Supplementary Material on "Localizing Strictly Proper Scoring Rules"

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

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

Roger J. A. Laeven
Department of Quantitative Economics
University of Amsterdam and Tinbergen Institute

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

October 13, 2023

*Corresponding author. Mailing Address: PO Box 15867, 1001 NJ Amsterdam, The Netherlands. Phone: +31 (0) 20 525 4252. Email: r.f.a.depunder@uva.nl

Contents

A Proofs

A.1 Proof Theorem 2

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

Lemma 1. *Consider the censored scoring rule defined in Definition 5.* $\forall w \in \mathcal{W}$ *and* $\mathrm{H} \in \mathcal{H}$*, 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 2. *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 5. 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

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

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

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

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

Corollary 1. *The censored scoring rule defined in Definition 5 is localising* $\forall \mathrm{H} \in \mathcal{H}$.

Proof. Suppose that $\mathrm{F}(E \cap \{w > 0\}) = \mathrm{G}(E \cap \{w > 0\}), \ \forall E \in \mathcal{G}$. Then, by Lemma 2,

$F^{\flat}_{w,H}(E) = G^{\flat}_{w,H}(E), \forall E \in \mathcal{G}$, whence it follows that $S^{\flat}_{w,H}(P, y) = S^{\flat}_{w,H}(F, y), \forall y \in \mathcal{Y}$. □

We now turn to the main body of the proof. The definition of a strictly locally proper scoring rule (Definition 3) and the definitions on which this definition is built, that is, the definition of a locally proper scoring rule (Definition 3) and a localising weighted scoring rule (Definition 2), reveal that we need to prove a list of three things $\forall H \in \mathcal{H}$: (i) $S^{\flat}_{w,H}(P, y)$ must be localising relative to $\mathcal{W}$, (ii) $S^{\flat}_{w,H}(P, y)$ must be proper relative to $\mathcal{P}$, $\forall w \in \mathcal{W}$ and (iii) the if and only if statement in Definition 3. We prove them one by one.

(i) $S^{\flat}_{w,H}(P, y)$ is localising relative to $\mathcal{W}$, $\forall H \in \mathcal{H}$, by Corollary 1.

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

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

which is by definition of the class $\mathcal{P}^{\flat}_{w,H} \equiv \{[P]^{\flat}_{w,H}, P \in \mathcal{P}\}$ equivalent to

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

and hence, by Lemma 1, also

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

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

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

and $\mathrm{H} \in \mathcal{H}$,

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

and thus, by Lemma 2,

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

$\forall E \in \mathcal{G}$, and hence, by Lemma 1, 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 3.

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

A.2 Proof Theorem 3

We start by rephrasing the hypotheses. 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 summarized 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.

Explicitizing 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

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

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

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

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

In terms of the index set of all observations $\mathcal{I} = \{1, \ldots, T\}$, this space can also be denoted as $\mathcal{Y}(\mathcal{I}) = \prod_{t \in \mathcal{I}} \mathcal{Y}_t$. So, the aim is to find a uniformly most powerful (UMP) test ϕ^* of size α for testing problem (6), i.e. a solution to the maximization problem

$$\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{p_1} \phi, \qquad \Phi(\alpha) = \{\phi : \sup_{p_0 \in \mathcal{P}_0} \mathbb{E}_{p_0} \phi \leq \alpha\}. \tag{A.1}$$

Now fix an $h_1 \in \mathcal{H}$ so that the distribution under the alternative is completely known. Given the fact that the hypotheses are, in the end, silent about the shape of the 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 (A.1) has integrated out the dependence on the nuisance densities.

Although it is obvious that marginalizing out the still assumed to be fixed density $h_1 \in \mathcal{H}$ is harmless in terms of power, it is non-trivial that this is an affordable strategy in terms of size for all $h_0 \in \mathcal{H}$. Lemma 3 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 2 then formalises the idea that we can restrict our attention to tests of the conjectured kind.

Lemma 3. *Consider testing problem* (6) *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* $h_1 \in \mathcal{H}$*, the test*

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

where $\phi_{h_1}^*$ *denotes a solution to problem* (A.1)*, is such that* $\psi_{h_1} \in \Phi(\alpha)$*.*

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

$\forall h_1 \in \mathcal{H}$, we have that

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

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

Corollary 2. *Consider testing problem* (6) *and suppose that the outcomes* $(y_t)_{t \in \mathcal{I}_A}$ *are in* A_t*, and the remaining* $T - k$*, with* $k = |\mathcal{I}_A|$*, observations* $(y_t)_{t \in \mathcal{I}_{A^c}}$ *in* A_t^c*. 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,*

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

Proof. Fix an arbitrary $h_1 \in \mathcal{H}$. Since $\Psi(\alpha) \subseteq \Phi(\alpha)$, we trivially have that $\max_{\phi \in \Phi(\alpha)} \mathbb{E}_{p_1} \phi \geq \max_{\psi \in \Psi(\alpha)} \mathbb{E}_{p_1} \psi$. 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 3, $\tilde{\psi} \in \Psi(\alpha)$, 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. □

For any fixed $h_1 \in \mathcal{H}$, the reduced optimization problem resulting from Corollary 2, simplifies to a simple versus simple hypothesis in terms of the censored measures $\mathrm{d}[\mathrm{F}_{jt}]^\flat_{A_t} = \mathbb{1}_{A_t}\mathrm{dF}_{jt} + \mathrm{F}_{jt}(A_t^c)\mathrm{d}\delta_*$, allowing us to apply Neyman and Pearson (1933).

Specifically, for any fixed $h_1 \in \mathcal{H}$, the most powerful test of size α is a solution to the following restricted maximization 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 maximization problem can be split into finding an optimal test on each of the partitioned parts conditional on the amount of size spent on each part and the optimal distribution of size over the partition of the outcome space.

The second equality holds by Corollary 2, 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 maximization problem is equivalent to finding the optimal test $\phi_A^\flat$ for the testing problem $\mathbb{H}_j : p_j = \prod_{t=0}^{T-1} [f_j]_{A_t}^\flat$, $j \in \{0, 1\}$, for which $\phi_A^\flat$ is the UMP test by the Fundamental Lemma of Neyman and Pearson (1933). By the equivalence, $\phi_A^\flat$ is, for any $h_1 \in \mathcal{H}$, also the most powerful test for testing problem (6). But, since the test $\phi^\flat$ is independent of h_1, it is the UMP test for testing problem (6). □

B Additional Proofs

B.1 Proof censored density

We defined the measures μ and F to the extended measurable space $(\mathcal{Y}^*, \mathcal{G}^*)$ by putting $\mu^*(E) = \mu(E \backslash \{*\})$, if $* \in E$ and $\mu^*(E) = \mu(E)$, otherwise, $\forall E \in \mathcal{G}^*$. To simplify the notation, we drop the subscript $*$ in the notation of the extended measures, while still considering all measures with respect to the extended measurable space $(\mathcal{Y}^*, \mathcal{G}^*)$

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

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

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

We can simplify this density as follows. Understanding that

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

we recall from the Radon-Nikodym theorem that $\frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\delta_*)}$ is the solution 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 $\frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\delta_*)} = \mathbb{1}_{\mathcal{Y}\setminus\{*\}}$ is the unique solution for the Radon-Nikodym derivative. By the same token, we conclude from

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

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

Put together, we arrive at

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

where f denotes the μ-density of F. □

B.2 Proof Corollary 2

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

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

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

B.3 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

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

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 (A.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

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

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. $\square$

C Derivations Table 1

The results in Table 1 hold under the assumption that all $\mathrm{F} \in \mathcal{P}$ are Borel measures on $\mathbb{R}^d$ satisfying $\mathrm{F}(r) = 0$, with $r \in \mathbb{R}^d$. Furthermore, the assumption on h and w in the generalised censored scoring rule examples can predominantly be simplified due to the observation that $w(y)h(y) = 0, \forall y \in \mathcal{Y}$ and $f_w(y) = 0, \forall y \in \{w = 0\}$.

