f, y) = \log f(y)$ by Equation (6) and $w(y)\mathrm{LogS}(f, y)$ is equivalent to the weighted likelihood score proposed by Amisano and Giacomini (2007), which is improper (Diks et al., 2011). As illustrated by Appendix B.2, this is no cause for concern as the aforementioned limit is a simpliciation of $\log \frac{f(y)}{\|f\|_1}$ and $\|f_w\|_1 = 1 - \bar{F}_w$. For the conditional rules, the linearity of limits straightforwardly implies that the limit result in Equation (6) survives conditioning. Appendix B.2 and B.3 show that the same is true for the censored PowS_α and PsSphS_α family, respectively. Consequently,

$$\lim_{\alpha \downarrow 1} \frac{1}{\alpha - 1} \mathrm{PsSphS}^x_{\alpha,w}(f, y) = \lim_{\alpha \downarrow 1} \frac{1}{\alpha - 1} \mathrm{PowS}^x_{\alpha,w}(f, y) = \mathrm{LogS}^x(f, y), \qquad \forall x \in \{\emptyset, \sharp, \flat\}, \tag{6}$$

where $x = \emptyset$ refers to a regular scoring rule (which does not depend on w).

Table 2: Regular, conditional and censored density-based scoring rules

S	$S(f,y)$	$S^\sharp_w(f,y)$	$S^\flat_w(f,y)$
LogS	$\log f(y)$	$w(y) \log\left(\frac{f(y)}{1-\bar{F}_w}\right)$	$w(y) \log f(y) + \big(1 - w(y)\big) \log \bar{F}_w$
PsSphS_α	$\frac{f(y)^{\alpha-1}}{\|f\|_\alpha^{\alpha-1}}$	$w(y) \frac{f_w(y)^{\alpha-1}}{\|f_w\|_\alpha^{\alpha-1}}$	$\frac{w(y) f_w(y)^{\alpha-1} + (1-w(y)) \bar{F}_w^{\alpha-1}}{\left(\|f_w\|_\alpha^\alpha + \bar{F}_w^\alpha\right)^{\frac{\alpha-1}{\alpha}}}$
PowS_α	$\alpha f(y)^{\alpha-1}$	$w(y)\Bigg(\alpha \left(\frac{f_w(y)}{1-\bar{F}_w}\right)^{\alpha-1}$	$w(y)\alpha f_w(y)^{\alpha-1} + \big(1 - w(y)\big)\alpha \bar{F}_w^{\alpha-1}$
	$-(\alpha-1)\|f\|_\alpha^\alpha$	$-(\alpha-1) \left\| \frac{f_w(y)}{1-\bar{F}_w} \right\|_\alpha^\alpha \Bigg)$	$-(\alpha-1)\left(\|f_w\|_\alpha^\alpha + \bar{F}_w^\alpha\right)$

Regular, conditional ($\sharp$) and censored ($\flat$) scoring rules based on the three density-based scoring rules introduced in Section 2.3. The derivations corresponding with the LogS, PsSphS_α and PowS_α rules are deferred to Appendix B.1, B.2 and B.3, respectively.

3.2 Generalised censored scoring rule

Since not every regular scoring rule S is well-defined relative to the censored measure in Equation (4), the censored scoring rule in Definition 7 puts a restriction on the class of admitted regular scoring rules $\mathcal{S}$. For instance, the CRPS is not compatible with the censored distribution in Equation (4) because the CRPS is solely defined for distributions that are absolutely continuous with respect to the Lebesgue measure λ on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ denotes the Borel σ-algebra on $\mathbb{R}$. One way to alleviate the restriction on $\mathcal{S}$ such that CRPS $\in \mathcal{S}$, is to flatten the shape of the distribution outside the region of interest such that the censored distribution remains absolutely continuous with respect to the Lebesgue measure. As will be clear from Theorem 2 below, a valid alternative to the standard notion of censoring is to 'store' the probability $\bar{F}_w$ on some interval on which the weight function is positive somewhere λ-a.e., instead of at $*$.

More precisely, replacing the Dirac distribution δ_* in Equation (4) by a $\mathrm{Unif}(r_w, r_w+\theta)$, $\theta > 0$, distribution such that $\lambda\big(\{w = 0\} \cap (r_w, r_w+\theta)\big) > 0$ does not affect the strict local propriety of the *generalised censored scoring rule* in Definition 7 based on this uniform choice for the *nuisance distribution* H. It is worth emphasising that the term 'generalised' is in place here since Definition 8 reduces to Definition 7 when choosing δ_* as nuisance distribution. As desired, the class of censored measures $\mathscr{P}^\flat = \{\mathrm{F}^\flat_{w,\mathrm{U}}, \mathrm{F} \in \mathscr{P}, w \in \mathcal{W}_\lambda\}$, where

$$\mathrm{dF}^\flat_{w,\mathrm{U}} = \mathrm{dF}_w + \bar{F}_w \mathrm{dU}, \qquad \mathrm{dU} = \frac{1}{\theta}\mathbb{1}_{(r_w, r_w+\theta)}\mathrm{d}\lambda, \tag{7}$$

and $\mathcal{W}_\lambda = \{w \in \mathcal{W} : \lambda(\{w = 0\}) > 0\}$ is now such that the CRPS is added to the class of admitted scoring rules $\mathcal{S}$. It is important to note, however, that the expansion of $\mathcal{S}$ involves a restriction on the class of admitted weight functions from $\mathcal{W}$ to $\mathcal{W}_\lambda$.This exchange between the restriction on $\mathcal{S}$ and $\mathcal{W}$ generalises to Assumption 1 of Theorem 2.

According to Theorem 2, for which Appendix A.2 includes a proof, we can choose any other nuisance distribution than δ_* and $\mathrm{Unif}(r_w, r_w + \theta)$ without losing strict local propriety, as long as the regular scoring rule remains well-defined and strictly proper relative to the implied class of censored distributions and the choice of nuisance distribution is compatible with the class of weight functions under consideration. We have illustrated the latter condition by pointing out that not all scoring rules are compatible with the Dirac choice like not all weight functions can be combined with the Uniform choice. Furthermore, the intuition behind the interplay between the choice of the nuisance distribution H and weight function w is that the original distribution can be inferred from the censored distribution on $\{w > 0\}$ if H is known. Hence, Assumption 1 requires the existence of an event E such that H reveals itself. For the sake of clarity, let us explicitly note that $S^\flat$ being strictly locally proper relative to a triplet $(\mathscr{P}, \mathcal{W}, \mathscr{H})$ means that $S^\flat_{\cdot,\mathrm{H}}$ is strictly locally proper relative to $(\mathscr{P}, \mathcal{W})$, $\forall \mathrm{H} \in \mathscr{H}$.

Definition 8 (Generalised censored scoring rule)**.** *Let $S : \mathscr{P}^\flat \times \mathcal{Y} \to \bar{\mathbb{R}}$ denote a scoring rule. The associated generalised censored scoring rule is given by the map $S^\flat : \mathscr{P} \times \mathcal{Y} \times \mathcal{W} \times \mathscr{H} \to \bar{\mathbb{R}}$,*

$$S^\flat_{w,\mathrm{H}}(\mathrm{F}, y) = w(y) S(\mathrm{F}^\flat_{w,\mathrm{H}}, y) + \big(1 - w(y)\big) \int_{\mathcal{Y}} S(\mathrm{F}^\flat_{w,\mathrm{H}}, q) \mathrm{H}(\mathrm{d}q), \quad \mathrm{dF}^\flat_{w,\mathrm{H}} = \mathrm{dF}_w + \bar{F}_w \mathrm{dH},$$

where $\mathrm{F}^\flat_{w,\mathrm{H}}$ is referred to as the generalised censored distribution of F*.*

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

Theorem 2. *Suppose that the regular scoring rule S in Definition 8 is strictly proper relative $(\mathscr{P}, \mathcal{W})$. Additionally assume that $\mathcal{W}$ and $\mathscr{H}$ are such that Assumption 1 is satisfied. Then, the generalised censored scoring rule $S^\flat$ in Definition 8 is strictly locally proper relative to $(\mathscr{P}, \mathcal{W}, \mathscr{H})$.*

For a class of weight functions $\mathcal{W}_\lambda$ and a class of λ-dominated distributions $\mathscr{P}_\lambda$, Assumption 1 permits a broad spectrum of nuisance distributions. For example, any Dirac measure δ_c, $c \in \mathbb{R}$, on $(\mathbb{R}, \mathcal{B})$ is allowed, since then $\mathrm{F}_w(c) = 0$, while $\mathrm{H}(c) = \delta_c(c) = 1$, independent of whether c is an element of $\{w = 0\}$ or not. Other examples include all continuous distributions that have (part of their) support on $\{w = 0\}$, as F_w is trivially zero on $\{w = 0\}$. The left-hand side panel of Table 3 presents the generalised censored scoring rules based on a density h satisfying the latter condition. When comparing the results with the censored scoring rules in Table 3, it is notable that the obtained $\mathrm{LogS}^\flat$ rule is independent of the choice of h. Due to their dependence of the α-norm, this invariance result only holds within a further restricted class of densities for the other two scoring rules, that is, for the normalised class of densities $\tilde{h} = h/\|h\|_\alpha$. The right-hand side panel of Table 3 lists the results for the specific choice $h = u$, where u is the density of a $\text{Unif}(r_w, r_w + \theta)$ distribution introduced above. Since the α-norm of a $\text{Unif}(r_w, r_w + 1)$ is one for all α, it it is not unforeseen that the generalised censored scoring rules coincide with the censored scoring rules in Table 2 if $\theta = 1$. For other values of $\theta > 0$, the parameter θ tunes the influence of the probability $\bar{F}_w$. In particular, the role of $\bar{F}_w$ diminishes for $\theta < 1$, while it intensifies for $\theta > 1$.

Table 3: Real valued examples of density-based generalised censored scoring rules

S	$S^{\flat}_{w,u}(f, y)$	$S^{\flat}_{w,h}(f, y)$
LogS	$w(y) \log f(y) + \big(1 - w(y)\big) \log \bar{F}_w$	$w(y) \log f(y) + \big(1 - w(y)\big) \log \bar{F}_w$
PsSphS_α	$\frac{w(y) f_w(y)^{\alpha-1} + \big(1-w(y)\big) \bar{F}_w^{\alpha-1} \theta^{1-\alpha}}{\big(\lVert f_w \rVert_\alpha^\alpha + \bar{F}_w^\alpha \theta^{1-\alpha}\big)^{\frac{\alpha-1}{\alpha}}}$	$\frac{w(y) f_w(y)^{\alpha-1} + \big(1-w(y)\big) \bar{F}_w^{\alpha-1} \lVert h \rVert_\alpha^\alpha}{\big(\lVert f_w \rVert_\alpha^\alpha + \bar{F}_w^\alpha \lVert h \rVert_\alpha^\alpha\big)^{\frac{\alpha-1}{\alpha}}}$
PowS_α	$w(y) \alpha f_w(y)^{\alpha-1} + \big(1 - w(y)\big) \alpha \bar{F}_w^{\alpha-1} \theta^{1-\alpha}$ $-(\alpha - 1) \big(\lVert f_w \rVert_\alpha^\alpha + \bar{F}_w^\alpha \theta^{1-\alpha}\big)$	$w(y) \alpha f_w(y)^{\alpha-1} + \big(1 - w(y)\big) \alpha \bar{F}_w^{\alpha-1} \lVert h \rVert_\alpha^\alpha$ $-(\alpha - 1) \big(\lVert f_w \rVert_\alpha^\alpha + \bar{F}_w^\alpha \lVert h \rVert_\alpha^\alpha\big)$
S	$S^{\flat}_{w,\mathrm{U}}(F, y)$	$S^{\flat}_{w,\mathrm{H}}(F, y)$
CRPS	$\int_{-\infty}^{r} \big(F_w(s) - w(y) \mathbb{1}_{(-\infty,s]}(y)\big)^2 \mathrm{d}s$ $+\frac{\theta}{3} \big(\bar{F}_w - (1 - w(y))\big)^2$	$\int_{-\infty}^{r} \big(F_w(s) - w(y) \mathbb{1}_{(-\infty,s]}(y)\big)^2 \mathrm{d}s$ $+\lVert H - 1 \rVert_2^2 \big(\bar{F}_w - (1 - w(y))\big)^2$

Generalised censored scoring rules based on the LogS, PsSphS_α, PowS_α and CRPS regular scoring rules (see Table 2 for the definitions) for two choices of the nuisance density (distribution) h (H). The norm $\lVert H - 1 \rVert_2^2$ is defined relative to support of H. The left panel uses a uniform distribution on $(r, r + \theta) \subseteq \{w = 0\}$, while the right handside panel departs from a density (distribution) h (H) which support is also a subset $\{w = 0\}$. Accordingly, the weight function $w \in \mathcal{W}_\theta \subseteq \mathcal{W}_\lambda$ is in both panels assumed to be zero on a subset of $\mathbb{R}$ with positive length. For the CRPS it is additionally assumed that $\{w > 0\} = (-\infty, r]$ and $\{w = 0\} = (r, \infty)$. The derivations corresponding with the LogS, PsSphS_α, PowS_α and CRPS rules are deferred to Appendix C.1, C.2, C.3 and C.4, respectively. Appendix C.5 additionally includes a censored CRPS example for the compact indicator weight function $w(y) = \mathbb{1}_{[a,b]}(y)$.

Table 3 also includes the generalised censored CRPS implied by Definition 8. Again, there exists a choice for the nuisance distribution for which the generalised censored scoring rule is functionally independent of this distribution, e.g. the Uniform choice with $\theta = 3$. The difference in interpretation between the latter $\text{CRPS}^{\flat}_{w,\mathrm{U}_3}$ rule and the twCRPS rule proposed by Gneiting and Ranjan (2011) is striking (see Table 4 for an overview of documented localised CRPS rules). Viewing the CRPS as an unweighted average of Brier probability scores for the probability forecasts $F(s)$ of the events $y \leq s$, the twCRPS is nothing but a weighted average of the same, strictly proper, individual

Brier scores. In contrast, the $\mathrm{CRPS}^{\flat}_{w,\mathrm{U}_3}$ rule applies the transformation implied by the weight function to the distribution function itself. For the indicator weight function $w(y) = \mathbb{1}_{(-\infty,r]}$ the consequence of the generalised censoring approach is that the scoring rule has an extra term compared to the twCRPS, which is the additional Brier score of the discrete probability $\bar{F}_w$ of the event $y > r$. Notably, this additional term is perfectly in line with the standard notion of censoring, in which we are no longer interested in the shape of distribution outside the region of interest, while still caring about the coverage probability of being in the region of interest.

Similarly, for the generality of weight functions introduced in Table 3, we note that the continuous and discrete component of the censored scoring rule extend naturally to more general weight functions. In particular, the weighted kernel F_w, which is now a weighted version of F on $\{w > 0\}$ rather than just F on $\{w > 0\}$, is now being compared with an indicator function that is weighted accordingly. Recalling that $\bar{F}_w = \int(1-w)\mathrm{dF}$, the generalisation of the Brier score based on the indicator weight function is also natural. Moreover, how the $\mathrm{CRPS}^{\flat}_{w,\mathrm{U}_3}$ rule splits into a continuous and discrete component is comparable to the censored likelihood score proposed by Diks et al. (2011), famed due to its favourable power properties. Hence, it is not unreasonable to conjecture that the censored CRPS also bears preferable power properties relative to the twCRPS.

Table 4: Localised CRPS variants

S_w	$S_w(F, y)$
$\mathrm{CRPS}^{\flat}_{w,\mathrm{U}_3}$	$\int_{-\infty}^{r} \big(F_w(s) - w(y)\mathbb{1}_{(-\infty,s]}(y)\big)^2 \mathrm{d}s + \big(\bar{F}_w - (1-w(y))\big)^2$
twCRPS_w	$\int_{-\infty}^{\infty} w(s)\big(F(s) - \mathbb{1}_{(-\infty,s]}(y)\big)^2 \mathrm{d}s$
$\mathrm{CRPS}^{\sharp}_w$	$w(y)\int_{-\infty}^{r} \left(\frac{F_w(s)}{1-\bar{F}_w} - \mathbb{1}_{(-\infty,s]}(y)\right)^2 \mathrm{d}s$
wsCRPS_w	$w(y)\int_{-\infty}^{r} \left(\frac{F_w(s)}{1-\bar{F}_w} - \mathbb{1}_{(-\infty,s]}(y)\right)^2 \mathrm{d}s + w(y)\bar{F}_w^2 + \big(1-w(y)\big)\big(1-\bar{F}_w\big)^2$

The $\mathrm{CRPS}^{\flat}_{w,\mathrm{U}_3}$ is a special case of Definition 8. For the twCRPS_w, $\mathrm{CRPS}^{\sharp}_w$ and $\mathrm{wsCRPS}^{\sharp}_w$ rules, we refer to Gneiting and Ranjan (2011) (Holzmann and Klar, 2017a, Theorem 1) and (Holzmann and Klar, 2017a, Theorem 3), respectively.

3.3 Z,Q-randomisation

It is also possible to formulate an explicit transformation of the random variable Y that leads to the censored random variable $Y^\flat_{w,\mathrm{H}}$ in Definition 8. For this transformation, we introduce – on top of the independent random variable Q with distribution H – the auxiliary random variable $Z|Y = y \sim \mathrm{BIN}(1, w(y))$. The specific transformation looks as follows

$$Y^\flat_{\mathrm{H},w} = \phi_{w,\mathrm{H}}(Y, Z, Q) = \begin{cases} Y, & \text{if } Z = 1 \\ Q, & \text{if } Z = 0 \end{cases}.$$

To keep the notation uncluttered, we use the censored density of Lemma **??** to discuss this equivalence briefly. Using the auxiliary variable Z, we first note that it is unchallenging to derive the bivariate density of $(Y^\flat_{\mathrm{H},w}, Z)$, which is of continuous-discrete type. We arrive at the following specification

$$\begin{aligned} f^\flat_{w,h}(y, Z = 0) &= \mathbb{P}(Z = 0)\frac{\mathrm{dH}}{\mathrm{d}\mu^\flat}(y) = \bar{F}_w \frac{\mathrm{dH}}{\mathrm{d}\mu^\flat}(y) \\ f^\flat_{w,h}(y, Z = 1) &= \mathbb{P}(Z = 1|Y = y)\frac{\mathrm{dF}}{\mathrm{d}\mu^\flat}(y) = w(y)\frac{\mathrm{dH}}{\mathrm{d}\mu^\flat}(y), \end{aligned}$$

with corresponding marginal density

$$f^\flat_{w,h}(y) = \sum_z f^\flat_{w,h}(y, z) = w(y)\frac{\mathrm{dF}}{\mathrm{d}\mu^\flat}(y) + \bar{F}_w \frac{\mathrm{dH}}{\mathrm{d}\mu^\flat}(y),$$

which indeed coincides with the censored density.

Similarly, the censored scoring rule of Definition 8 can alternatively be written in terms of the same Z and Q random variables in the following way

$$S^\flat_{h,w}(f, y) := \mathbb{E}_{Z|Y=y,Q} S([f]^\flat_{h,w}, \phi_{w,h}(y, Z, Q)),$$

since

$$\begin{aligned}\mathbb{E}_{Z|Y=y,Q}S([f]^{\flat}_{h,w},\phi_{w,h}(y,Z,Q)) =&\mathbb{E}_{Z|Y=y,Q}\Big(S(f^{\flat}_{h,w},y)\mathbb{1}_{Z=1}+S(f^{\flat}_{h,w},Q)\mathbb{1}_{Z=0}\Big),\\ =&w(y)S(f^{\flat}_{h,w},y)+(1-w(y))\mathbb{E}_{Q}S(f^{\flat}_{h,w},Q),\end{aligned}$$

coincides with the density version of Definition 8.

4 Hypothesis testing

4.1 Localised Neyman–Pearson

In anticipation of our favourite applications, we now switch to an explicit time series context. In particular, 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 $\mathcal{G}$, respectively. The process generates the filtration $\{\mathcal{F}_t\}_{t=1}^{T}$, in which $\mathcal{F}_t = \sigma(Y_1, \dots, Y_t)$ is the information set at time t, satisfying $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}$, $\forall t$. We denote predictive distributions of Y_{t+1} based on $\mathcal{F}_t$ by F_t, predictive distribution functions by F_t and predictive μ_t-densities by f_t. The existence of the sequence of densities f_t is implied by the existence of a sequence of measures $\{\mu_t\}$ such that $\mathrm{F}_t \ll \mu_t, \forall t$. Furthermore, the regions of interest $A_t \subseteq \mathcal{Y}$ are always assumed to be $\mathcal{F}_t$-measurable.

The aim of this section is to derive a uniformly most powerful (UMP) test for the following null and alternative hypothesis

$$\mathbb{H}_0 : p_{0t}\mathbb{1}_{A_t} = f_{0t}\mathbb{1}_{A_t},\ \forall t \qquad \text{vs} \qquad \mathbb{H}_1 : p_{1t}\mathbb{1}_{A_t} = f_{1t}\mathbb{1}_{A_t}, \quad \forall t. \tag{8}$$

Although the predictive densities $f_{jt} = \frac{\mathrm{F}_{jt}}{\mathrm{d}\mu_t}$, $j \in \{0, 1\}$, are assumed to be known, the testing problem remains a multiple versus multiple hypothesis test due to the lacking specification of the density outside the regions of interest A_t. Yet, since the densities p_{jt} must integrate to one on $A_t \cup A_t^c$, the null hypothesis does imply that these densities integrate to $\mathrm{F}_{jt}(A_t^c)$ on A_t^c. Therefore, the implied specification on A_t^c can be summarised as

$$\frac{\mathrm{F}_{jt}(A^c)}{\mathrm{H}_{jt}(A^c)} h_{jt} \mathbb{1}_{A_t^c} = \mathrm{F}_{jt}(A^c)[h_{jt}]^{\sharp}_{A_t^c} \mathbb{1}_{A_t^c}, \quad j \in \{0,1\},$$

where the unknown densities $h_{jt} = \frac{\mathrm{H}_{jt}}{\mathrm{d}\mu_t}$ can be seen as infinite dimensional nuisance parameters.

Explicitising 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 following equivalent hypotheses

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

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

$$\mathscr{p}_j = \left\{ \prod_{t=0}^{T-1} \Big(f_j(y_{t+1}) \mathbb{1}_{A_t}(y_{t+1}) + \mathrm{F}_{jt}(A^c)[h_{jt}]^{\sharp}_{A_t^c}(y_{t+1}) \mathbb{1}_{A_t^c}(y_{t+1}) \Big), h_j \in \mathscr{h} \right\}, \quad j \in \{0,1\},$$

in which $\mathscr{h}$ denotes the space of all densities on $A^c = \prod_{t=0}^{T-1} A_t^c$.

Let $\phi : \mathcal{Y}^T \to [0,1]$ denote a test function determining which values should be included in the critical region. In terms of the index set of all observations $\mathcal{I} = \{1,\ldots,T\}$, this space can also be denoted as $\mathcal{Y}(\mathcal{I}) = \prod_{t \in \mathcal{I}} \mathcal{Y}_t$. The aim of this section is to find a uniformly most powerful (UMP) test ϕ^* of size α for testing problem (8), i.e. a solution to the maximisation problem

$$\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{p_1} \phi, \qquad \Phi(\alpha) = \{\phi : \sup_{p_0 \in \mathscr{p}_0} \mathbb{E}_{p_0} \phi \le \alpha\}. \tag{9}$$

As a first step toward the solution, given by Theorem 3, let us fix an $h_1 \in \mathscr{h}$ 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 density on A^c, we conjecture that a UMP test neglects the information about the shape of the density on A^c. If $T = 2$, for example, and we consider the optimal test on $A_1 \times A_2^c$, our intuition is that an optimal test does not care about the shape of $[h_2]^\sharp_{A_2{}^c}$, that is, the specific values $[h_2]^\sharp_{A_2{}^c}(y_2)$ for all $y_2 \in A_2{}^c$, but just about the total probability of an outcome falling into A_2^c. In other words, we expect that a solution to problem (9) has integrated out the dependence on the nuisance densities.

Although it is obvious that marginalising out the still assumed to be fixed density $h_1 \in \hbar$ is harmless in terms of power, it is non-trivial that this is an affordable strategy in terms of size for all $h_0 \in \hbar$. Lemma 1 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 1 then formalises the idea that we can restrict our attention to tests of the conjectured kind.

Lemma 1. *Consider testing problem (8) and suppose that the outcomes $(y_t)_{t\in\mathcal{I}_A}$ are in A_t, and the remaining $n-k$, with $k = |\mathcal{I}_A|$, observations $(y_t)_{t\in\mathcal{I}_{A^c}}$ in A_t^c. For an arbitrary but fixed density $h_1 \in \hbar$, the test*

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

where $\phi^*_{h_1}$ denotes a solution to problem (9), is such that $\psi_{h_1} \in \Phi(\alpha)$.

Corollary 1. *Consider testing problem (8) 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. 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,*

$$\max_{\phi\in\Phi(\alpha)} \mathbb{E}_{p_1}\phi = \max_{\psi\in\Psi(\alpha)} \mathbb{E}_{p_1}\psi, \qquad \forall h_1 \in \hbar.$$

For any fixed $h_1 \in \hbar$, the reduced optimisation problem resulting from Corollary 1, simplifies to a simple versus simple hypothesis in terms of the censored measures $\mathrm{d}[\mathrm{F}_{jt}]^\flat_{A_t} = \mathbb{1}_{A_t}\mathrm{dF}_{jt} + \mathrm{F}_{jt}(A_t^c)\mathrm{d}\delta_*$, enabling us to formalise a localised version of the Fundamental Lemma of Neyman and Pearson (1933), included below as Theorem 3.

Theorem 3 (Localised Neyman-Pearson)**.** *The UMP test for testing problem* (8) *is given by*

$$\phi^\flat_A(\mathbf{y}) = \begin{cases} 1, & \text{if } \lambda(\mathbf{y}) > c \\ \gamma & \text{if } \lambda(\mathbf{y}) = c \\ 0, & \text{if } \lambda(\mathbf{y}) < c \end{cases} \qquad \lambda(\mathbf{y}) = \frac{[f_1]^\flat_A(\mathbf{y})}{[f_0]^\flat_A(\mathbf{y})}, \qquad [f_j]^\flat_A(\mathbf{y}) = \prod_{t=0}^{T-1} [f_{jt}]^\flat_{A_t}(y_{t+1}), \quad j \in \{0,1\},$$

where c is the largest constant such that $[\mathrm{F}_0]^\flat_A\big(\lambda(\mathbf{y}) \geq c\big) \geq \alpha$ and $[\mathrm{F}_0]^\flat_A\big(\lambda(\mathbf{y}) \leq c\big) \geq 1-\alpha$, and $\gamma \in [0,1]$ is such that $\alpha = [\mathrm{F}_0]^\flat_A\big(\lambda(\mathbf{y}) > c\big) + \gamma[\mathrm{F}_0]^\flat_A\big(\lambda(\mathbf{y}) = c\big)$.

It is worth emphasising that the obtained equivalence between testing problem (8) and $\mathbb{H}_j:\ p_j = [f_j]^\flat_{A_t}$, $j \in \{0,1\}$, is a priori unobvious, since

$$p_j = \Big(f_j \mathbb{1}_A + \mathrm{F}_j(A^c)[h_j]^\sharp_{A^c}\mathbb{1}_{A^c}\Big) \implies p_j = [f_j]^\flat_A,$$

but not the other way around. Formulated differently,

$$\mathbb{H}_j : [p_j]^\flat_A = [f_j]^\flat_A$$

is a multiple versus multiple hypothesis about p_j (for example satisfied if $p_j = [f_j]^\flat_A$), but a simple versus simple hypothesis about $[p_j]^\flat_A$.

Corollary 2. *Another formulation of the UMP test for testing problem* (8) *is given by the test defined in Theorem 3, with $\lambda(\mathbf{y})$ replaced by $\tilde{\lambda}(\mathbf{y}) = \sum_{t=0}^{T-1}\big(S^{csl}_{A_t}(f_{1t}, y_{t+1}) - S^{csl}_{A_t}(f_{0t}, y_{t+1})\big)$, where $S^{csl}_{A_t}$ denotes the censored likelihood score (csl) proposed by Diks et al. (2011).*

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

$$\tilde{\lambda}(\mathbf{y}) = \sum_{t=0}^{T-1}\big(S^{\mathrm{csl}}_{A_t}(f_{1t}, y_{t+1}) - S^{\mathrm{csl}}_{A_t}(f_{0t}, y_{t+1})\big) = \sum_{t=0}^{T-1}\Bigg(\log\Big([f_{1t}]^\flat_{A_t}(y_{t+1})\Big) - \log\Big([f_{0t}]^\flat_{A_t}(y_{t+1})\Big)\Bigg) = \log\lambda(\mathbf{y})$$

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

Example 2. *Consider the special case $T = 1$. For one observation, it is straightforward to derive a most powerful test on A^c. For any $h_1 \in \hbar$, maximisation problem (9) simplifies to*

$$\begin{aligned}\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{p_1} \phi(y) &= \max_{\phi \in \Phi(\alpha)} \left\{ \mathbb{E}_{f_1} \phi_A(y) + \mathrm{F}_1(A^c) \mathbb{E}_{[h_1]^\sharp_{A_c}} \phi_{A^c}(y) \right\} \\ &= \max_{\alpha_A \le \alpha} \left\{ \max_{\phi_A \in \Phi_A(\alpha_A)} \{\mathbb{E}_{f_1} \phi_A(y)\} + \mathrm{F}_1(A^c) \max_{\phi_{A^c} \in \Phi_{A^c}(\alpha - \alpha_A)} \left\{ \mathbb{E}_{[h_1]^\sharp_{A_c}} \phi_{A^c}(y) \right\} \right\} \\ &= \max_{\alpha_A \le \alpha} \left\{ \max_{\phi_A \in \Phi_A(\alpha_A)} \{\mathbb{E}_{f_1} \phi_A(y)\} + \mathrm{F}_1(A^c) \frac{\alpha - \alpha_A}{\mathrm{F}_0(A^c)} \mathbb{1}_{A^c} \right\}.\end{aligned}$$

After all, rejecting with probability $\frac{\alpha - \alpha_A}{\mathrm{F}_0(A^c)}$ if $y \in A^c$ is optimal since this is size correct and any more complicated test function ϕ_{A^c} has lower power. This can be verified as follows. For all level $\alpha - \alpha_A$ tests ϕ_{A^c}, i.e. $\phi_{A^c} \in \Phi_{A^c}(\alpha - \alpha_A)$, we have that

$$\begin{aligned}\mathrm{F}_1(A^c) \mathbb{E}_{[h_1]^\sharp_{A_c}} \phi_{A^c}(y) &\le \mathrm{F}_1(A^c) \sup_{h_1 \in \hbar} \{\mathbb{E}_{[h_1]^\sharp_{A_c}} \phi_{A^c}(y)\} \\ &= \mathrm{F}_1(A^c) \sup_{h_0 \in \hbar} \{\mathbb{E}_{[h_0]^\sharp_{A_c}} \phi_{A^c}(y)\} \\ &\le \mathrm{F}_1(A^c) \frac{\alpha - \alpha_A}{\mathrm{F}_0(A^c)}.\end{aligned}$$

Consequently, the test

$$\phi^*_{A^c}(y) = \frac{\alpha - \alpha_A}{\mathrm{F}_0(A^c)}, \qquad y \in A^c,$$

is most powerful against any other test $\phi^*_{A^c}(y)$ of size $\alpha - \alpha_A$.