C.1 LogS

$$
\begin{aligned}
\mathrm{LogS}(\tilde{f},\tilde{y}) &= \log \tilde{f}(\tilde{y}) = \log f(y) - \log|b| \overset{\text{eqv.}}{=} \log f(y). \\
\mathrm{LogS}^{\sharp}_{w}(f,y) &= w(y)\log\left(\frac{w(y)f(y)}{1-\bar{F}_w}\right) \\
&\overset{\text{eqv.}}{=} w(y)\log\left(\frac{f(y)}{1-\bar{F}_w}\right) \\
&= S^{\text{cl}}_{w}(f,y), \\
\mathrm{LogS}^{\flat}_{w}(f,y) &= w(y)\log\big(w(y)f(y)\mathbb{1}_{y\neq r} + \bar{F}_w\mathbb{1}_{y=r}\big) + \big(1-w(y)\big)\log\bar{F}_w \\
&\overset{\mu\text{-a.e.}}{=} w(y)\log\big(w(y)f(y)\big) + \big(1-w(y)\big)\log\bar{F}_w \\
&\overset{\text{eqv.}}{=} w(y)\log\big(f(y)\big) + \big(1-w(y)\big)\log\bar{F}_w \\
&= S^{\text{csl}}_{w}(f,y). \\
\mathrm{LogS}^{\flat}_{w,h}(f,y) &= w(y)\log f^{\flat}_{w,h}(y) + \big(1-w(y)\big)\int_{\mathcal{Y}} \log f^{\flat}_{w,h}(q)h(q)\mathrm{d}q, \\
&= w(y)\Big(\log\big(f_w(y)\big)\mathbb{1}_{w>0} + \log\big(\bar{F}_w h(y)\big)\mathbb{1}_{w=0}\Big) \\
&\qquad + \big(1-w(y)\big)\int_{\{w=0\}}\Big(\log\big(f_w(q)\big)\mathbb{1}_{w>0} + \log\big(\bar{F}_w h(q)\big)\mathbb{1}_{w=0}\Big)h(q)\mathrm{d}q, \\
&= w(y)\log f_w(y) + \big(1-w(y)\big)\int_{\{w=0\}}\log\big(\bar{F}_w h(q)\big)\,h(q)\mathrm{d}q, \\
&\overset{\text{eqv.}}{=} w(y)\log f(y) + \big(1-w(y)\big)\log\bar{F}_w, \\
&= S^{\text{csl}}(f,y).
\end{aligned}
$$

C.2 PsSphS$_\alpha$

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

Next, we show 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

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

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 (A.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

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

Moreover, for the censored PsSphS_α family, it follows that

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

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

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

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

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

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

as

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

Next, we verify the limit for the non-focused family. Specifically,

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

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 1 is nothing but a direct application of the conditioning procedure. For the limit of the $\mathrm{PowS}^\sharp_{\alpha,w}$, we recall Equation (A.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 C.2

that $\|f_w^\flat\|_\alpha^\alpha = \|f_w(y)\|_\alpha^\alpha + \bar{F}_w^\alpha$. Using this result, we obtain

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

which bears the following limit

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

Therefore,

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

D Examples

D.1 Localised NP for $T = 1$

Consider the special case $T = 1$. For one observation, it is straightforward to derive a most powerful test on A^c. For any $h_1 \in \mathcal{H}$, the relevant maximization problem 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

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

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

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

D.2 CRPS

$$\begin{aligned}
\mathrm{CRPS}^\flat_{w_1}(\mathrm{F}, y) &= w_1(y)\mathrm{CRPS}(\mathrm{F}^\flat_{w_1}, y) + \big(1 - w_1(y)\big)\mathrm{CRPS}(\mathrm{F}^\flat_{w_1}, r) \\
&= \mathbb{1}_{(-\infty,r)}(y) \left(\int_{-\infty}^{r} \big(F(s) - \Delta_y(s)\big)^2 \mathrm{d}s + \int_{r}^{\infty} (1-1)^2 \mathrm{d}s \right) \\
&\quad + \big(1 - \mathbb{1}_{(-\infty,r)}(y)\big) \left(\int_{-\infty}^{r} \big(F(s) - \Delta_r(s)\big)^2 \mathrm{d}s + \int_{r}^{\infty} (1-1)^2 \mathrm{d}s \right) \\
&= \mathbb{1}_{(-\infty,r)}(y) \left(\int_{-\infty}^{r} \big(F(s) - \Delta_y(s)\big)^2 \mathrm{d}s + \int_{r}^{\infty} (1-1)^2 \mathrm{d}s \right) \\
&\quad + \big(1 - \mathbb{1}_{(-\infty,r)}(y)\big) \left(\int_{-\infty}^{r} \big(F(s) - \Delta_y(s)\big)^2 \mathrm{d}s + \int_{r}^{\infty} (1-1)^2 \mathrm{d}s \right) \\
&= \int_{-\infty}^{\infty} w_1(s)\big(F(s) - \Delta_y(s)\big)^2 \mathrm{d}s
\end{aligned}$$

$$\begin{aligned}
\mathrm{CRPS}^\flat_{w_2}(\mathrm{F}, y) &= w_2(y)\mathrm{CRPS}(\mathrm{F}^\flat_{w_2}, y) + \big(1 - w_2(y)\big)\mathrm{CRPS}(\mathrm{F}^\flat_{w_2}, r) \\
&= \mathbb{1}_{(r,\infty)}(y) \left(\int_{-\infty}^{r} (0-0)^2 \mathrm{d}s + \int_{r}^{\infty} \big(F(s) - \Delta_y(s)\big)^2 \mathrm{d}s \right) \\
&\quad + \big(1 - \mathbb{1}_{(r,\infty)}(y)\big) \left(\int_{-\infty}^{r} (0-0)^2 \mathrm{d}s + \int_{r}^{\infty} \big(F(s) - \Delta_r(s)\big)^2 \mathrm{d}s \right) \\
&= \mathbb{1}_{(r,\infty)}(y) \left(\int_{-\infty}^{r} (0-0)^2 \mathrm{d}s + \int_{r}^{\infty} \big(F(s) - \Delta_y(s)\big)^2 \mathrm{d}s \right) \\
&\quad + \big(1 - \mathbb{1}_{(r,\infty)}(y)\big) \left(\int_{-\infty}^{r} (0-0)^2 \mathrm{d}s + \int_{r}^{\infty} \big(F(s) - \Delta_y(s)\big)^2 \mathrm{d}s \right) \\
&= \int_{-\infty}^{\infty} w_2(s)\big(F(s) - \Delta_y(s)\big)^2 \mathrm{d}s
\end{aligned}$$

E Monte Carlo simulation

In this section, we compare the size and power properties of the conditional and censored scoring rules of a selection of regular scoring rules based on the Giacomini and White (2006) test, for the null hypothesis

$$\mathbb{H}_0 : \mathbb{E}_{\mathrm{P}} S(\mathrm{F}, Y) = \mathbb{E}_{\mathrm{P}} S(\mathrm{G}, Y),$$