This solution, also documented by Holzmann and Klar (2016), coincides with the UMP test given by Theorem 3. Indeed, suppose that the size α is such that $\frac{\mathrm{F}_1(A^c)}{\mathrm{F}_0(A^c)} = c$, i.e. not all of the size is spent on A, then the randomisation probability γ in Theorem 3 is such that

$$\alpha = \alpha_A + \gamma \mathrm{F}_0 \left(\lambda(y) = \frac{\mathrm{F}_1(A^c)}{\mathrm{F}_0(A^c)} \right) = \alpha_A + \gamma \mathrm{F}_0 \left(A^c \right) \implies \gamma = \frac{\alpha - \alpha_A}{\mathrm{F}_0(A^c)}.$$

Corollary 3. *For testing problem (8), the test*

$$\phi_A^\sharp(\mathbf{y}) = \begin{cases} 1, & \text{if } \lambda^\sharp(\mathbf{y}) > c \\ \gamma & \text{if } \lambda^\sharp(\mathbf{y}) = c \\ 0, & \text{if } \lambda^\sharp(\mathbf{y}) < c \end{cases} \qquad \lambda_A^\sharp(\mathbf{y}) = \frac{[f_1]_A^\sharp(\mathbf{y})}{[f_0]_A^\sharp(\mathbf{y})}\mathbb{1}_A(\mathbf{y}), \qquad [f_j]_A^\sharp(\mathbf{y}) = \prod_{t=1}^T [f_{jt}]_{A_t}^\sharp(y_t), \quad j \in \{0,1\},$$

where 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 $\alpha = [\mathrm{F}_0]_A^\flat\big(\lambda(\mathbf{y}) > c\big) + \gamma[\mathrm{F}_0]_A^\flat\big(\lambda(\mathbf{y}) = c\big)$, is not UMP.

4.2 Giacomini and White test: Theoretical Example

Let $\mathcal{P}$ be given by the parametric Laplace(θ,μ) family, with density function

$$p(y;\theta,\mu) = \frac{1}{2\theta}\mathrm{e}^{-\frac{1}{\theta}|y-\mu|}, \qquad \theta > 0,$$

and, for convenience, the additional restriction that $\mu > r$, where r is the upper bound of the region of interest. In other words, $A = (-\infty, r]$ and $\mathcal{P} = \{p(y;\theta,\mu), (\mu,\theta) \in \mathbf{\Omega}\}$, where $\mathbf{\Omega} = \Omega_\mu \times \Omega_\theta = (r,\infty] \times (0,\infty)$. The two candidate densities f and g are also from the Laplace family with parameter values μ_f, θ_f, μ_g and θ_g. For the region of interest, we set $r = -2$. The null sets based on the cl and csl scoring rule are given by

$$\mathbf{\Omega}_0^\sharp = \left\{(\mu,\theta) \in \mathbf{\Omega} : -\left(\theta\left(\frac{1}{\theta_f} - \frac{1}{\theta_g}\right) + \log\left(\frac{\theta_f}{\theta_g}\right)\right)\frac{1}{2}\mathrm{e}^{-\frac{1}{\theta}(\mu-r)} = 0\right\}$$

$$\mathbf{\Omega}_0^\flat = \left\{(\mu,\theta) \in \mathbf{\Omega} : -\left((\theta - r)\left(\frac{1}{\theta_f} - \frac{1}{\theta_g}\right) + \log\left(\frac{\theta_f}{\theta_g}\right) + \left(\frac{\mu_f}{\theta_f} - \frac{\mu_g}{\theta_g}\right)\right.\right.$$
$$\left.\left. - \log\left(\frac{1 - \frac{1}{2}\mathrm{e}^{-\frac{1}{\theta_f}(\mu_f - r)}}{1 - \frac{1}{2}\mathrm{e}^{-\frac{1}{\theta_g}(\mu_g - r)}}\right)\right)\frac{1}{2}\mathrm{e}^{-\frac{1}{\theta}(\mu-r)} + \log\left(\frac{1 - \frac{1}{2}\mathrm{e}^{-\frac{1}{\theta_f}(\mu_f - r)}}{1 - \frac{1}{2}\mathrm{e}^{-\frac{1}{\theta_g}(\mu_g - r)}}\right) = 0\right\}.$$

Now consider the special case in which $\theta_f = \theta_g$. Then, $\mathbf{\Omega}_0^\sharp = \mathbf{\Omega}$, meaning that the GW test based on the conditional rule has no asymptotic power. On the other hand, the censored null set remains a lower-dimensional subspace of the parameter space. The lack of power for the conditional rule is a consequence of the result that the conditional rule is not strictly locally proper but proportionally locally proper. More specifically, since the conditional scoring rule is unable to discriminate between proportional densities, such as the left-tail of the candidate densities for $\theta_f = \theta_g$, the statistical distance from any density in the density space to candidate f and g is equivalent.

5 Simulation study

5.1 Laplace

Figure 1: Rejection rates $c = 20$ for $\theta_f = 1$ and $\theta_g = 1.1$

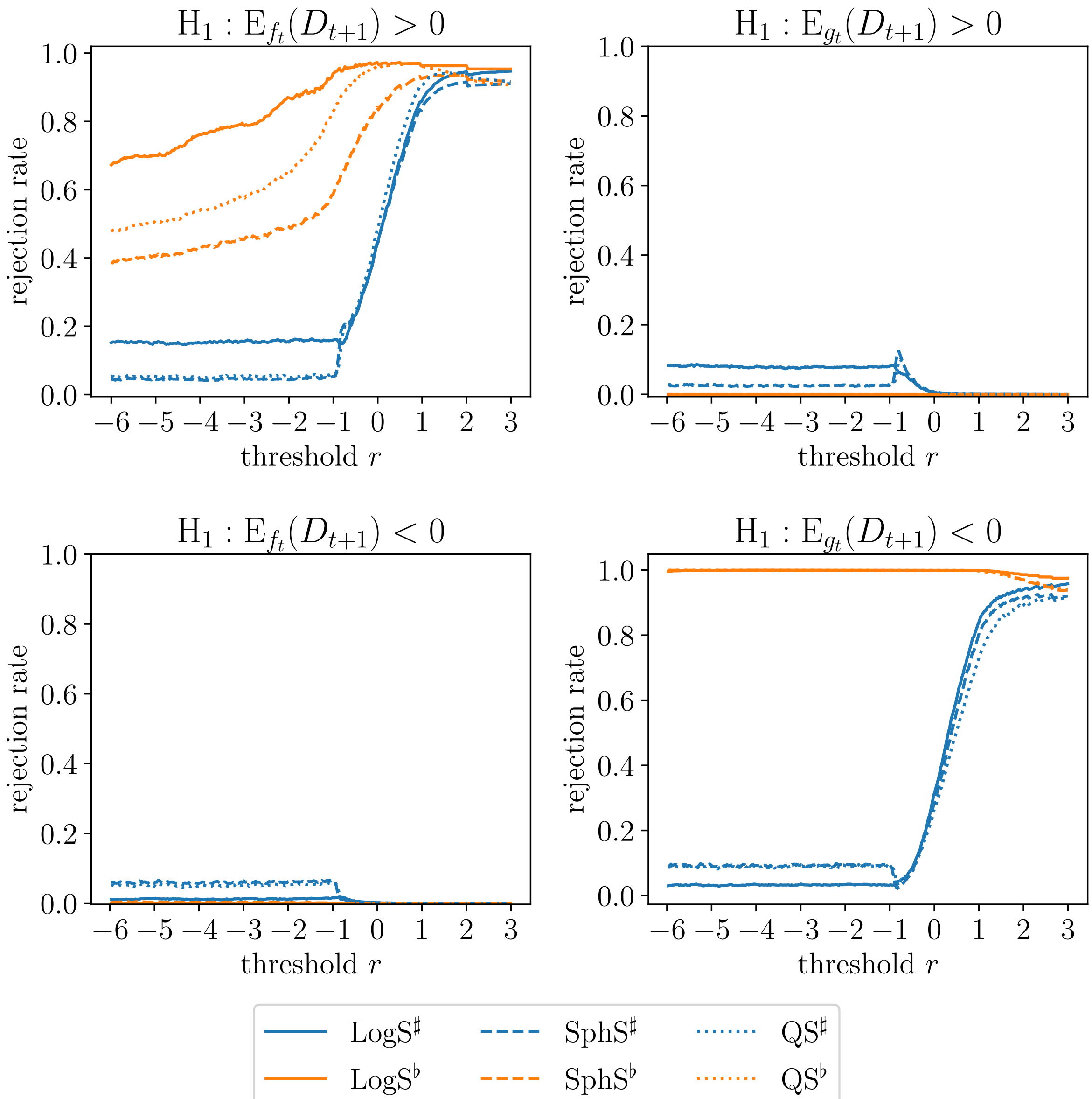


Figure 2: Rejection rates $c = 20$ for $\theta_f = 1$ and $\theta_g = 1.1$

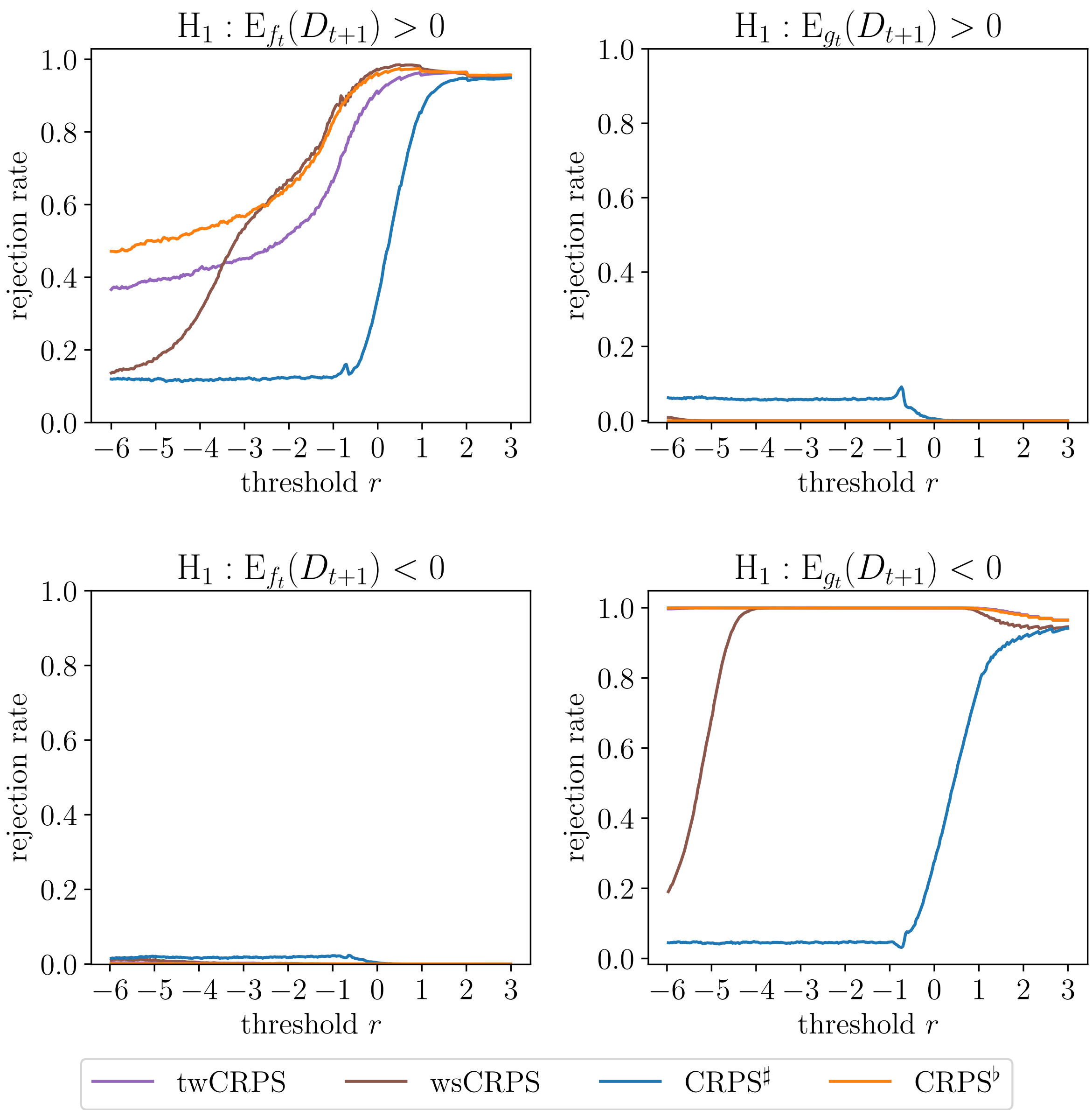


Figure 3: Rejection rates $c = 40$ for $\theta_f = 1$ and $\theta_g = 1.1$

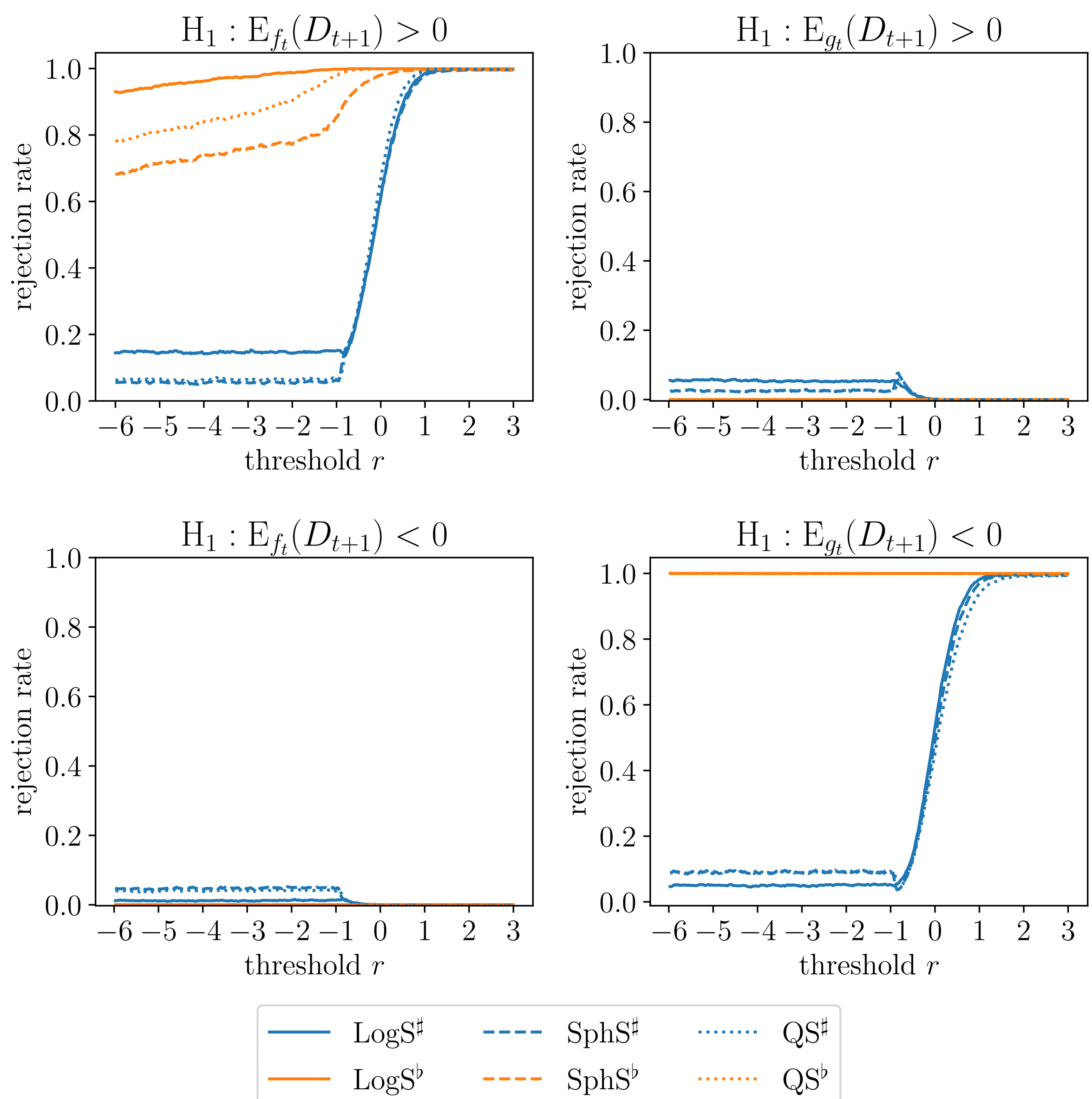
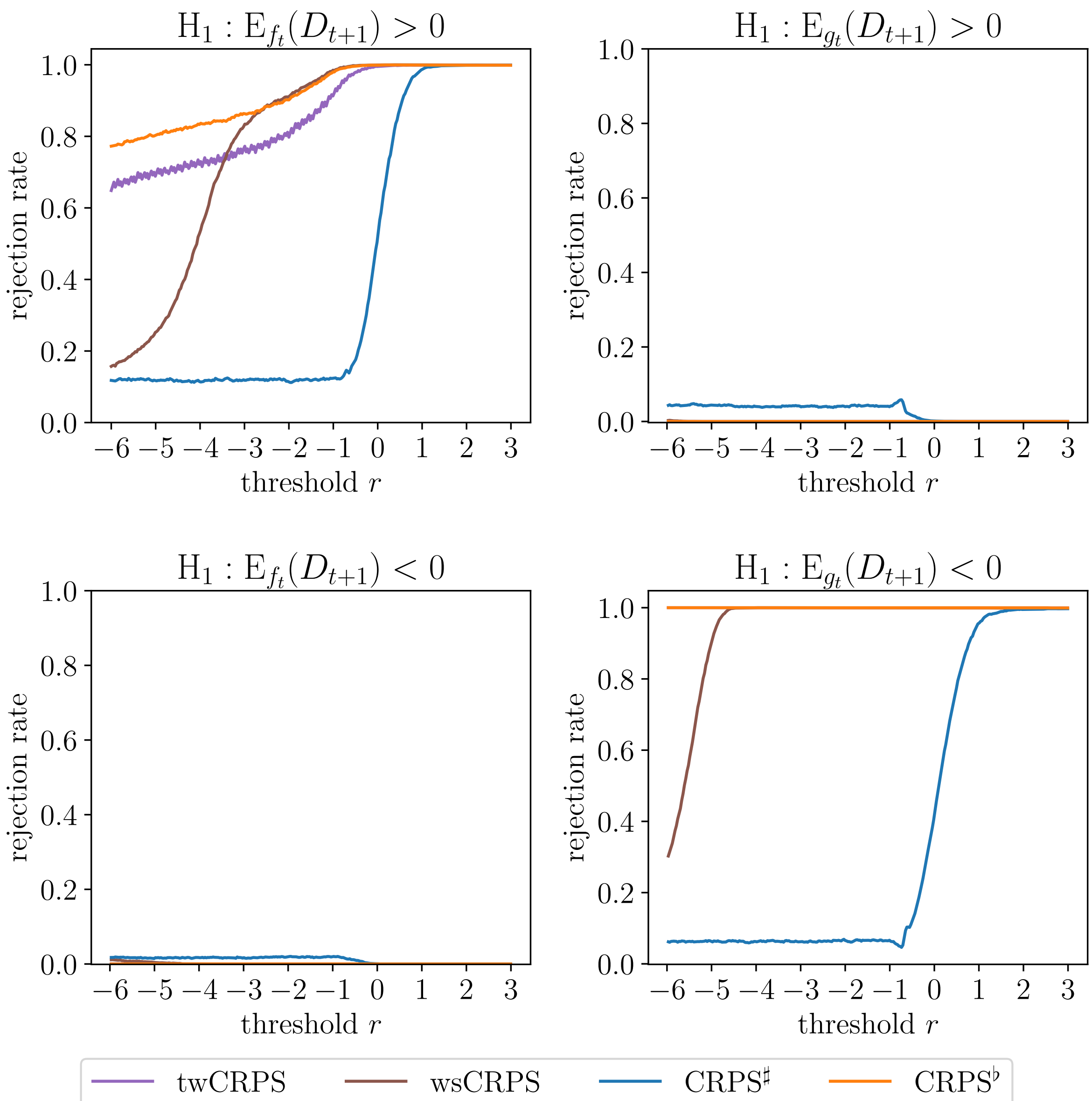