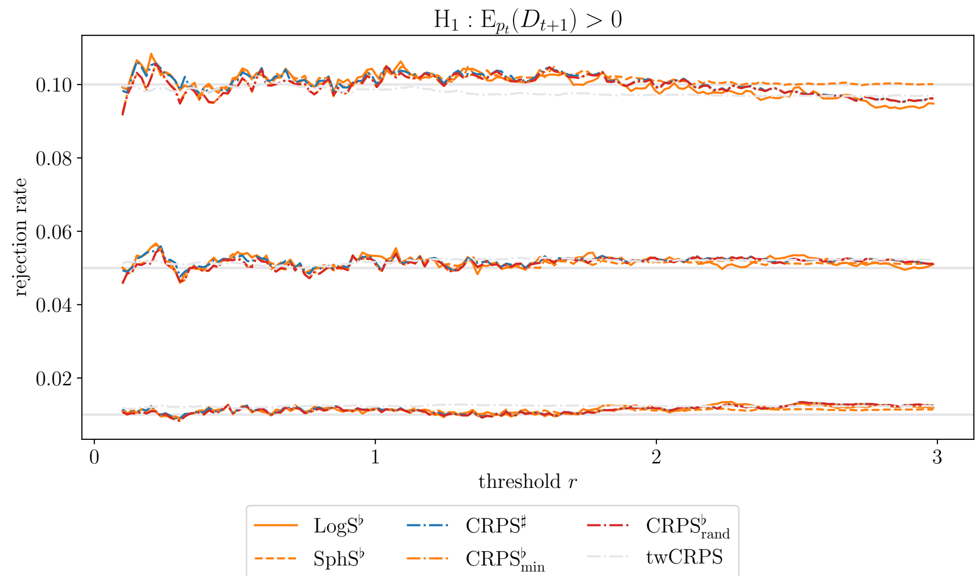
by means of the Diebold Mariano-type statistic $t_T = \frac{\frac{1}{T}\sum_{t=1}^{T} d_t}{\sqrt{\hat{\sigma}_t^2/T}}$, where d_t denotes realisations of $D_t = S(\mathrm{F}, Y_t) - S(\mathrm{G}, Y_T)$, $\hat{\sigma}_t$ should be a heteroskedasticity and autocorrelation-consistent (HAC) variance estimator in non-i.i.d. settings. This null hypothesis, which is equivalent to $\mathbb{H}_0 : \mathbb{D}_S(\mathrm{P}\|\mathrm{F}) = \mathbb{D}_S(\mathrm{F}\|\mathrm{G})$, is rejected if it is unlikely enough that quoting F instead of P leads to the same information loss as quoting G instead of P.

E.1 Size

As carefully explained by Diks et al. (2011), the null hypothesis of the GW test forces a particularly symmetric design. We adopt the design of Diks et al. (2011), using a center-indicator weight function $w(y) = \mathbb{1}_{[-r,r]}(y)$ combined with and i.i.d. standard normal DGP and normal candidates with unit variance and means $\mu_f = -0.2$ and $\mu_g = 0.2$. Due to the symmetry, the norms and $\bar{F}_w$-probabilities of the candidates are equivalent, leading to coinciding DM statistics based on QS and SphS scoring rules. Additionally, the equal norms and discrete probabilities also imply the censoring and conditioning rules to be equivalent within a semi-local scoring rule family since observations outside the region of interest obtain the same scores under both candidates in this case.

Figure 1 displays the rejection rates for rejection the null of equal predictive ability against the one-sided alternative that candidate f is statistically closer to p than g. The

Figure 1: Size properties GW test



rejection rates are given at nominal significance levels of 0.01, 0.05 and 0.10, for focused versions of the LogS, SphS and CRPS scoring rules, based on 10,000 simulations. Given the discussion above, this gives a complete picture of the selection $\{\text{LogS}, \text{SphS}, \text{QS}, \text{CRPS}\} \times \{\sharp, \flat\}$. The twCRPS is added since it will also be included as benchmark in the power studies based on weight functions for which the censored CRPS variants do not reduce to the twCRPS. None of the displayed rejection rates give reason to doubt the size correctness of the tests.

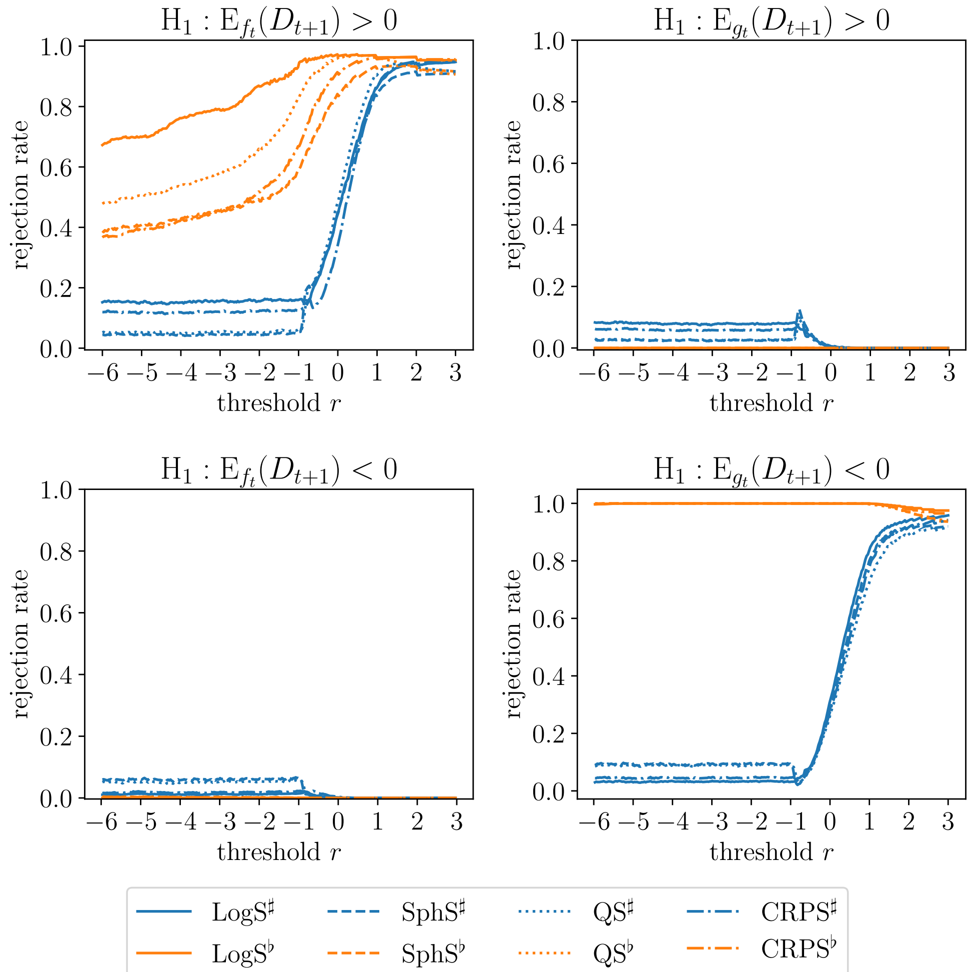
E.2 Power

Laplace tails Our first experiment studies the consequences of the lack of the conditional rule to disentangle two proportional tails when using the left tail indicator function $w(y) = \mathbb{1}_{(-\infty,r)}(y)$. In particular, we we analyze two Laplace candidates with different location $\mu_f = -1$ and $\mu_g = 1$ but equivalent scale $\theta_f = \theta_g = 1$. Interestingly, even if $\mu_p \to \mu_f$, the conditional scoring rule does not have any power against the null of the candidates being statically equally far away from p, that is, for thresholds $r < \mu_f$, for which the conditional distributions on $(-\infty, r)$ coincide. Since movements of p in terms of μ_p are invisible through the lens of a conditional score divergence, this is essentially not a lack of power against $\mathbb{H}_0$, which is based on the conditional scoring rule. Yet, it is a lack of power against the distributions being statistically equally far away from the actual density on $\{w > 0\}$ through the lens of the regular score divergence and, therefore, still a lack of local discriminative ability. More fundamentally, the GW test degenerates in this case, as the score differences are exactly zero.

Leaving this extreme case, we analyse what happens if the scale parameters are not exactly the same, but close. Specifically, we let $\theta_f = 1$ and $\theta_g = 1.1$. Figure 2 shows the rejection rates of the GW test if the DGP is f (left-hand side) or g (right-hand side) in favor of f (top) or g (bottom). Since both candidates are now also different through the lens of the conditional rule, the subfigures on the diagonal display actual power while the off-diagonal ones show spurious power. Concerning the selection of scoring rules, it is good to remember that the censored CRPS coincides with the twCRPS for the selected weight function.