Figure 4: Rejection rates $c = 40$ for $\theta_f = 1$ and $\theta_g = 1.1$



5.2 Normal-Student-$t(5)$

Figure 5: Rejection rates $c = 20$

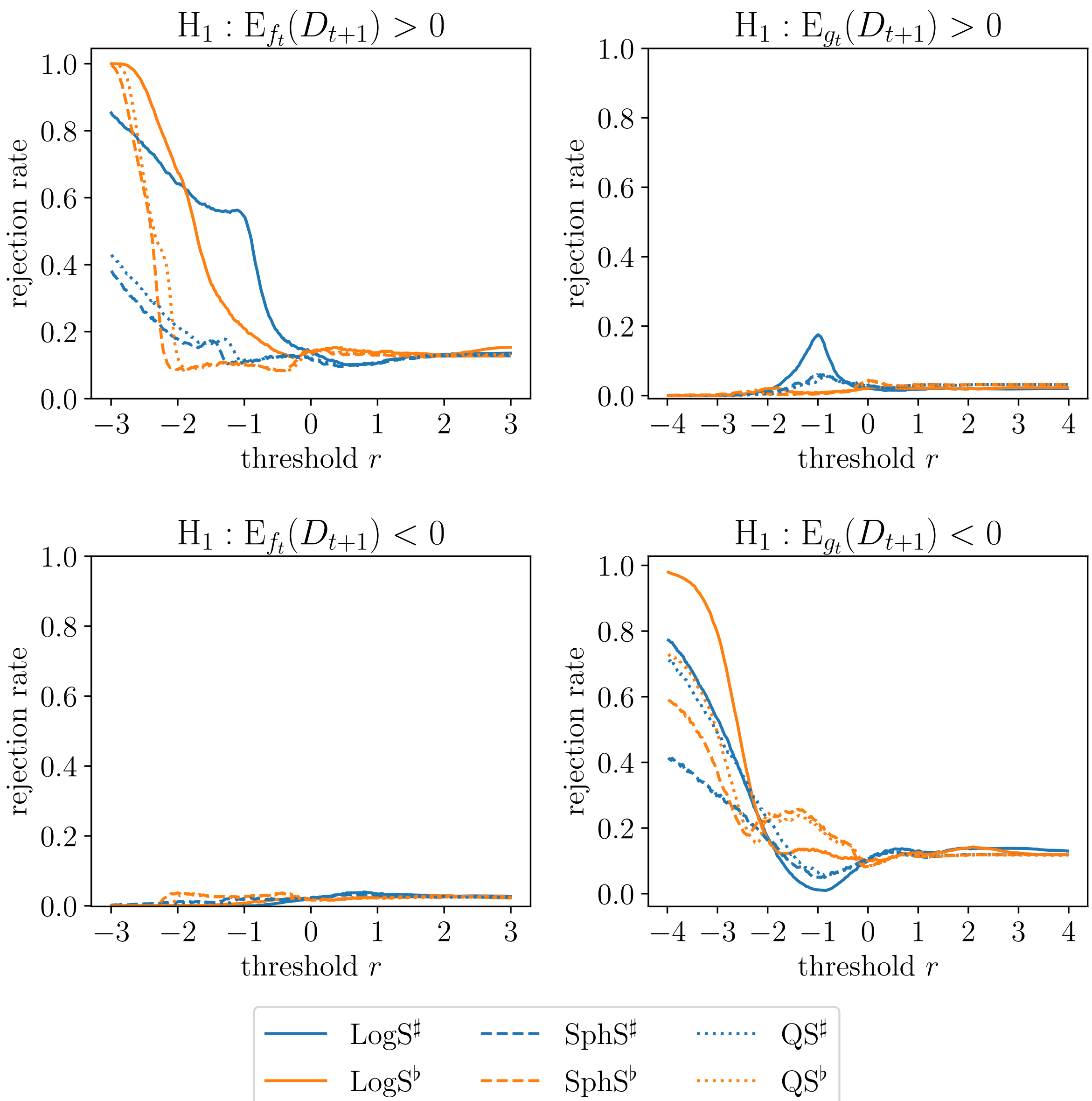


Figure 5: Rejection rates $c = 20$

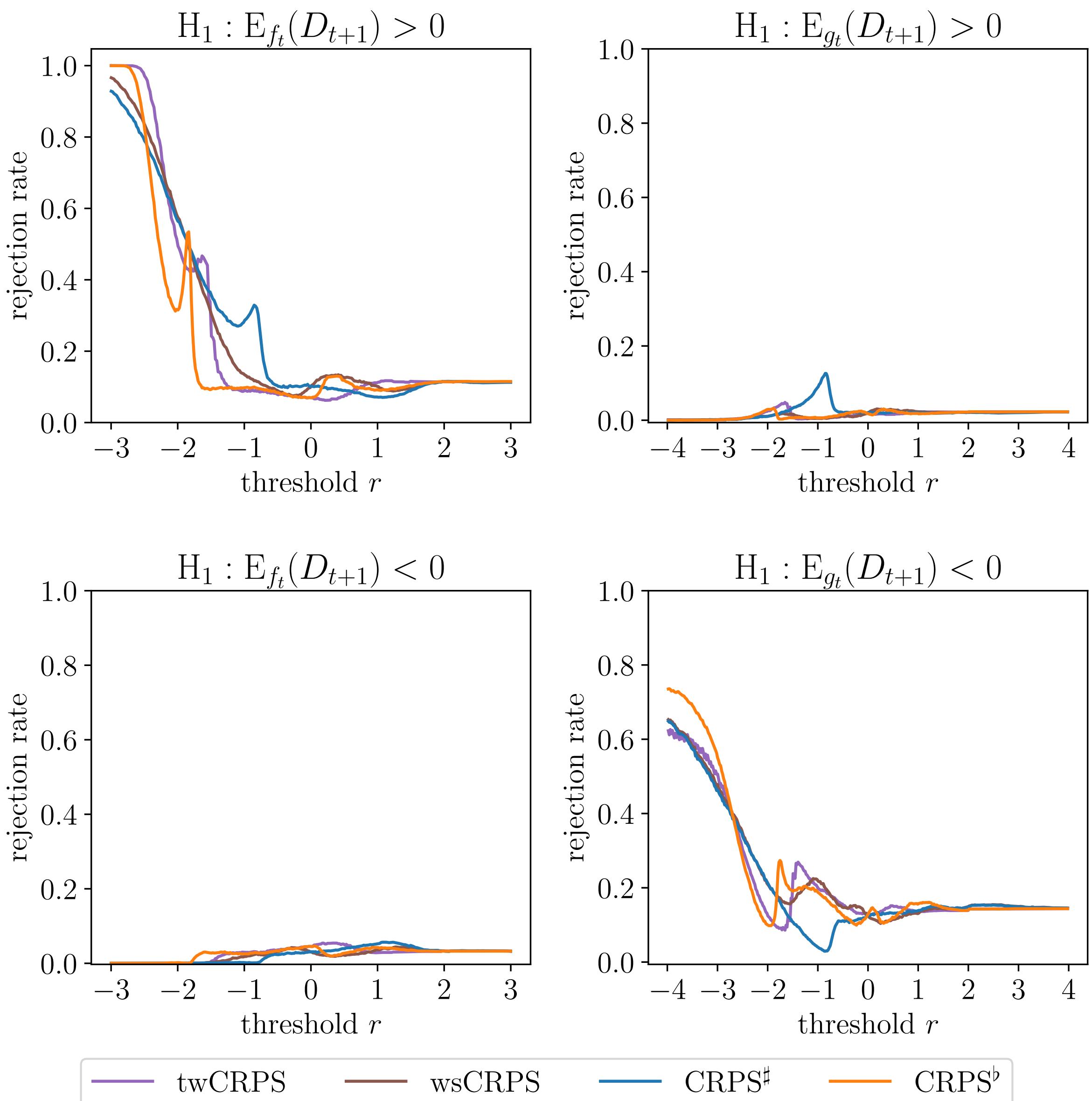


Figure 6: Rejection rates $c = 5$

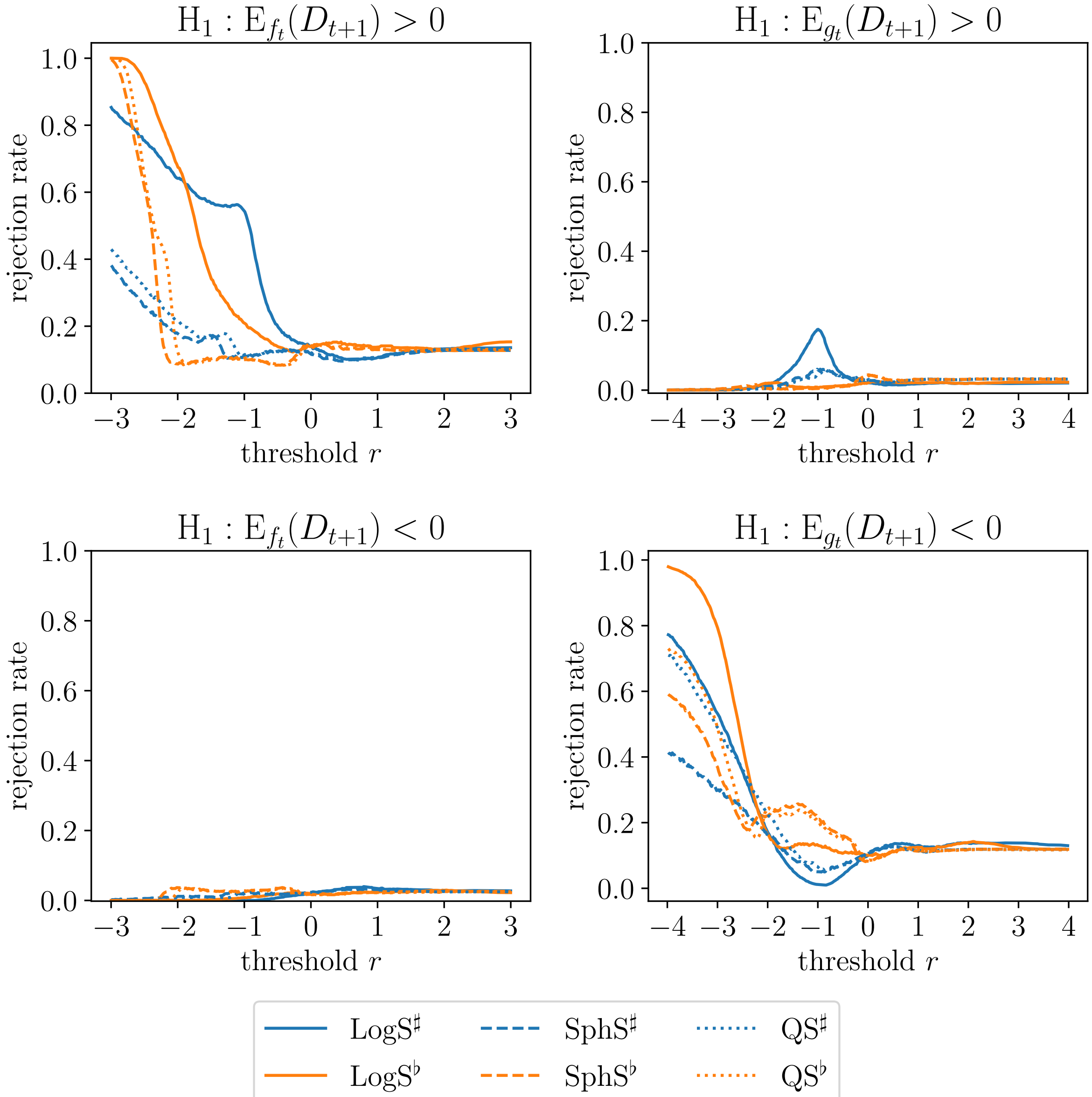
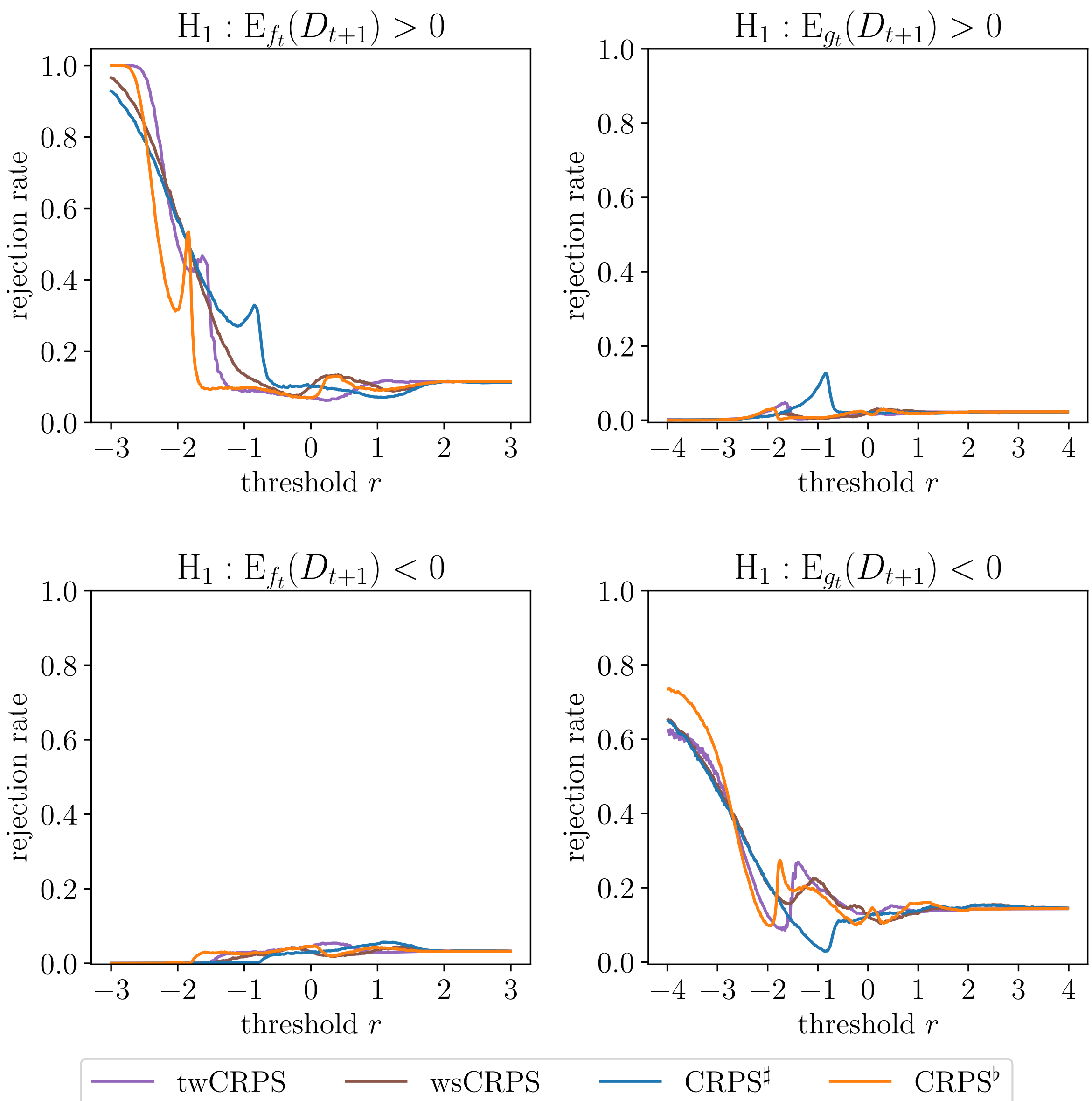


Figure 7: Rejection rates $c = 5$



6 Conclusion

In many applications, forecasters are not equally interested in all possible outcomes of the random variable of interest. For such cases, we have motivated the use of censoring as focusing device. In particular, we have shown that focusing scoring rules by applying them to censored distributions leads to strictly locally proper scoring rules. To the best of our knowledge, we are first in deriving a transformation of the original scoring rule that preserves strict propriety.

Our method is very flexible with regard to the choice of the original scoring rule and weight function. For specific choices, the generalised censored scoring rule delivers intuitively sound scoring rules that can easily be implemented by practitioners. When applied to the logarithmic scoring rule, our focusing procedure implies the well-known censored likelihood score.

The censored likelihood score also appears in a second important result of this paper. In particular, we have shown that the UMP test for a localised version of the standard simple versus simple Neyman Pearson testing problem is based on the censored likelihood ratio. Furthermore, the results of our Monte Carlo study suggest that our theoretical findings spill over to the finite sample properties of other forecast evaluation tests. In our experiments, striking differences in power always favour censoring.

Appendix

A Proofs

A.1 Proof Theorem 1

As discussed in Section 3.1, the censoring process can be viewed as the result of two intermeditate steps in which the first step extends the definitions of the orginal measures and weight functions to the extended measurable space, i.e. mapping $(\mathscr{P}, \mathcal{Y}, \mathcal{W})$ to $(\mathscr{P}^*, \mathcal{Y}^*, \mathcal{W}^*)$. Here, we consider all outcomes y, events E, measures F and weight w functions with respect to the extended classes $\mathcal{Y}^*, \mathcal{G}^*, \mathscr{P}^*$ and $\mathcal{W}^*$. Correspondingly, we will show that $S^\flat$ is strictly locally proper relative to $(\mathscr{P}^*, \mathcal{W}^*)$. According to Definition 5, we then need to prove a list of three things: *(i)* $S^\flat$ must be localising relative to $\mathcal{W}^*$, *(ii)* $S^\flat$ must be proper relative to $\mathscr{P}^*$, $\forall w \in \mathcal{W}^*$, and *(iii)* the if and only if statement in Definition 5.

We start by proving an if and only if statement that will serve as a building block for assertions *(i)* and *(iii)*. In particular, it will be helpful to know that

$$\mathrm{F}^\flat_w(E) = \mathrm{G}^\flat_w(E),\ \forall E \in \mathcal{G}^* \iff \mathrm{F}(E \cap \{w > 0\}) = \mathrm{G}(E \cap \{w > 0\}),\ \forall E \in \mathcal{G}^*, \tag{A.1}$$

for which we prove both implications in isolation.

" $\Longrightarrow$ " Suppose that $\mathrm{F}^\flat_w(E) = \mathrm{G}^\flat_w(E),\ \forall E \in \mathcal{G}^*$. Then, since $* \in \mathcal{G}^*$ and $\mathrm{F}^*_w(*) = 0,\ \forall w \in \mathcal{W}^*, \mathrm{F} \in \mathscr{P}^*$, it immediately follows that $\bar{\mathrm{F}}_w = \bar{\mathrm{G}}_w$. This in turn implies that $\mathrm{F}^*_w(E) = \mathrm{G}^*_w(E)$, $\forall E \in \mathcal{G}^*$, which holds if and only if $\mathrm{F}(E \cap \{w > 0\}) = \mathrm{G}(E \cap \{w > 0\}),\ \forall E \in \mathcal{G}^*$.

" $\Longleftarrow$ " Suppose that $\mathrm{F}(E \cap \{w > 0\}) = \mathrm{G}(E \cap \{w > 0\}),\ \forall E \in \mathcal{G}^*$. Then, using the same if and only if relation as in the proof of the other direction, we also have that $\mathrm{F}^*_w(E) = \mathrm{G}^*_w(E)$, $\forall E \in \mathcal{G}^*$, and hence in particular $1 - \bar{F}_w = \mathrm{F}^*_w(\mathcal{Y}) = \mathrm{G}^*_w(\mathcal{Y}) = 1 - \bar{G}_w$. But then, $\mathrm{F}^\flat_w(E) = \mathrm{G}^\flat_w(E),\ \forall E \in \mathcal{G}^*$.

We can now easily proof the three listed items:

(i) $S^\flat$ is localising relative to $\mathcal{W}^*$, as $\mathrm{F}(E \cap \{w > 0\}) = \mathrm{G}(E \cap \{w > 0\}),\ \forall E \in \mathcal{G}^*$ implies, by Equation (A.1), that $\mathrm{F}_w^\flat(E) = \mathrm{G}_w^\flat(E),\ \forall E \in \mathcal{G}^*$, whence it follows that $S_w^\flat(\mathrm{F}, y) = S_w^\flat(\mathrm{G}, y)$.

(ii) To show that $S_w^\flat$ is proper for all $w \in \mathcal{W}^*$, we start by fixing an arbitrary $w \in \mathcal{W}^*$. Since S is proper relative to $\mathscr{P}^\flat \supseteq \mathscr{P}_w^\flat$, the following inequality holds

$$\int_{\mathcal{Y}} S(\mathrm{P}_w^\flat, y)\mathrm{P}_w^\flat(\mathrm{d}y) \geq \int_{\mathcal{Y}} S(\mathrm{F}_w^\flat, y)\mathrm{P}_w^\flat(\mathrm{d}y), \quad \forall \mathrm{P}_w^\flat, \mathrm{F}_w^\flat \in \mathscr{P}_w^\flat. \tag{A.2}$$

By definition of the class $\mathscr{P}_w^\flat \equiv \{[\mathrm{F}]_w^\flat, \mathrm{F} \in \mathscr{P}\}$ and the fact that $\mathrm{F}(*) = 0,\ \forall \mathrm{F} \in \mathscr{P}^*$, this is equivalent to

$$\int_{\mathcal{Y}} S([\mathrm{P}]_w^\flat, y)[\mathrm{P}]_w^\flat(\mathrm{d}y) \geq \int_{\mathcal{Y}} S([\mathrm{F}]_w^\flat, y)[\mathrm{P}]_w^\flat(\mathrm{d}y), \quad \forall \mathrm{P}, \mathrm{F}^* \in \mathscr{P}^*. \tag{A.3}$$

For these integrals, verify that

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

$\forall \mathrm{P}, \mathrm{F} \in \mathscr{P}^*$. Consequently, Equation (A.2) is equivalent to

$$\int_{\mathcal{Y}^*} S_w^\flat(\mathrm{P}, y)\mathrm{P}(\mathrm{d}y) \geq \int_{\mathcal{Y}^*} S_w^\flat(\mathrm{F}, y)\mathrm{P}(\mathrm{d}y), \quad \forall \mathrm{P}, \mathrm{F} \in \mathscr{P}^*, \tag{A.4}$$

and hence $S_w^\flat$ is proper relative to $\mathscr{P}^*$ by Definition 2.

(iii) Since S is also strictly proper relative to $\mathscr{P}^\flat \supseteq \mathscr{P}^\flat_w$, we find that, $\forall w \in \mathcal{W}^*$,

$$\int_{\mathcal{Y}^*} S(\mathrm{P}^\flat_w, y)\mathrm{P}^\flat_w(\mathrm{d}y) = \int_{\mathcal{Y}^*} S(\mathrm{F}^\flat_w, y)\mathrm{P}^\flat_w(\mathrm{d}y) \iff \mathrm{P}^\flat_w = \mathrm{F}^\flat_w,$$

and thus, by Equation (A.1),

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

$\forall E \in \mathcal{G}^*$, and hence, by the equality $\int_{\mathcal{Y}^*} S(\mathrm{F}^\flat_w, y)\mathrm{P}^\flat_w(\mathrm{d}y) = \int_{\mathcal{Y}^*} S^\flat_w(\mathrm{F}, y)\mathrm{P}(\mathrm{d}y)$ derived above, also that

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

$\forall E \in \mathcal{G}^*$, which is the desired if and only if statement of Definition 5.

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

A.2 Proof Theorem 2

For clarity of exposition, we first prove the main ingredients of the proof via two isolated lemmas and a corollary.

Lemma 2. *Consider the censored scoring rule defined in Definition 8. $\forall w \in \mathcal{W}$ and $\mathrm{H} \in \mathscr{H}$, the following identity holds $\int_{\mathcal{Y}} S^\flat_{w,\mathrm{H}}(\mathrm{F}, y)\mathrm{P}(\mathrm{d}y) = \int_{\mathcal{Y}} S(\mathrm{F}^\flat_{w,\mathrm{H}}, y)\mathrm{P}^\flat_{w,\mathrm{H}}(\mathrm{d}y)$.*

Proof.

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

□

Lemma 3. *Consider two distributions* P *and* F *on the same measurable space* $(\mathcal{Y},\mathcal{G})$*. On the same space, let their censored counterparts* $\mathrm{P}^{\flat}_{w,\mathrm{H}}$ *and* $\mathrm{F}^{\flat}_{w,\mathrm{H}}$ *be given by Definition 8. Then,*

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

Proof. " $\Longrightarrow$ " We start with the most challenging direction, for which Assumption 1 is of critical importance. First, note that

$$\begin{aligned}&\mathrm{F}^{\flat}_{w,\mathrm{H}}(E) = \mathrm{G}^{\flat}_{w,\mathrm{H}}(E), \qquad \forall E\in\mathcal{G}\\ &\quad\Longrightarrow\ \mathrm{F}^{\flat}_{w,\mathrm{H}}(E\cap\{w=c\}) = \mathrm{G}^{\flat}_{w,\mathrm{H}}(E\cap\{w=c\}), \qquad \forall E\in\mathcal{G}\\ &\quad\Longrightarrow\ \int_{\mathcal{Y}}(1-w)\mathrm{dFH}(E\cap\{w=c\}) = \int_{\mathcal{Y}}(1-w)\mathrm{dGH}(E\cap\{w=c\}), \qquad \forall E\in\mathcal{G}\\ &\quad\Longrightarrow\ \int_{\mathcal{Y}}(1-w)\mathrm{dFH}(\{w=c\}) = \int_{\mathcal{Y}}(1-w)\mathrm{dGH}(\{w=c\}),\\ &\quad\Longrightarrow\ \int_{\mathcal{Y}}(1-w)\mathrm{dF} = \int_{\mathcal{Y}}(1-w)\mathrm{dG},\end{aligned}$$

where c denotes a constant such that Assumption 1 is satisfied. Then, exploit this equality to conclude

$$\begin{aligned}
&\mathrm{F}^{\flat}_{w,\mathrm{H}}(E) = \mathrm{G}^{\flat}_{w,\mathrm{H}}(E), \qquad \forall E \in \mathcal{G} \\
&\implies \int_{\mathcal{Y}} w(y)\mathbb{1}_{y\in E}\mathrm{F}(\mathrm{d}y) = \int_{\mathcal{Y}} w(y)\mathbb{1}_{y\in E}\mathrm{G}(\mathrm{d}y), \qquad \forall E \in \mathcal{G} \\
&\implies \mathrm{F}(E \cap \{w > 0\}) = \mathrm{G}(E \cap \{w > 0\}), \qquad \forall E \in \mathcal{G}.
\end{aligned}$$

" $\Longleftarrow$ " The other direction is somewhat trivial. Indeed,

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

and the two implied results jointly imply $\mathrm{F}^{\flat}_{w,\mathrm{H}}(E) = \mathrm{G}^{\flat}_{w,\mathrm{H}}(E),\ \forall E \in \mathcal{G}, \forall \mathrm{H} \in \mathscr{H}$. □

Corollary 4. *The censored scoring rule defined in Definition 8 is localising* $\forall \mathrm{H} \in \mathscr{H}$*.*

Proof. Suppose that $\mathrm{F}(E \cap \{w > 0\}) = \mathrm{G}(E \cap \{w > 0\}),\ \forall E \in \mathcal{G}$. Then, by Lemma 3, $\mathrm{F}^{\flat}_{w,\mathrm{H}}(E) = \mathrm{G}^{\flat}_{w,\mathrm{H}}(E), \forall E \in \mathcal{G}$, 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. The definition of a strictly locally proper scoring rule (Definition 5) and the definitions on which this definition is built, that is, the definition of a locally proper scoring rule (Definition 5) and a localising weighted scoring rule (Definition 4), reveal that we need to prove a list of three things $\forall \mathrm{H} \in \mathscr{H}$: (i) $S^{\flat}_{w,\mathrm{H}}(\mathrm{P}, y)$ must be localising relative to $\mathcal{W}$, (ii) $S^{\flat}_{w,\mathrm{H}}(\mathrm{P}, y)$ must be proper relative to $\mathscr{P}$, $\forall w \in \mathcal{W}$ and (iii) the if and only if statement in Definition 5. We prove them one by one.

(i) $S^{\flat}_{w,\mathrm{H}}(\mathrm{P}, y)$ is localising relative to $\mathcal{W}$, $\forall \mathrm{H} \in \mathscr{H}$, by Corollary 4.

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

$$\int_{\mathcal{Y}} S(\mathrm{P}^\flat_{w,\mathrm{H}}, y)\mathrm{P}^\flat_{w,\mathrm{H}}(\mathrm{d}y) \geq \int_{\mathcal{Y}} S(\mathrm{F}^\flat_{w,\mathrm{H}}, y)\mathrm{P}^\flat_{w,\mathrm{H}}(\mathrm{d}y), \quad \forall \mathrm{P}^\flat_{w,\mathrm{H}}, \mathrm{F}^\flat_{w,\mathrm{H}} \in \mathscr{P}^\flat_{w,\mathrm{H}}, \tag{A.5}$$

which is by definition of the class $\mathscr{P}^\flat_{w,\mathrm{H}} \equiv \{[\mathrm{P}]^\flat_{w,\mathrm{H}}, \mathrm{P} \in \mathscr{P}\}$ equivalent to

$$\int_{\mathcal{Y}} S([\mathrm{P}]^\flat_{w,\mathrm{H}}, y)[\mathrm{P}]^\flat_{w,\mathrm{H}}(\mathrm{d}y) \geq \int_{\mathcal{Y}} S([\mathrm{F}]^\flat_{w,\mathrm{H}}, y)[\mathrm{P}]^\flat_{w,\mathrm{H}}(\mathrm{d}y), \quad \forall \mathrm{P}, \mathrm{F} \in \mathscr{P}, \tag{A.6}$$

and hence, by Lemma 2, also

$$\int_{\mathcal{Y}} S^\flat_{w,\mathrm{H}}(\mathrm{P}, y)\mathrm{P}(\mathrm{d}y) \geq \int_{\mathcal{Y}} S^\flat_{w,\mathrm{H}}(\mathrm{F}, y)\mathrm{P}(\mathrm{d}y), \quad \forall \mathrm{P}, \mathrm{F} \in \mathscr{P}. \tag{A.7}$$

Therefore, $S^\flat_{w,\mathrm{H}}(\mathrm{P}, y)$ is proper relative to $\mathscr{P}$ by Definition 2.

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

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

and thus, by Lemma 3,

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

$\forall E \in \mathcal{G}$, and hence, by Lemma 2, also

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

which is the desired if and only if statement of Definition 5.

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 $(\mathscr{P}, \mathcal{W})$, $\forall \mathrm{H} \in \mathscr{H}$. $\square$

A.3 Proof censored density in Equation (5)

We defined the measures μ and F to the extended measurable space $(\mathcal{Y}^*, \mathcal{G}^*)$ in the exact same way as we proposed to do in Section 3.1, that is, $\mu^*(E) = \mu(E\backslash\{*\})$, if $* \in E$ and $\mu^*(E) = \mu(E)$, otherwise, $\forall E \in \mathcal{G}^*$. Again, we drop the subsript $*$ in the notation of the extended measures, while still considering all measures with respect to the extended measurable space $(\mathcal{Y}^*, \mathcal{G}^*)$, equivalent to the proof of Theorem 1 in Appendix A.1.

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,

$$f^\flat_{w,h} := \frac{\mathrm{dF}^\flat_w}{\mathrm{d}(\mu+\delta_*)} = w\frac{\mathrm{dF}}{\mathrm{d}(\mu+\delta_*)} + \bar{F}_w\frac{\mathrm{d}\delta_*}{\mathrm{d}(\mu+\delta_*)}$$

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

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 of

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

By the same theorem, the solution of 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 $\frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\delta_*)} = \mathbb{1}_{\mathcal{Y}\backslash\{*\}}$ 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

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

where f denotes the μ-density of F. $\square$

A.4 Proof Lemma 1

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, $\forall h_1 \in \hbar$, we have that

$$\begin{aligned}
\sup_{p_0\in\wp_0} \mathbb{E}_{p_0}\psi_{h_1} &= \left(\prod_{t\in\mathcal{I}_{A^c}} \mathrm{F}_0(A_t^c)\right) \sup_{h_0\in\hbar}\int_{\mathcal{Y}^T} \psi_{h_1} \prod_{t\in\mathcal{I}_A} f_{0t}\mathbb{1}_{A_t} \prod_{t\in\mathcal{I}_{A^c}} [h_{0t}]^{\sharp}_{A_t^c}\mathbb{1}_{A_t^c}\mathrm{d}\mu_t \\
&= \left(\prod_{t\in\mathcal{I}_{A^c}} \mathrm{F}_0(A_t^c)\right) \int_{\mathcal{Y}(\mathcal{I}_A)} \psi_{h_1} \prod_{t\in\mathcal{I}_A} f_{0t}\mathbb{1}_{A_t}\mathrm{d}\mu_t \\
&= \left(\prod_{t\in\mathcal{I}_{A^c}} \mathrm{F}_0(A_t^c)\right) \int_{\mathcal{Y}^T} \phi^*_{h_1} \prod_{t\in\mathcal{I}_{A^c}} [h_{1t}]^{\sharp}_{A_t^c}\mathbb{1}_{A_t^c}\mathrm{d}\mu_t \prod_{t\in\mathcal{I}_A} f_{0t}\mathbb{1}_{A_t}\mathrm{d}\mu_t \\
&\leq \left(\prod_{t\in\mathcal{I}_{A^c}} \mathrm{F}_0(A_t^c)\right) \sup_{h_0\in\hbar}\int_{\mathcal{Y}^T} \phi^*_{h_1} \prod_{t\in\mathcal{I}_{A^c}} [h_{0t}]^{\sharp}_{A_t^c}\mathbb{1}_{A_t^c}\mathrm{d}\mu_t \prod_{t\in\mathcal{I}_A} f_{0t}\mathbb{1}_{A_t}\mathrm{d}\mu_t \\
&= \sup_{p_0\in\wp_0} \mathbb{E}_{p_0}\phi^*_{h_1} \\
&\leq \alpha,
\end{aligned}$$

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

A.5 Proof Corollary 1

Fix an arbitrary $h_1 \in \hbar$. Since $\Psi(\alpha) \subseteq \Phi(\alpha)$, we trivially have that $\max_{\phi\in\Phi(\alpha)} \mathbb{E}_{p_1}\phi \geq \max_{\psi\in\Psi(\alpha)} \mathbb{E}_{p_1}\psi$. Now suppose that $\max_{\phi\in\Phi(\alpha)} \mathbb{E}_{p_1}\phi < \max_{\psi\in\Psi(\alpha)} \mathbb{E}_{p_1}\psi$. Then, we can always define the test $\tilde{\psi} = \int_{\mathcal{Y}(\mathcal{I}_{A^c})} \phi^* \prod_{t\in\mathcal{I}_{A^c}} [h_{1t}]^{\sharp}_{A_t^c}\mathbb{1}_{A_t^c}\mathrm{d}\mu_t$, with $\phi^* \in \arg\max_{\phi\in\Phi(\alpha)} \mathbb{E}_{p_1}\phi$, satisfying $\mathbb{E}_{p_1}\phi^* = \mathbb{E}_{p_1}\tilde{\psi}$. But, by Lemma 1, $\tilde{\psi} \in \Psi(\alpha)$, in which case $\max_{\phi\in\Phi(\alpha)} \mathbb{E}_{p_1}\phi = \max_{\psi\in\Psi(\alpha)} \mathbb{E}_{p_1}\tilde{\psi}$, contradicting the assumed strict inequality. $\square$

A.6 Proof Theorem 3

For any fixed $h_1 \in \hbar$, the most powerful test of size α is a solution to the following restricted maximisation problem

$$
\begin{aligned}
\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{p_1} \phi &= \max_{\boldsymbol{\alpha} \in \Delta_{\bar{T}}(\alpha)} \sum_{k=0}^{T} \sum_{s=1}^{\binom{T}{k}} \max_{\phi_{k,s} \in \Phi(\alpha_{k,s})} \mathbb{E}_{p_1} \left(\phi_{k,s} | y_t \in A_t, \forall i \in \mathcal{I}_A(k,s) \wedge y_t \in A_t^c, \forall i \in \mathcal{I}_{A^c}(k,s)\right) \\
&= \max_{\boldsymbol{\alpha} \in \Delta_{\bar{T}}(\alpha)} \sum_{k=0}^{T} \sum_{s=1}^{\binom{T}{k}} \max_{\phi_{k,s} \in \Psi(\alpha_{k,s})} \mathbb{E}_{p_1} \left(\phi_{k,s} | y_t \in A_t, \forall i \in \mathcal{I}_A(k,s) \wedge y_t \in A_t^c, \forall i \in \mathcal{I}_{A^c}(k,s)\right) \\
&= \max_{\boldsymbol{\alpha} \in \Delta_{\bar{T}}(\alpha)} \sum_{k=0}^{T} \sum_{s=1}^{\binom{T}{k}} \max_{\phi_{k,s} \in \Psi(\alpha_{k,s})} \left(\prod_{t \in \mathcal{I}_{A^c}} \mathrm{F}_1(A_t^c) \right) \int_{\mathcal{Y}(\mathcal{I}_A)} \phi_{k,s} \prod_{t \in \mathcal{I}_A} f_{1t} \mathbb{1}_{A_t} \mathrm{d}\mu_t \\
&= \max_{\boldsymbol{\alpha} \in \Delta_{\bar{T}}(\alpha)} \sum_{k=0}^{T} \sum_{s=1}^{\binom{T}{k}} \max_{\phi_{k,s} \in \Psi(\alpha_{k,s})} \int_{\mathcal{Y}^T} \phi_{k,s} \prod_{t=0}^{T-1} \mathrm{d}[\mathrm{F}_t]^\flat_{A_t} \\
&= \max_{\boldsymbol{\alpha} \in \Delta_{\bar{T}}(\alpha)} \sum_{k=0}^{T} \sum_{s=1}^{\binom{T}{k}} \max_{\phi_{k,s} \in \Phi(\alpha_{k,s})} \int_{\mathcal{Y}^T} \phi_{k,s} \prod_{t=0}^{T-1} \mathrm{d}[\mathrm{F}_t]^\flat_{A_t},
\end{aligned}
$$

where $\bar{T} = \sum_{k=0}^{T} \binom{T}{k}$ and $\Delta_{\bar{T}}(\alpha_0) = \{\boldsymbol{\alpha}_0 \in [0, \alpha_0]^{\bar{T}} : \boldsymbol{\iota}_{\bar{T}}' \boldsymbol{\alpha}_0 = \alpha_0\}$, with $\boldsymbol{\iota}_{\bar{T}}$ denoting column vector of ones of length $\bar{T}$. The first equality exploits 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 maximisation 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 1, the third equality uses that the optimal test is constant in arguments varying in A^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 maximisation problem is equivalent to finding the optimal test $\phi_A^\flat$ for the testing problem $\mathbb{H}_j : p_j = \prod_{t=0}^{T-1} [f_j]^\flat_{A_t}$, $j \in \{0, 1\}$, 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 $h_1 \in \hbar$, also the most powerful test for testing problem (8). But, since the test $\phi^\flat$ is independent of h_1, it is the UMP test for testing problem (8). $\square$

A.7 Proof Corollary 3

We show that $\phi_A^\sharp$ is not UMP by a specific counterexample in which the power of $\phi_A^\sharp$ is strictly smaller than the power of $\phi_A^\flat$. In particular, suppose that $T = 1$ and consider two densities f_0 and f_1 that are different on $A = [r, \infty)$, for some constant $r > 0$. Furthermore, assume that

$$\int_{\{y:\lambda(y)>r\}}^{\infty} \mathrm{F}_0(\mathrm{d}y) > \alpha, \qquad \lambda(y) = \frac{f_1(y)}{f_0(y)}. \tag{B.1}$$

For $T = 1$, the likelihood ratios of the conditional and censored test simplify to

$$\begin{aligned}
\lambda_A^\sharp(y) &= \frac{[f_1]_A^\sharp(y)}{[f_0]_A^\sharp(y)} = \frac{\frac{f_1(y)}{\mathrm{F}_1(A)}}{\frac{f_0(y)}{\mathrm{F}_0(A)}} \mathbb{1}_A(y) = \frac{\mathrm{F}_0(A^c)}{\mathrm{F}_1(A^c)} \frac{f_1(y)}{f_0(y)} \mathbb{1}_A(y) \\
\lambda_A^\flat(y) &= \frac{[f_1]_A^\flat(y)}{[f_0]_A^\flat(y)} = \frac{f_1(y)}{f_0(y)} \mathbb{1}_A(y) + \frac{\mathrm{F}_1(A^c)}{\mathrm{F}_0(A^c)} \mathbb{1}_{A^c}(y).
\end{aligned}$$

Due to restriction (B.1), the corresponding critical regions $C^\sharp = \left[c^\sharp, \infty\right)$ and $C^\flat = \left[c^\flat, \infty\right)$ are both contained in A. [Hence, an example in which $\sharp$ has higher power than $\flat$, would not only be a counterexample to Theorem 3 but also to the fundamental lemma of Neyman and Pearson (1933).]