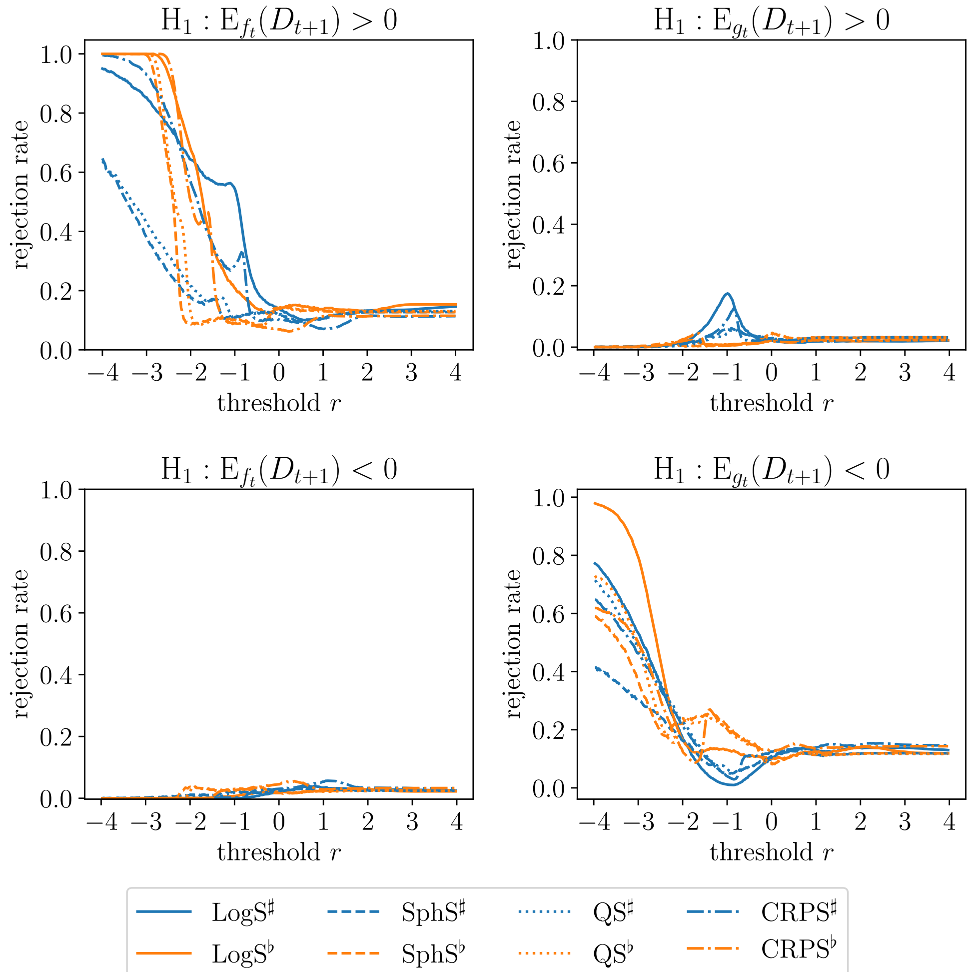
Three observations strike us. First of all, the increase in power from the conditional operator to the censoring operator is immense for all four scoring rules and thresholds $r < \mu_f$. The difference decreases over the interval $r \in (\mu_f, \mu_g)$, after which both conditioning and censoring have close to unit power. This observation is in line with the lack of discriminative ability of proportional and apparently close to proportional tails. Second, there is a clear difference in spurious power between the focusing operators: The censoring operator does seemingly not suffer from spurious power at all, whereas the conditional rules have spurious power up to 0.10 for thresholds smaller than $\mu_f = -1$. Third, we note that the censored likelihood score dominates the other scoring rules in terms of power.

Figure 2: Laplace experiment ($c = 20$)



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

Figure 3: $\mathcal{N}(0,1)$ versus Student-$t(5)$: Left-tail ($c = 20$)



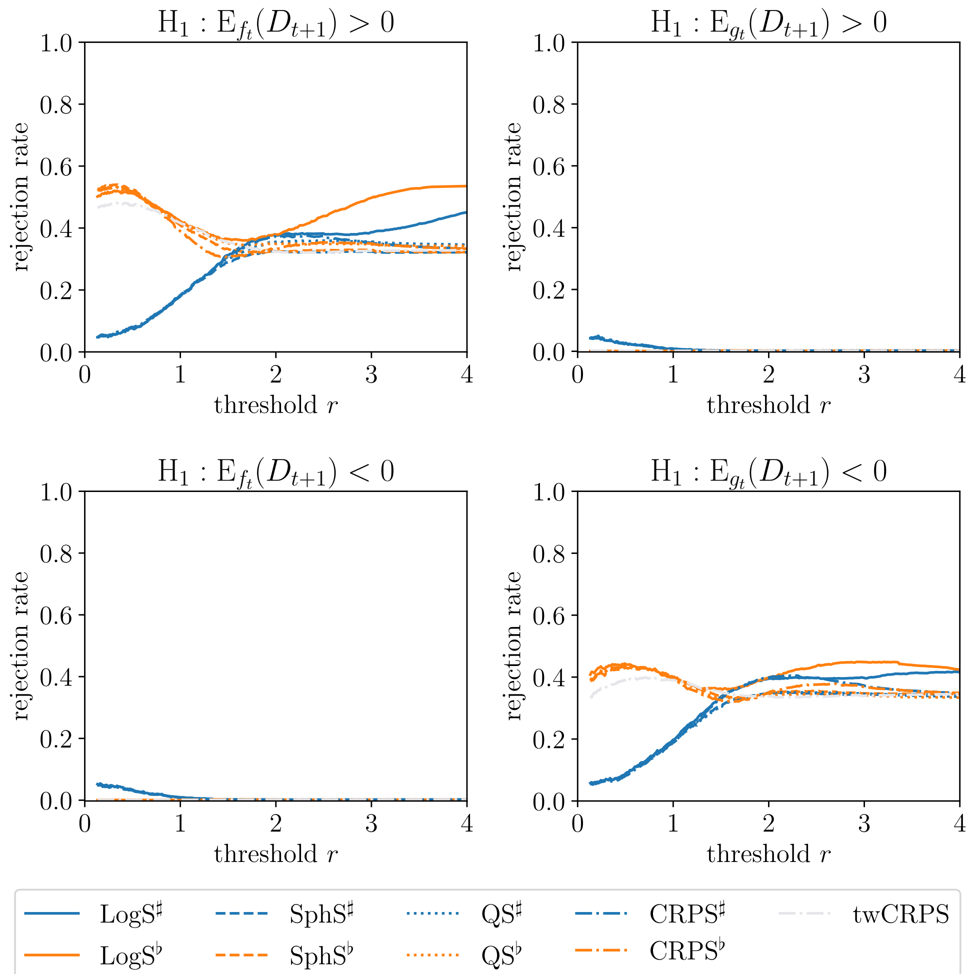
One-sided rejection rates of the GW-test of equal predictive ability of the candidates f_t (standard normal) and g_t (Student-$t(5)$) at a nominal significance level of 0.05 based on 10,000 simulations. The DGP is either f_t (left-hand side) or g_t (right-hand side). Moreover, rejections in the top panels are in favor of f_t, while rejections in the bottom panels are in favor of g_t. The incorporated weight function is $w(y) = 1_{(-\infty,r)}(y)$ and the number of expected observations in the region of interest is kept constant at $c = 20$.

Normal versus Student-t: Left-tail Figure 3 shows the rejection rates of the GW test, where f is standard normal and g Student-$t(5)$. Again, we consider the left-tail indicator function $w(y) = 1_{(-\infty,r)}(y)$ for varying values of r. The combination of the selected candidates and the left-tail region of interest make the current setting particularly interesting for financial risk management applications. As revealed by the figure, the rejection rate plots are now less monotonic, intersecting the graph of the competing focusing operator rejection rates. The latter occurs by construction since the densities of the candidates intersect as well, see Diks et al. (2011) for a discussion. Starting with the clearest differences, we note the spurious power humps of the conditional rules if the Student-$t(5)$ distribution is the DGP. In contrast, the censored scoring rules have almost no spurious power. Staying in the right column of Figure 3, the rejection rates in the bottom row reveal a clear preference for the censoring operator. Indeed, the exceptions of higher conditional power are rather weak, while the difference between the rejection rates (far) into the left-tail is particularly large for the Logarithmic and Spherical scoring rule. On the other hand, if the standard normal distribution is the DGP, then there is hardly (a difference in) spurious power. The differences between the rejection rates representing power are more extreme when the data is generated from the standard Normal distribution, yet so is their drop between $r = -2$ and $r = -1$, clouding a clear preference for one of the focusing operators for these intermediate tail values of r.

Normal versus Student-t: Center In our third Monte Carlo experiment, we focus on the center of the candidate distributions by implementing the weight function $w(y) = \mathbb{1}_{[-r,r]}(y)$. Figure 4 displays the rejection rates for the same selection of regular scoring rules as in the previous experiments. Based on Figure 4, the added value of censoring relative to conditioning is overwhelming: Censoring leads to higher power and lower spurious power, in particular for values smaller than $r = 1$, which are of particular interest in applications.

The $\mathrm{CRPS}_w^\flat$ displayed in the Figure 4 is the generalized censored scoring rule based on the generalized censored measure in Equation (5). Due to the symmetry of the set up, there is visually no difference between using the suggested value $\gamma = \frac{1}{2}$ or the estimated proportion $\hat{\gamma}$. We have also calculated the $\mathrm{CRPS}^\dagger(\mathrm{F}, y)$ introduced in Subsection 3.3.2, which visually coincides with the twCRPS in this case.

Figure 4: $\mathcal{N}(0,1)$ versus Student-$t(5)$: Center ($c = 200$)



One-sided rejection rates of the GW-test of equal predictive ability of the candidates f_t (standard normal) and g_t (Student-$t(5)$) at a nominal significance level of 0.05 based on 10,000 simulations. The DGP is either f_t (left-hand side) or g_t (right-hand side). Moreover, rejections in the top panels are in favor of f_t, while rejections in the bottom panels are in favor of g_t. The incorporated weight function is $w(y) = 1_{(-r,r)}(y)$ and the number of expected observations in the region of interest is kept constant at $c = 200$.

F Additional Tables

F.1 Risk management application

Table F.1.a: Model Confidence Sets for risk management examples.

			LogS		QS		SphS		CRPS	
q	h	Method	♭	♯	♭	♯	♭	♯	♭	♯
0.01	1	RGARCH-t	1.00	0.60	0.45	0.95	0.65	0.88	0.73	0.97
		TGARCH-t	0.99	1.00	0.63	1.00	0.88	1.00	0.81	1.00
		GARCH-t	0.53	0.69	0.34	0.84	0.65	0.88	0.81	0.91
		RGARCH-$\mathcal{N}$	**0.09**	**0.19**	1.00	0.95	1.00	0.66	0.81	0.97
		TGARCH-$\mathcal{N}$	**0.03**	**0.09**	0.63	0.95	0.88	0.38	1.00	0.97
		GARCH-$\mathcal{N}$	**0.01**	**0.09**	0.45	0.84	0.65	0.45	0.81	0.64
	5	RGARCH-t	0.37	0.87	**0.12**	1.00	**0.24**	1.00	0.57	1.00
		TGARCH-t	0.83	1.00	0.86	0.45	1.00	0.40	0.65	0.63
		GARCH-t	1.00	0.96	**0.17**	0.38	0.47	0.40	0.65	0.44
		RGARCH-$\mathcal{N}$	**0.01**	**0.05**	**0.12**	0.81	**0.18**	0.40	0.57	**0.18**
		TGARCH-$\mathcal{N}$	**0.01**	**0.05**	1.00	0.75	1.00	**0.23**	1.00	**0.16**
		GARCH-$\mathcal{N}$	**0.01**	**0.04**	**0.17**	0.75	0.41	0.27	0.65	**0.09**
0.05	1	RGARCH-t	1.00	0.79	**0.05**	1.00	**0.01**	1.00	0.40	0.79
		TGARCH-t	**0.11**	1.00	**0.02**	0.83	**0.01**	0.74	0.40	1.00
		GARCH-t	**0.01**	0.79	**0.00**	0.83	**0.00**	0.74	**0.14**	0.79
		RGARCH-$\mathcal{N}$	**0.09**	**0.06**	1.00	1.00	1.00	0.60	1.00	0.79
		TGARCH-$\mathcal{N}$	**0.00**	**0.03**	**0.05**	0.83	**0.01**	**0.24**	0.50	0.65
		GARCH-$\mathcal{N}$	**0.00**	**0.01**	**0.00**	0.83	**0.00**	**0.25**	0.40	0.37
	5	RGARCH-t	0.75	**0.23**	0.31	0.26	**0.25**	0.29	0.49	0.55
		TGARCH-t	0.98	0.75	0.98	1.00	1.00	1.00	1.00	1.00
		GARCH-t	1.00	1.00	0.31	0.36	0.41	0.57	0.56	0.55
		RGARCH-$\mathcal{N}$	**0.01**	**0.01**	0.98	**0.14**	0.92	**0.01**	0.49	**0.03**
		TGARCH-$\mathcal{N}$	**0.01**	**0.01**	1.00	0.36	0.81	**0.02**	0.74	**0.06**
		GARCH-$\mathcal{N}$	**0.01**	**0.01**	0.31	0.26	**0.25**	**0.01**	0.49	**0.06**
0.1	1	RGARCH-t	1.00	0.73	1.00	0.70	0.35	0.88	**0.12**	0.74
		TGARCH-t	**0.05**	1.00	**0.16**	1.00	**0.02**	1.00	**0.05**	1.00
		GARCH-t	**0.00**	0.40	**0.01**	0.43	**0.00**	0.37	**0.01**	**0.20**
		RGARCH-$\mathcal{N}$	**0.05**	**0.03**	0.49	0.70	1.00	**0.21**	1.00	0.74
		TGARCH-$\mathcal{N}$	**0.00**	**0.01**	**0.06**	0.39	**0.00**	**0.03**	**0.12**	**0.04**
		GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.06**	**0.00**	**0.01**	**0.03**	**0.01**
	5	RGARCH-t	0.46	**0.15**	0.35	**0.16**	0.55	**0.11**	0.47	0.43
		TGARCH-t	1.00	0.36	1.00	1.00	1.00	0.41	1.00	1.00
		GARCH-t	0.56	1.00	0.35	0.70	0.55	1.00	0.60	0.43
		RGARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.01**	**0.16**	0.26	**0.03**	0.53	**0.01**
		TGARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.01**	**0.16**	**0.21**	**0.00**	0.60	**0.00**
		GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.01**	**0.16**	**0.20**	**0.00**	0.46	**0.00**

Table F.1.a (Continued): Model Confidence Sets for risk management examples.

0.15	1	RGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	**0.08**	0.84
		TGARCH-t	**0.03**	0.79	**0.07**	0.64	**0.01**	**0.19**	**0.01**	0.88
		GARCH-t	**0.00**	**0.18**	**0.00**	**0.19**	**0.00**	**0.10**	**0.00**	**0.11**
		RGARCH-$\mathcal{N}$	**0.03**	**0.02**	**0.07**	0.64	0.52	**0.10**	1.00	1.00
		TGARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	0.38	**0.00**	**0.01**	**0.03**	**0.11**
		GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.04**	**0.00**	**0.00**	**0.01**	**0.00**
	5	RGARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	**0.08**	0.84
		TGARCH-t	**0.03**	0.79	**0.07**	0.64	**0.01**	**0.19**	**0.01**	0.88
		GARCH-t	**0.00**	**0.18**	**0.00**	**0.19**	**0.00**	**0.10**	**0.00**	**0.11**
		RGARCH-$\mathcal{N}$	**0.03**	**0.02**	**0.07**	0.64	0.52	**0.10**	1.00	1.00
		TGARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	0.38	**0.00**	**0.01**	**0.03**	**0.11**
		GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.04**	**0.00**	**0.00**	**0.01**	**0.00**
0.2	1	GARCH-$\mathcal{N}$	1.00	1.00	1.00	0.40	1.00	0.26	**0.10**	0.42
		GARCH-t	**0.02**	**0.23**	**0.10**	**0.06**	**0.02**	**0.02**	**0.01**	**0.14**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.06**	**0.06**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**
		QGARCH-I-t	**0.01**	**0.04**	**0.00**	1.00	**0.02**	1.00	1.00	1.00
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.01**	**0.09**
		QGARCH-II-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**
	5	RGARCH-t	**0.15**	0.36	**0.00**	0.89	**0.02**	0.97	0.37	0.67
		TGARCH-t	1.00	0.36	0.77	1.00	1.00	0.97	1.00	1.00
		GARCH-t	0.78	1.00	1.00	0.89	0.50	0.97	0.37	0.67
		RGARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	0.89	**0.00**	1.00	0.32	0.67
		TGARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	0.79	**0.00**	0.90	0.37	**0.02**
		GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	0.79	**0.00**	0.76	**0.23**	**0.01**
0.25	1	RGARCH-t	1.00	1.00	1.00	**0.04**	1.00	**0.02**	**0.17**	**0.11**
		TGARCH-t	**0.03**	0.28	0.74	**0.08**	**0.03**	**0.02**	**0.01**	**0.10**
		GARCH-t	**0.00**	**0.04**	0.74	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**
		RGARCH-$\mathcal{N}$	**0.01**	**0.04**	**0.00**	1.00	**0.00**	1.00	1.00	1.00
		TGARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.08**	**0.00**	**0.02**	**0.01**	**0.11**
		GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**
	5	RGARCH-t	**0.12**	0.33	**0.00**	0.76	**0.00**	0.96	0.28	0.49
		TGARCH-t	1.00	0.49	**0.16**	1.00	0.84	0.96	1.00	1.00
		GARCH-t	0.99	1.00	1.00	0.76	1.00	0.93	0.28	0.52
		RGARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	0.76	**0.00**	0.96	**0.12**	0.50
		TGARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	0.76	**0.00**	1.00	**0.13**	0.44
		GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	0.66	**0.00**	0.71	**0.12**	**0.08**