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 $y \sim \mathrm{Exp}(\theta_j)$, $j \in \{0, 1\}$, with $\theta_0 > \theta_1$. Then, the critical regions follow from the equation

$$\alpha = \int_{\{y:\lambda(y)>c^*\}}^{\infty} \theta_0 \mathrm{e}^{-\theta_0 y} \mathrm{d}y = \int_{\{y:a^*\left(\frac{\theta_1}{\theta_0}\right)\mathrm{e}^{-(\theta_1-\theta_0)y}>c^*\}}^{\infty} \theta_0 \mathrm{e}^{-\theta_0 y} \mathrm{d}y = 1 - F_0\left(\frac{1}{\theta_0 - \theta_1} \log\left(\frac{\theta_0}{\theta_1}\right) \frac{c^*}{a^*}\right),$$

where $a^\sharp = \frac{1-F_0(r)}{1-F_1(r)} = \mathrm{e}^{-(\theta_0-\theta_1)r}$ and $a^\flat = 1$. Isolating c^*, gives

$$c^* = ba^*, \qquad b = \frac{\theta_0}{\theta_1}\mathrm{e}^{(\theta_0-\theta_1)F_0^{-1}(1-\alpha)} > 0.$$

Now, the power of the conditional test is only weakly larger than the power of the censored test, if

$$\int_{\{y:\lambda(y)>c^\sharp\}}^{\infty} \theta_1 \mathrm{e}^{-\theta_1 y}\mathrm{d}y \geq \int_{\{y:\lambda(y)>c^\flat\}}^{\infty} \theta_1 \mathrm{e}^{-\theta_1 y}\mathrm{d}y \iff c^\sharp \geq c^\flat \iff (\theta_0-\theta_1)r \leq 0.$$

But then, as $\theta_0 > \theta_1$ and $r > 0$, it follows that the power of the conditional test is always strictly smaller than the power of the censored test. Consequently, the conditional test $\phi_A^\sharp$ is not UMP. $\square$

B Examples Table 2

B.1 LogS

$$\begin{aligned}
\mathrm{LogS}_w^\sharp(f,y) &= w(y)\log\left(\frac{w(y)f(y)}{1-\bar{F}_w}\right) \\
&\overset{\text{eqv.}}{=} w(y)\log\left(\frac{f(y)}{1-\bar{F}_w}\right) \\
&= S_w^{\mathrm{cl}}(f,y), \\
\mathrm{LogS}_w^\flat(f,y) &= w(y)\log\big(w(y)f(y)\mathbb{1}_{y\neq *} + \bar{F}_w\mathbb{1}_{y=*}\big) + \big(1-w(y)\big)\log\bar{F}_w \\
&\overset{\text{eqv.}}{=} w(y)\log\big(f(y)\mathbb{1}_{y\neq *} + \bar{F}_w\mathbb{1}_{y=*}\big) + \big(1-w(y)\big)\log\bar{F}_w \\
&\overset{\mu\text{-a.e.}}{=} w(y)\log f(y) + \big(1-w(y)\big)\log\bar{F}_w \\
&= S_w^{\mathrm{csl}}(f,y).
\end{aligned}$$

B.2 PsSphS_α

We start by showing the limit. Rescaling the PsSphS_α family by a factor $\frac{1}{\alpha-1}$, we obtain

$$\begin{aligned}
\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-1} &= \lim_{\alpha\downarrow 1}\frac{(\alpha-1)\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-1}}{(\alpha-1)^2}\\
&= \lim_{\alpha\downarrow 1}\frac{\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-1}+(\alpha-1)\left(\log\left(\frac{f(y)}{\|f\|_\alpha}\right)+(\alpha-1)\left(\frac{f(y)}{\|f\|_\alpha}\right)^{-1}\frac{\partial}{\partial\alpha}\frac{f(y)}{\|f\|_\alpha}\right)\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-1}}{2(\alpha-1)}\\
&= \frac{1}{2}\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-1}+\frac{1}{2}\lim_{\alpha\downarrow 1}\log\left(\frac{f(y)}{\|f\|_\alpha}\right)\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-1}\\
&\quad+\frac{1}{2}\lim_{\alpha\downarrow 1}(\alpha-1)\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-2}\frac{\partial}{\partial\alpha}\frac{f(y)}{\|f\|_\alpha},
\end{aligned}$$

and hence

$$\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\left(\frac{f(y)}{\|f\|_\alpha}\right)^{\alpha-1} = \log f(y), \tag{B.2}$$

since $\|f\|_1 = 1$. It might be helpful to note that the second equality in the first display follows from L'Hôpital's rule combined with the following derivative

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

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 close similarity with Equation (B.2), it is uncomplicated to obtain the following limit

$$\begin{aligned}\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\mathrm{PsSphS}^{\sharp}_{\alpha,w}(f,y) &= w(y)\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\left(\frac{f_w(y)}{\|f_w\|_\alpha}\right)^{\alpha-1}\\ &= w(y)\log\left(\frac{f_w(y)}{\|f_w\|_1}\right)\\ &= w(y)\log f^{\sharp}_w(y)\\ &= \mathrm{LogS}^{\sharp}_w(f,y),\end{aligned}$$

since $\|f_w\|_1 = \int_{\mathcal{Y}} wf\,\mathrm{d}\mu = 1-\bar{F}_w$. Clearly, this result also follows directly from the linearity of limits, as

$$\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\mathrm{PsSphS}^{\sharp}_{\alpha}(f,y) = w(y)\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\mathrm{PsSphS}_{\alpha}(f^{\sharp}_w,y) = w(y)\log f^{\sharp}_w(y) = \mathrm{LogS}^{\sharp}_w(f,y). \tag{B.3}$$

Moreover, for the censored PsSphS_α family, it follows that

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

For the limit of $\alpha\downarrow 1$, we cannot directly apply Equation (B.2) as we did for the conditional case. Nevertheless, we obtain a similarly satisfying result, namely

$$
\begin{aligned}
\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\mathrm{PsSphS}_w^\flat(f,y) &= w(y)\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\left(\frac{f_w(y)}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)^{\alpha-1} \\
&\quad + \big(1-w(y)\big)\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\left(\frac{\bar{F}_w}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)^{\alpha-1} \\
&= w(y)\left(\lim_{\alpha\downarrow 1}\log\left(\frac{f_w(y)}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)\left(\frac{f_w(y)}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)^{\alpha-1}\right. \\
&\quad \left. + \lim_{\alpha\downarrow 1}(\alpha-1)\left(\frac{f_w(y)}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)^{\alpha-2}\frac{\partial}{\partial\alpha}\frac{f_w(y)}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right) \\
&\quad + \big(1-w(y)\big)\left(\lim_{\alpha\downarrow 1}\log\left(\frac{\bar{F}_w}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)\left(\frac{\bar{F}_w}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)^{\alpha-1}\right. \\
&\quad \left. + \lim_{\alpha\downarrow 1}(\alpha-1)\left(\frac{\bar{F}_w}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right)^{\alpha-2}\frac{\partial}{\partial\alpha}\frac{\bar{F}_w}{\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\right)^{\frac{1}{\alpha}}}\right) \\
&= w(y)\log f_w(y) + \big(1-w(y)\big)\log\bar{F}_w \\
&= \mathrm{LogS}_w^\flat(f,y),
\end{aligned}
$$

where we have used that $\|f_w\|_1+\bar{F}_w = 1-\bar{F}_w+\bar{F}_w = 1$.

B.3 $\mathbf{PowS}_\alpha$

We start by verifying the limit in Equation (6) for the non-focused family. Specifically,

$$
\begin{aligned}
\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\mathrm{PowS}_\alpha &= \lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\left(\alpha f(y)^{\alpha-1}-(\alpha-1)\|f\|_\alpha^\alpha\right) \\
&= \lim_{\alpha\downarrow 1}\frac{(\alpha-1)\alpha f(y)^{\alpha-1}}{(\alpha-1)^2}-1 \\
&= \lim_{\alpha\downarrow 1}\frac{\alpha f(y)^{\alpha-1}+(\alpha-1)f(y)^{\alpha-1}\big(1+\alpha\log f(y)\big)}{2(\alpha-1)}-1 \\
&= \frac{1}{2}\left(\lim_{\alpha\downarrow 1}\frac{1}{\alpha-1}\alpha f(y)^{\alpha-1}-1\right)+\frac{1}{2}\left(\lim_{\alpha\downarrow 1}f(y)^{\alpha-1}\big(1+\alpha\log f(y)\big)-1\right)
\end{aligned}
$$

and hence

$$\lim_{\alpha \downarrow 1} \frac{1}{\alpha - 1} \mathrm{PowS}_\alpha(f, y) = \log f(y).$$

Furthermore, the conditional version of the PowS_α family displayed in Table 2 is nothing but a direct application of the conditioning procedure introduced in Example 1. For the limit of the $\mathrm{PowS}^\sharp_{\alpha,w}$, we recall Equation (B.3) and immediately conclude that $\lim_{\alpha\downarrow 1} \frac{1}{\alpha-1}\mathrm{PowS}^\sharp_{\alpha,w}(f, y) = \mathrm{LogS}^\sharp_w(f, y)$.

Turning to the censored focusing method, we recall from the analysis in Appendix B.2 that $\|f^\flat_w\|^\alpha_\alpha = \|f_w(y)\|^\alpha_\alpha + \bar{F}^\alpha_w$. Using this result, we obtain

$$\begin{aligned}
\mathrm{PowS}^\flat_{\alpha,w}(f, y) &= w(y)\alpha \left(f_w(y)\mathbb{1}_{y\neq *} + \bar{F}_w \mathbb{1}_{y=c}\right)^{\alpha-1} + \big(1 - w(y)\big)\alpha \bar{F}^{\alpha-1}_w - (\alpha - 1)\|f^\flat_w\|^\alpha_\alpha \\
&\overset{\mu\text{-a.e.}}{=} w(y)\alpha f_w(y)^{\alpha-1} + \big(1 - w(y)\big)\alpha \bar{F}^{\alpha-1}_w - (\alpha - 1)\left(\|f_w\|^\alpha_\alpha + \bar{F}^\alpha_w\right),
\end{aligned}$$

which bears the following limit

$$\begin{aligned}
\lim_{\alpha\downarrow 1} \frac{1}{\alpha - 1}\mathrm{PowS}^\flat_{\alpha,w}(f, y) &= w(y) \lim_{\alpha\downarrow 1} \frac{(\alpha - 1)\alpha f_w(y)^{\alpha-1}}{(\alpha - 1)^2} + \big(1 - w(y)\big) \lim_{\alpha\downarrow 1} \frac{(\alpha - 1)\alpha \bar{F}^{\alpha-1}_w}{(\alpha - 1)^2} - 1 \\
&= \frac{1}{2} \lim_{\alpha\downarrow 1} \left(w(y)\left(\frac{1}{\alpha - 1}\alpha f_w(y)^{\alpha-1} - 1\right) + \big(1 - w(y)\big)\left(\frac{1}{\alpha - 1}\alpha \bar{F}^{\alpha-1}_w - 1\right)\right) \\
&\quad + \frac{1}{2} \lim_{\alpha\downarrow 1} \Bigg(w(y)\left(f_w(y)^{\alpha-1}\big(1 + \alpha \log f_w(y)\big) - 1\right) \\
&\quad + \big(1 - w(y)\big)\left(\bar{F}^{\alpha-1}_w\big(1 + \alpha \log \bar{F}_w\big) - 1\right)\Bigg).
\end{aligned}$$

Therefore,

$$\lim_{\alpha\downarrow 1} \frac{1}{\alpha - 1}\mathrm{PowS}^\flat_{\alpha,w}(f, y) = w(y) \log f_w(y) + \big(1 - w(y)\big) \log \bar{F}_w = \mathrm{LogS}^\flat_w(f, y).$$

C Examples Table 3

C.1 LogS

We start with the derivation based on the density h. In particular,

$$\begin{aligned}
\mathrm{LogS}^{\flat}_{w,h}(f,y) &= w(y)\log f^{\flat}_{w,h}(y) + \big(1-w(y)\big)\int_{\mathcal{Y}} \log f^{\flat}_{w,h}(q)h(q)\mathrm{d}q, \\
&= w(y)\Big(\log\big(f_w(y)\big)\mathbb{1}_{w>0} + \log\big(\bar{F}_w h(y)\big)\mathbb{1}_{w=0}\Big) \\
&\quad\; + \big(1-w(y)\big)\int_{\{w=0\}} \Big(\log\big(f_w(q)\big)\mathbb{1}_{w>0} + \log\big(\bar{F}_w h(q)\big)\mathbb{1}_{w=0}\Big)h(q)\mathrm{d}q, \\
&= w(y)\log f_w(y) + \big(1-w(y)\big)\int_{\{w=0\}} \log\big(\bar{F}_w h(q)\big)\,h(q)\mathrm{d}q, \\
&\overset{\text{eqv.}}{=} w(y)\log f(y) + \big(1-w(y)\big)\log \bar{F}_w, \\
&= S^{\mathrm{csl}}(f,y).
\end{aligned}$$

As this generalised censored scoring rule does not depend on the nuisance density h, we obtain the same result for the specific choice $h = u$.

C.2 PsSphS$_\alpha$

As illustrated more explicitly in C.1, the simplification below follows predominantly from the observation that $w(y)h(y) = 0$, $\forall y \in \mathcal{Y}$ and $f_w(y) = 0, \forall y \in \{w = 0\}$. More specifically,

$$\begin{aligned}
\mathrm{PsSphS}^{\flat}_{w,h}(f,y) &= w(y)\frac{\big(f_w(y)+\bar{F}_w h(y)\big)^{\alpha-1}}{\Big(\int_{\mathcal{Y}}\big(f_w(y)+\bar{F}_w h(y)\big)^{\alpha}\,\mathrm{d}y\Big)^{\frac{\alpha-1}{\alpha}}}, \\
&\quad + \big(1-w(y)\big)\frac{\int_{\{w=0\}}\big(f_w(q)+\bar{F}_w h(q)\big)^{\alpha-1}\,h(q)\mathrm{d}q}{\Big(\int_{\mathcal{Y}}\big(f_w(y)+\bar{F}_w h(y)\big)^{\alpha}\,\mathrm{d}y\Big)^{\frac{\alpha-1}{\alpha}}}, \\
&= \frac{w(y)f_w(y)^{\alpha-1} + \big(1-w(y)\big)\bar{F}_w^{\alpha-1}\|h\|_\alpha^\alpha}{\big(\|f_w\|_\alpha^\alpha + \bar{F}_w^\alpha\|h\|_\alpha^\alpha\big)^{\frac{\alpha-1}{\alpha}}}.
\end{aligned}$$

The expression for $\mathrm{PsSphS}^{\flat}_{w,u}(f,y)$ in Table 3 follows directly from the observation that $\|\frac{1}{\theta}\mathbb{1}_{(r,r+\theta)}\|_\alpha^\alpha = \theta^{1-\alpha}$.