NOTE: This table presents the MCS p-values implied by censored ($\flat$) and conditional ($\sharp$) scoring rules, based on h-step ahead density forecasts. The emphasis is on the left tail, incorporated by the weight function $w_t(y_t) = \mathbb{1}_{(-infty,r_t^q}(y_t)$, where r_t^q is the empirical q-th quantile based on the estimation window. Bold (and underlined) p-values signify a forecast method's elimination from $\text{MCS}_{0.75}$ (and $\text{MCS}_{0.90}$). TR is the selected statistic, using $B = 10,000$ simulations and block length $k = 5$.

Table F.1.b: Overview and robustness of left-tail application

		$\text{MCS}_{0.90}$							$\text{MCS}_{0.75}$						
		|MCS|			VaR		ES		**|MCS|**			VaR		ES	
$T_{\textbf{est}}$	**Stat.**	$<$	$>$	**Ratio**	$\flat$	$\sharp$	$\flat$	$\sharp$	$<$	$>$	**Ratio**	$\flat$	$\sharp$	$\flat$	$\sharp$
							$h = 1$								
1000	TR_{20}	**17**	1	**2.28**	4	4	4	**2**	**15**	2	**2.04**	**4**	5	6	**3**
	TR_{100}	**17**	1	**2.30**	4	4	4	**2**	**15**	2	**2.04**	**4**	4	4	**0**
	Tmax_{20}	**11**	3	**1.58**	**0**	3	1	1	**17**	2	**2.07**	**4**	3	4	**1**
	Tmax_{100}	**11**	3	**1.48**	**0**	3	1	1	**16**	3	**2.02**	**4**	3	4	**1**
750	TR_{20}	**13**	3	**1.81**	**3**	6	3	3	**12**	3	**1.80**	**3**	7	7	**4**
	Tmax_{20}	**8**	7	**1.19**	**0**	3	1	1	**15**	4	**2.00**	**1**	4	5	**2**
1250	TR_{100}	**15**	2	**1.22**	4	4	4	**0**	**12**	1	**1.99**	4	4	5	**4**
	Tmax_{20}	**14**	4	**1.74**	2	2	3	**0**	**12**	4	**1.83**	**2**	3	5	**4**
							$h = 5$								
1000	TR_{20}	**9**	6	**1.69**	4	**0**	1	**0**	**12**	10	**1.72**	**4**	5	1	**0**
	TR_{100}	**9**	6	**1.69**	4	**0**	1	**0**	**12**	10	**1.66**	**4**	5	1	**0**
	Tmax_{20}	**10**	6	**1.43**	0	0	0	0	**12**	10	**1.59**	**0**	4	0	0
	Tmax_{100}	**9**	4	**1.44**	0	0	0	0	**11**	9	**1.59**	**0**	4	0	0
750	TR_{100}	8	**11**	**1.44**	**0**	2	0	0	8	**12**	**1.54**	**0**	6	0	0
	Tmax_{20}	**7**	6	**1.53**	**0**	1	0	0	**10**	9	**1.53**	**0**	4	0	0
1250	TR_{20}	**9**	8	**1.61**	1	1	0	0	**12**	8	**1.61**	**2**	6	0	0
	Tmax_{20}	**10**	7	**1.46**	**0**	1	0	0	**11**	10	**1.46**	**1**	4	0	0

NOTE: The table summarises MCS and backtesting results using varying values for estimation window T_{est}, equivalence test statistics TR_k and Tmax_k, blocklength k across forecast horizons $h = 1$ and $h = 5$, based on $B = 10,000$ bootstrap replications. All values are percentages. Columns labelled with |MCS| refer to MCS cardinality. Across 24 combinations of $q \in \{0.01, 0.05, 0.10, 0.15, 0.20, 0.25\}$ and $S \in \{\text{LogS}, \text{QS}, \text{SphS}, \text{CRPS}\}$, the $<$ ($>$)-column displays the frequency of $|\text{MCS}^{\flat}_{1-\alpha}| < (>) |\text{MCS}^{\sharp}_{1-\alpha}|$ and the %-column indicates average relative increase in cardinality from $\text{MCS}^{\sharp}_{1-\alpha}$ to $\text{MCS}^{\flat}_{1-\alpha}$ (in %). The VaR (ES) column shows the frequency of the $\text{MCS}_{1-\alpha}$ containing one of the top three models based on VaR (ES) backtesting results. Bolded numbers indicate strictly strictly smaller ($<$) or larger ($>$) $\text{MCS}^{\flat}_{1-\alpha}$ as well as strictly less times the $\text{MCS}_{1-\alpha}$ contains a top 3 VaR or ES method.

F.2 Inflation application

Table F.2.a: MSE and MAE of inflation forecasting applications

	MSE			MAE		
Method	$h = 6$	$h = 12$	$h = 24$	$h = 6$	$h = 12$	$h = 24$
Random Walk	8.74	**3.75**	**1.60**	1.97	**1.48**	**1.00**
AR	7.91	5.52	**3.67**	1.89	1.71	**1.46**
Bagging	5.74	**3.58**	4.67	1.83	1.56	1.69
CSR	**5.70**	5.18	8.11	**1.60**	1.76	2.14
LASSO	**5.66**	4.32	5.21	**1.66**	**1.55**	1.74
Random Forest	**5.40**	**2.76**	**1.93**	**1.54**	**1.18**	**1.11**

NOTE: MSE and MAE of forecast methods used in inflation examples, for the incorporated horizons $h \in \{6, 12, 24\}$. Bolded numbers indicate the best three models per performance measure.

Table F.2.b: Model Confidence Sets for inflation examples.

q	h	Method	LogS ♭	LogS ♯	QS ♭	QS ♯	SphS ♭	SphS ♯	CRPS ♭	CRPS ♯	CRPS tw
						Centre					
1	6	Random Walk	**0.09**	0.55	**0.04**	0.70	**0.01**	0.69	0.37	0.69	**0.18**
		AR	**0.09**	0.95	**0.02**	0.70	**0.00**	0.85	**0.05**	0.82	0.26
		Bagging	**0.00**	**0.00**	**0.00**	**0.06**	**0.00**	**0.01**	**0.01**	**0.00**	**0.15**
		CSR	0.37	0.99	**0.04**	0.70	**0.02**	0.85	0.50	0.82	0.93
		LASSO	**0.09**	0.26	**0.04**	0.66	**0.02**	0.45	0.50	0.39	1.00
		Random Forest	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93
	12	Random Walk	1.00	1.00	0.32	1.00	**0.10**	1.00	0.43	1.00	0.72
		AR	**0.10**	**0.22**	0.32	**0.20**	**0.10**	**0.13**	0.43	**0.23**	**0.17**
		Bagging	**0.00**	**0.17**	0.32	0.26	**0.03**	**0.08**	0.43	**0.23**	0.72
		CSR	**0.03**	0.30	0.32	0.38	**0.07**	0.31	0.43	0.41	**0.15**
		LASSO	**0.10**	0.30	0.32	0.51	**0.10**	0.33	0.43	0.41	0.79
		Random Forest	0.46	0.30	1.00	0.51	1.00	0.33	1.00	0.41	1.00
	24	Random Walk	1.00	0.64	**0.06**	0.74	**0.01**	0.99	**0.25**	0.83	0.92
		AR	**0.22**	1.00	**0.06**	0.74	**0.01**	1.00	**0.03**	1.00	**0.05**
		Bagging	**0.16**	0.32	**0.13**	1.00	**0.01**	0.37	0.41	**0.17**	0.92
		CSR	**0.20**	0.52	**0.06**	0.74	**0.01**	0.60	**0.09**	0.74	**0.04**
		LASSO	**0.22**	0.47	**0.06**	0.74	**0.01**	0.67	**0.25**	0.76	0.92
		Random Forest	0.84	0.66	1.00	0.74	1.00	0.99	1.00	0.96	1.00
1.5	6	Random Walk	0.27	0.68	**0.06**	0.31	**0.04**	0.60	0.70	0.61	**0.11**
		AR	**0.19**	0.97	**0.04**	0.26	**0.04**	0.60	**0.08**	0.62	**0.18**
		Bagging	**0.00**	**0.01**	**0.00**	**0.01**	**0.00**	**0.00**	**0.01**	**0.01**	**0.10**
		CSR	0.68	1.00	**0.17**	0.44	**0.09**	0.60	0.74	0.62	0.99
		LASSO	**0.19**	0.54	**0.06**	0.44	**0.01**	0.60	0.74	0.59	0.99
		Random Forest	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	12	Random Walk	1.00	1.00	**0.07**	1.00	**0.12**	1.00	0.46	1.00	0.54
		AR	**0.07**	0.37	**0.07**	0.52	**0.12**	0.47	0.26	0.39	**0.08**
		Bagging	**0.00**	**0.02**	**0.02**	**0.01**	**0.01**	**0.00**	0.26	**0.04**	0.54
		CSR	**0.00**	0.37	**0.07**	0.40	**0.09**	0.31	0.26	0.39	**0.08**
		LASSO	**0.02**	0.37	**0.07**	0.54	**0.09**	0.47	0.46	0.39	0.54
		Random Forest	0.61	0.37	1.00	0.81	1.00	0.54	1.00	0.39	1.00
	24	Random Walk	1.00	0.91	**0.00**	0.79	0.38	1.00	**0.14**	0.92	0.64
		AR	**0.16**	0.91	**0.00**	0.79	0.26	0.89	**0.00**	0.92	**0.04**
		Bagging	**0.00**	**0.06**	**0.19**	1.00	0.26	**0.00**	0.32	**0.08**	0.64
		CSR	**0.00**	0.91	**0.00**	0.79	**0.03**	0.80	**0.00**	0.92	**0.04**
		LASSO	**0.02**	0.72	**0.00**	0.79	**0.13**	0.80	**0.14**	0.92	0.64
		Random Forest	0.38	1.00	1.00	0.79	1.00	0.89	1.00	1.00	1.00

Table F.2.b (Continued): Model Confidence Sets for inflation examples.

			LogS		QS		SphS		CRPS		
q	h	Method	♭	♯	♭	♯	♭	♯	♭	♯	tw
2	6	Random Walk	0.36	0.66	**0.02**	0.51	**0.06**	0.48	0.39	0.70	**0.07**
		AR	**0.12**	0.83	**0.03**	0.53	**0.06**	0.69	**0.01**	0.89	**0.11**
		Bagging	**0.00**	**0.01**	**0.00**	**0.03**	**0.00**	**0.00**	**0.00**	**0.02**	**0.07**
		CSR	1.00	1.00	**0.20**	0.69	0.28	0.81	0.87	1.00	0.95
		LASSO	0.36	0.62	**0.06**	0.69	**0.01**	0.66	0.87	0.77	0.95
		Random Forest	0.68	0.66	1.00	1.00	1.00	1.00	1.00	0.93	1.00
	12	Random Walk	1.00	1.00	**0.01**	0.66	**0.07**	0.97	0.38	1.00	0.41
		AR	**0.17**	0.43	**0.01**	0.66	**0.07**	0.44	**0.06**	0.66	**0.03**
		Bagging	**0.00**	**0.02**	**0.00**	**0.03**	**0.00**	**0.00**	**0.06**	**0.07**	0.41
		CSR	**0.11**	0.26	**0.00**	**0.23**	**0.04**	**0.15**	**0.13**	0.38	**0.03**
		LASSO	**0.17**	0.43	**0.00**	0.66	**0.04**	0.44	0.35	0.66	0.41
		Random Forest	0.76	0.43	1.00	1.00	1.00	1.00	1.00	0.87	1.00
	24	Random Walk	1.00	0.74	**0.00**	0.74	0.58	1.00	0.48	1.00	1.00
		AR	0.37	1.00	**0.00**	0.74	**0.14**	0.29	**0.00**	0.92	**0.09**
		Bagging	**0.07**	**0.00**	**0.06**	1.00	**0.14**	**0.00**	0.42	**0.00**	**0.09**
		CSR	**0.19**	0.57	**0.00**	0.74	**0.02**	**0.19**	**0.00**	0.90	**0.09**
		LASSO	0.37	0.42	**0.00**	0.74	**0.14**	**0.17**	0.40	0.90	**0.09**
		Random Forest	0.48	0.57	1.00	0.74	1.00	**0.17**	1.00	0.90	0.91
		Tails									
1	6	Random Walk	**0.24**	0.46	**0.02**	0.26	**0.03**	0.31	0.33	**0.18**	**0.06**
		AR	**0.00**	**0.13**	**0.00**	0.31	**0.01**	0.44	1.00	**0.21**	**0.03**
		Bagging	**0.00**	**0.07**	**0.00**	0.26	**0.00**	**0.08**	0.30	0.84	**0.03**
		CSR	1.00	1.00	**0.13**	1.00	**0.11**	1.00	**0.09**	0.84	0.29
		LASSO	0.53	0.78	**0.03**	0.94	**0.03**	0.81	0.29	1.00	**0.16**
		Random Forest	0.53	0.46	1.00	0.94	1.00	0.81	**0.05**	0.84	1.00
	12	Random Walk	1.00	0.50	**0.18**	**0.07**	**0.12**	0.28	**0.08**	**0.00**	**0.06**
		AR	0.27	0.46	**0.05**	**0.07**	**0.08**	0.28	1.00	**0.00**	**0.06**
		Bagging	**0.03**	**0.07**	**0.00**	**0.07**	**0.00**	**0.03**	0.48	0.88	**0.06**
		CSR	**0.14**	0.46	**0.02**	**0.04**	**0.04**	**0.16**	0.47	**0.00**	**0.06**
		LASSO	0.27	1.00	**0.04**	**0.07**	**0.05**	0.28	0.45	0.88	**0.06**
		Random Forest	0.56	0.53	1.00	1.00	1.00	1.00	**0.08**	1.00	1.00
	24	Random Walk	1.00	0.58	1.00	0.74	0.80	0.69	0.43	0.31	1.00
		AR	**0.11**	0.58	**0.00**	0.74	**0.00**	0.69	0.94	0.31	**0.00**
		Bagging	**0.00**	**0.15**	**0.00**	0.69	**0.00**	0.69	1.00	**0.19**	**0.00**
		CSR	**0.02**	0.44	**0.00**	0.69	**0.00**	0.58	0.94	**0.09**	**0.00**
		LASSO	**0.11**	1.00	**0.00**	1.00	**0.00**	1.00	0.94	1.00	**0.00**
		Random Forest	**0.20**	0.31	**0.07**	0.66	1.00	0.50	0.28	0.31	0.31

Table F.2.b (Continued): Model Confidence Sets for inflation examples.