C.3 PowS$_\alpha$

Using the hints in Appendix C.2, we easily arrive at

$$\begin{aligned}\mathrm{Pow}^\flat_{w,h}(f,y) &= w(y)\alpha\left(f_w(y)+\bar{F}_w h(y)\right)^{\alpha-1}\\ &\qquad + \big(1-w(y)\big)\alpha\int_{\{w=0\}}\left(f_w(q)+\bar{F}_w h(q)\right)^{\alpha-1}h(q)\mathrm{d}q\\ &\qquad - (\alpha-1)\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\|h\|_\alpha^\alpha\right)\\ &= w(y)\alpha f_w(y)^{\alpha-1}+\big(1-w(y)\big)\alpha\bar{F}_w^{\alpha-1}\|h\|_\alpha^\alpha\\ &\qquad - (\alpha-1)\left(\|f_w\|_\alpha^\alpha+\bar{F}_w^\alpha\|h\|_\alpha^\alpha\right).\end{aligned}$$

Like in Appendix C.2, the expression for $\mathrm{Pow}^\flat_{w,u}(f,y)$ in Table 3 follows directly from the observation that $\|\frac{1}{\theta}\mathbb{1}_{(r,r+\theta)}\|_\alpha^\alpha=\theta^{1-\alpha}$.

C.4 CRPS

The additional assumption on the weight functions for the CRPS, that is, $\{w>0\}=(-\infty,r]$ and $\{w=0\}=(r,\infty)$ yields that $H(y)=0,\forall y\in\{w>0\}$ and $F_w(y)=1-\bar{F}_w,\forall y\in\{w=0\}$. Consequently, we arrive at the following intuitive expression for the censored CRPS

$$\begin{aligned}&\mathrm{CRPS}^\flat_{w,H}(F,y)\\ &= w(y)\mathrm{CRPS}(F^\flat_w,y)+\big(1-w(y)\big)\int_r^\infty \mathrm{CRPS}(F^\flat_w,q)\mathrm{d}H(q)\\ &= w(y)\int_{-\infty}^\infty\big(F^\flat_w(s)-\mathbb{1}_{(-\infty,s]}(y)\big)^2\mathrm{d}s+\big(1-w(y)\big)\int_r^\infty\int_{-\infty}^\infty\big(F^\flat_w(s)-\mathbb{1}_{(-\infty,s]}(q)\big)^2\mathrm{d}s\mathrm{d}H(q)\\ &= w(y)\left(\int_{-\infty}^r\big(F_w(s)-\mathbb{1}_{(-\infty,s]}(y)\big)^2\mathrm{d}s+\int_r^\infty\big(\bar{F}_w(H(s)-1)+\mathbb{1}_{(s,\infty)}(y)\big)^2\mathrm{d}s\right)\\ &\qquad+\big(1-w(y)\big)\left(\int_{-\infty}^r\big(F_w(s)\big)^2\mathrm{d}s+\int_r^\infty\int_r^\infty\big(\bar{F}_w(H(s)-1)+\mathbb{1}_{(s,\infty)}(q)\big)^2\mathrm{d}s\mathrm{d}H(q)\right)\\ &= w(y)\int_{-\infty}^r\big(F_w(s)-\mathbb{1}_{(-\infty,s]}(y)\big)^2\mathrm{d}s+\big(1-w(y)\big)\int_{-\infty}^r\big(F_w(s)\big)^2\mathrm{d}s\\ &\qquad+\bar{F}_w^2\|H-1\|_2^2-2\big(1-w(y)\big)\bar{F}_w\|H-1\|_2^2\\ &\overset{\text{eqv.}}{=}\int_{-\infty}^r\big(F_w(s)-w(y)\mathbb{1}_{(-\infty,s]}(y)\big)^2\mathrm{d}s+\|H-1\|_2^2\big(\bar{F}_w-(1-w(y)\big)^2\end{aligned}$$

For the uniform nuisance distribution, with distribution function $U(y) = (y-r)\mathbb{1}_{(r,r+\theta)} + \mathbb{1}_{[r+\theta,\infty)}$, it follows that

$$\mathrm{CRPS}^{\flat}_{w,U}(F,y) = \int_{-\infty}^{r} \big(F_w(s) - w(y)\mathbb{1}_{(-\infty,s]}(y)\big)^2 \mathrm{d}s + \frac{\theta}{3}\big(\bar{F}_w - (1-w(y))\big)^2$$

Furthermore, if we additionally assume that $w(y) = \mathbb{1}_{(-\infty,r]}(y)$, the scoring rule simplifies to

$$\mathrm{CRPS}^{\flat}_{w,U}(F,y) = \int_{-\infty}^{\infty} \mathbb{1}_{(-\infty,r]}(y)\big(F(s) - \mathbb{1}_{(-\infty,s]}(y)\big)^2 \mathrm{d}s + \frac{\theta}{3}\big(\bar{F}_w - \mathbb{1}_{(r,\infty)}(y)\big)^2$$

C.5 CRPS for $w(y) = \mathbb{1}_{[a,b]}(y)$

For the compact indicator weight function $w(y) = \mathbb{1}_{[a,b]}(y)$, $a \leq b$, with $a, b \in \mathbb{R}$ and Unif(b,$b+\theta$) nuisance distribution, where $\theta \in \mathbb{R}_{++}$, we obtain the following censored

CRPS

$$
\begin{aligned}
&\mathrm{CRPS}_w^\flat(\mathrm{F}, y)\\
&= \mathbb{1}_{[a,b]}(y)\mathrm{CRPS}(\mathrm{F}_w^\flat, y) + \big(1 - \mathbb{1}_{[a,b]}(y)\big)\frac{1}{\theta}\int_b^{b+\theta} \mathrm{CRPS}(\mathrm{F}_w^\flat, q)\mathrm{d}q\\
&= \mathbb{1}_{[a,b]}(y)\int_{-\infty}^{\infty} \big(F_w^\flat(s) - \mathbb{1}_{(-\infty,s]}(y)\big)^2\mathrm{d}s + \big(1 - \mathbb{1}_{[a,b]}(y)\big)\int_b^{b+\theta}\int_{-\infty}^{\infty} \big(F_w^\flat(s) - \mathbb{1}_{(-\infty,s]}(q)\big)^2\mathrm{d}s\mathrm{d}q\\
&= \mathbb{1}_{[a,b]}(y)\int_{-\infty}^{\infty} \left(F_w(s) + \bar{F}_w\left(\frac{s-b}{\theta}\mathbb{1}_{(b,b+\theta)}(s) + \mathbb{1}_{[b+\theta,\infty)}(s)\right) - \mathbb{1}_{(-\infty,s]}(y)\right)^2\mathrm{d}s\\
&\quad + \big(1 - \mathbb{1}_{[a,b]}(y)\big)\frac{1}{\theta}\int_b^{b+\theta}\int_{-\infty}^{\infty} \left(F_w(s) + \bar{F}_w\left(\frac{s-b}{\theta}\mathbb{1}_{(b,b+\theta)}(s) + \mathbb{1}_{[b+\theta,\infty)}(s)\right) - \mathbb{1}_{(-\infty,s]}(q)\right)^2\mathrm{d}s\mathrm{d}q\\
&= \mathbb{1}_{[a,b]}(y)\left(\int_a^b \big(F(s) - \mathbb{1}_{(-\infty,s]}(y)\big)^2\mathrm{d}s + \bar{F}_w^2\int_b^{b+\theta}\left(\frac{s-b}{\theta} - 1\right)^2\mathrm{d}s\right)\\
&\quad + \big(1 - \mathbb{1}_{[a,b]}(y)\big)\frac{1}{\theta}\left(\int_b^{b+\theta}\left(\int_a^b \big(F(s)\big)^2\mathrm{d}s + \int_b^{b+\theta}\left(\bar{F}_w\left(\frac{s-b}{\theta} - 1\right) + \mathbb{1}_{(s,\infty)}(q)\right)^2\mathrm{d}s\right)\mathrm{d}q\right)\\
&= \mathbb{1}_{[a,b]}(y)\left(\int_{-\infty}^{\infty}\mathbb{1}_{[a,b]}(s)\big(F(s) - \mathbb{1}_{(-\infty,s]}(y)\big)^2\mathrm{d}s + \frac{\theta}{3}\bar{F}_w^2\right)\\
&\quad + \big(1 - \mathbb{1}_{[a,b]}(y)\big)\left(\int_{-\infty}^{\infty}\mathbb{1}_{[a,b]}(s)\big(F(s) - \mathbb{1}_{(-\infty,s]}(y) - \mathbb{1}_{(-\infty,a]}(y)\big)^2\mathrm{d}s + \frac{\theta}{3}\bar{F}_w^2\right.\\
&\quad \left. + \frac{1}{\theta}\int_b^{b+\theta}\int_b^{b+\theta}\bar{F}_w\left(\frac{s-b}{\theta} - 1\right)\mathbb{1}_{(s,\infty)}(q)\mathrm{d}s\mathrm{d}q + \frac{1}{\theta}\int_b^{b+\theta}\int_b^{b+\theta}\mathbb{1}_{(s,\infty)}(q)\mathrm{d}s\mathrm{d}q\right)\\
&= \int_{-\infty}^{\infty}\mathbb{1}_{[a,b]}(s)\big(F(s) - \mathbb{1}_{(-\infty,s]}(y)\big)^2\mathrm{d}s + \frac{\theta}{3}\bar{F}_w^2 + \mathbb{1}_{(-\infty,a]}(y)\left(b - a - 2\int_a^b \big(F(s) - \mathbb{1}_{(-\infty,s]}(y)\big)\mathrm{d}s\right)\\
&\quad + \big(1 - \mathbb{1}_{[a,b]}(y)\big)\left(-\frac{\theta}{3}\bar{F}_w + \frac{\theta}{2}\right)\\
&\overset{\text{eqv.}}{=} \int_{-\infty}^{\infty}\mathbb{1}_{[a,b]}(s)\big(F(s) - \mathbb{1}_{(-\infty,s]}(y)\big)^2\mathrm{d}s + \frac{\theta}{3}\bar{F}_w^2 - \frac{\theta}{3}\bar{F}_w\big(1 - \mathbb{1}_{[a,b]}(y)\big) - 2\mathbb{1}_{(-\infty,a]}(y)\int_a^b F(s)\mathrm{d}s\\
&\overset{\text{eqv.}}{=} \mathrm{twCRPS}(F, y) + \frac{\theta}{3}\Big(\bar{F}_w - \big(1 - \mathbb{1}_{[a,b]}(y)\big)\Big)^2 - 2\mathbb{1}_{(-\infty,a]}(y)\int_a^b F(s)\mathrm{d}s,
\end{aligned}
$$

where twCRPS denotes the weighted CRPS scoring rule proposed by Gneiting and Ranjan (2011). Although the twCRPS scoring rule remains proper for every weight function, this scoring rule fails to be strictly locally proper if the class of weight functions contains indicators on compact intervals (Holzmann and Klar, 2017a). Therefore, it is unsurprising that the censored CRPS, which is strictly locally proper relative to the class of distributions and weight functions for which it delivers a scoring rule (Definition 1) by Theorem 2, is not equivalent to the twCRPS scoring rule. The 'correction term' consists of two parts. The first one is simply the Brier score of the probability forecast $\bar{F}_w = \int_{\mathcal{Y}}(1 - \mathbb{1}_{[a,b]}(y))\mathrm{d}F(y)$ of the event $y \notin [a, b]$, multiplied by $\frac{\theta}{3}$. This additional term is somewhat similar to the additional log-score applied to the discrete probability assigned to the complement of the region of interest in the csl rule, necessary for strict local propriety. Furthermore, the factor $\frac{\theta}{3}$ determines the weight with which the Brier score of the complement of the region of interest enters the scoring rule.

The third term can be seen as a by-product of the specific choice of the location of the uniform distribution, which puts the censored distribution function to zero on $(-\infty, a]$. Notably, this term drops out of the equation in the popular left-tail application, i.e. $a \downarrow -\infty$, in which case the censored CRPS reduces to

$$\mathrm{CRPS}^{\flat}_{w}(\mathrm{F}, y) \overset{\text{eqv.}}{=} \mathrm{twCRPS}(F, y) + \frac{\theta}{3}\big(\bar{F}_w - \mathbb{1}_{(b,\infty)}(y)\big)^2.$$

According to Theorem 3 of Holzmann and Klar (2017a), the twCRPS scoring rule is also strictly locally proper without the second term if $a \downarrow -\infty$. Moreover, right-tail applications for which $b \uparrow \infty$ are obviously not compatible with the current choice of the location of the uniform reference distribution.

References

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 mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Histoire de l'Acad., Roy. Sci.(Paris) avec Mem*, 1–45.

Bregman, L. M. (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. 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.

Dawid, A. P. (1984). Present position and potential developments: Some personal views statistical theory the prequential approach. *Journal of the Royal Statistical Society: Series A (General) 147*(2), 278–290.

Dawid, A. P. (2007). The geometry of proper scoring rules. *Annals of the Institute of Statistical Mathematics 59*(1), 77–93.

Diks, C. and H. Fang (2020). Comparing density forecasts in a risk management context. *International Journal of Forecasting 36*(2), 531–551.

Diks, C., V. Panchenko, O. Sokolinskiy, and D. van Dijk (2014). Comparing the accuracy of multivariate density forecasts in selected regions of the copula support. *Journal of Economic Dynamics and Control 48*, 79–94.

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.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 222*(594-604), 309–368.

Gatarek, L. T., L. F. Hoogerheide, K. Hooning, and H. K. van Dijk (2013). Censored posterior and predictive likelihood in bayesian left-tail prediction for accurate value at risk estimation.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Gneiting, T. and A. 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.

Good, I. (1952). Rational decisions. *Journal of the Royal Statistical Society 14*, 107–114.

Good, I. (1971). Comment on "measuring information and uncertainty.". *Foundation of Statistical Inference*, 265–273.

Holzmann, H. and B. Klar (2016). Weighted scoring rules and hypothesis testing. Available at arXiv:1611.07345v2.

Holzmann, H. and B. Klar (2017a). Focusing on regions of interest in forecast evaluation. *Annals of applied statistics 11*(4), 2404–2431.

Holzmann, H. and B. Klar (2017b). Supplementary material to: Focusing on regions of interest in forecast evaluation.

Iacopini, M., F. Ravazzolo, and L. Rossini (2022). Proper scoring rules for evaluating density forecasts with asymmetric loss functions. *Journal of Business & Economic Statistics*, 1–15.

Jose, V. R. (2009). A characterization for the spherical scoring rule. *Theory and Decision 66*(3), 263–281.

Jose, V. R. R., R. F. Nau, and R. L. Winkler (2008). Scoring rules, generalized entropy, and utility maximization. *Operations research 56*(5), 1146–1157.

Kole, E., T. Markwat, A. Opschoor, and D. Van Dijk (2017). Forecasting value-at-risk under temporal and portfolio aggregation. *Journal of Financial Econometrics 15*(4), 649–677.

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*, 106–127.

Liese, F. and I. Vajda (2006). On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory 52*(10), 4394–4412.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. pp. 621–661. Elsevier.

Neyman, J. and E. S. 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.

Opschoor, A., D. van Dijk, and M. van der Wel (2017). Combining density forecasts using focused scoring rules. *Journal of Applied Econometrics 32*(7), 1298–1313.

Ovcharov, E. (2018). Proper scoring rules and bregman divergence. *Bernoulli 24*(1), 53–79.

Painsky, A. and G. W. Wornell (2019). 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.

Roby, T. B. (1964). Belief states: A preliminary empirical study. Technical report, TUFTS UNIV MEDFORD MA.

Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association 66*(336), 783–801.

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics 1*(1), 43–61.

Shuford, E. H., A. Albert, and H. E. Massengill (1966). Admissible probability measurement procedures. *Psychometrika 31*(2), 125–145.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 24–36.

Toda, M. (1963). Measurement of subjective probability distribution. rep. no. 3, 1963, state college, pennsylvania. *Institute for Research*.