			LogS		QS		SphS		CRPS		
q	h	Method	$\flat$	$\sharp$	$\flat$	$\sharp$	$\flat$	$\sharp$	$\flat$	$\sharp$	tw
1.5	6	Random Walk	**0.15**	0.25	**0.02**	0.43	**0.01**	0.55	0.39	**0.24**	**0.14**
		AR	**0.00**	0.39	**0.02**	0.61	**0.00**	0.66	1.00	0.46	**0.06**
		Bagging	**0.01**	0.25	**0.00**	0.38	**0.00**	**0.02**	0.27	0.63	**0.03**
		CSR	1.00	0.68	**0.21**	0.80	**0.16**	0.82	**0.16**	0.63	**0.21**
		LASSO	0.52	1.00	**0.02**	1.00	**0.01**	1.00	0.27	1.00	**0.15**
		Random Forest	0.49	0.39	1.00	0.80	1.00	0.82	**0.08**	0.63	1.00
	12	Random Walk	1.00	0.40	0.33	**0.01**	**0.06**	0.43	**0.25**	**0.02**	**0.11**
		AR	**0.20**	0.40	**0.07**	**0.02**	**0.06**	0.43	1.00	**0.14**	**0.06**
		Bagging	**0.06**	0.40	**0.01**	**0.17**	**0.01**	0.26	0.31	1.00	**0.06**
		CSR	**0.12**	0.40	**0.03**	**0.01**	**0.04**	0.38	0.31	**0.14**	**0.06**
		LASSO	**0.20**	1.00	**0.05**	**0.17**	**0.04**	0.43	0.28	0.27	**0.06**
		Random Forest	0.52	0.40	1.00	1.00	1.00	1.00	**0.16**	0.27	1.00
	24	Random Walk	1.00	0.32	1.00	0.53	0.89	0.49	0.49	0.38	1.00
		AR	**0.06**	0.32	**0.00**	0.58	**0.01**	0.49	0.75	0.26	**0.00**
		Bagging	**0.06**	0.32	**0.00**	0.91	**0.02**	0.99	1.00	0.38	**0.01**
		CSR	**0.06**	0.32	**0.00**	0.41	**0.00**	0.44	0.75	0.26	**0.00**
		LASSO	**0.06**	1.00	**0.00**	1.00	**0.02**	1.00	0.75	1.00	**0.01**
		Random Forest	**0.08**	0.35	**0.10**	0.91	1.00	0.89	0.49	0.38	0.31
2	6	Random Walk	**0.16**	**0.24**	**0.10**	0.67	**0.02**	0.64	0.31	0.32	**0.18**
		AR	**0.01**	**0.23**	**0.05**	0.67	**0.01**	0.64	1.00	0.32	**0.10**
		Bagging	**0.04**	**0.24**	**0.01**	0.92	**0.00**	0.58	0.31	0.76	**0.04**
		CSR	1.00	**0.24**	0.68	0.92	0.34	1.00	0.28	0.68	**0.22**
		LASSO	0.67	1.00	**0.06**	0.92	**0.01**	0.96	0.31	1.00	**0.19**
		Random Forest	0.60	**0.24**	1.00	1.00	1.00	0.96	**0.21**	0.68	1.00
	12	Random Walk	1.00	0.26	0.86	0.28	**0.24**	0.62	**0.22**	**0.17**	0.31
		AR	**0.22**	**0.22**	**0.03**	0.34	**0.06**	0.62	1.00	**0.13**	**0.11**
		Bagging	**0.22**	0.66	**0.03**	1.00	**0.01**	0.94	0.31	1.00	**0.08**
		CSR	**0.22**	0.27	**0.03**	0.34	**0.06**	0.62	0.31	**0.17**	**0.10**
		LASSO	0.34	1.00	**0.07**	0.58	**0.06**	1.00	0.31	0.29	**0.11**
		Random Forest	0.46	0.46	1.00	0.58	1.00	0.94	0.27	0.29	1.00
	24	Random Walk	1.00	0.26	1.00	**0.24**	1.00	**0.24**	0.42	0.36	1.00
		AR	**0.07**	0.26	**0.00**	**0.24**	**0.00**	**0.24**	1.00	0.36	**0.01**
		Bagging	**0.03**	0.35	**0.00**	**0.24**	**0.00**	0.31	0.34	0.42	**0.01**
		CSR	**0.02**	0.26	**0.00**	**0.24**	**0.00**	**0.24**	0.34	0.36	**0.00**
		LASSO	**0.07**	1.00	**0.00**	1.00	**0.00**	1.00	0.34	1.00	**0.03**
		Random Forest	**0.18**	0.26	**0.11**	**0.24**	0.84	**0.22**	0.42	0.36	0.52

NOTE: This table mimics the setup of Table F.1.a, albeit with one additional column for the twCRPS, as the twCRPS and CRPS$^\flat$ no longer coincide by construction. The emphasis is on the centre or tails, incorporated by the weight function $w_t(y_t) = \mathbb{1}_{[2-q,2+q]}(y_t)$ and its complement. All other settings are consistent with Table F.1.a.

F.3 Climate application

Table F.3.a: MSE and MAE of climate forecasting applications

	MSE			MAE		
Method	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
GARCH-$\mathcal{N}$	5.13	10.14	13.92	1.80	2.54	2.99
GARCH-t	**5.10**	**10.06**	**13.74**	**1.79**	**2.52**	**2.97**
QGARCH-I-$\mathcal{N}$	5.17	10.30	14.27	1.81	2.56	3.03
QGARCH-I-t	5.15	10.26	14.18	1.80	2.55	3.02
QGARCH-II-$\mathcal{N}$	**4.82**	**8.26**	**9.93**	**1.73**	**2.26**	**2.49**
QGARCH-II-t	**4.82**	**8.26**	**9.94**	**1.73**	**2.26**	**2.49**

NOTE: MSE and MAE of forecast methods used in climate examples, for the incorporated horizons $h \in \{1, 2, 5\}$. Bolded numbers indicate the best three models per performance measure.

Table F.3.b: Model Confidence Sets for climate data examples.

			LogS		QS		SphS		CRPS		
q	h	Method	♭	♯	♭	♯	♭	♯	♭	♯	tw
					Centre						
1	1	GARCH-$\mathcal{N}$	**0.00**	0.27	**0.00**	0.48	**0.00**	0.43	**0.00**	0.44	**0.00**
		GARCH-t	**0.00**	0.62	**0.00**	0.48	**0.00**	0.43	**0.00**	0.48	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.04**	**0.00**	**0.02**	**0.00**	**0.02**	**0.00**	**0.03**	**0.00**
		QGARCH-I-t	**0.00**	**0.05**	**0.00**	**0.02**	**0.00**	**0.02**	**0.00**	**0.04**	**0.00**
		QGARCH-II-$\mathcal{N}$	0.76	**0.21**	1.00	0.51	1.00	0.43	1.00	0.41	1.00
		QGARCH-II-t	1.00	1.00	0.49	1.00	0.36	1.00	**0.08**	1.00	**0.13**
	2	GARCH-$\mathcal{N}$	**0.01**	**0.03**	**0.00**	**0.02**	**0.00**	**0.02**	**0.00**	**0.04**	**0.00**
		GARCH-t	**0.01**	1.00	**0.00**	0.43	**0.00**	0.69	**0.00**	0.76	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	**0.04**	**0.02**	1.00	**0.07**	1.00	**0.02**	1.00	**0.04**	1.00
		QGARCH-II-t	1.00	0.97	**0.00**	1.00	**0.00**	1.00	**0.00**	1.00	**0.04**
	3	GARCH-$\mathcal{N}$	**0.01**	**0.05**	**0.00**	**0.01**	**0.00**	**0.01**	**0.00**	**0.03**	**0.00**
		GARCH-t	**0.01**	0.31	**0.00**	**0.06**	**0.00**	**0.15**	**0.00**	**0.19**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	**0.01**	**0.05**	1.00	**0.02**	1.00	**0.01**	1.00	**0.03**	1.00
		QGARCH-II-t	1.00	1.00	**0.00**	1.00	**0.00**	1.00	**0.00**	1.00	**0.01**
2	1	GARCH-$\mathcal{N}$	**0.00**	**0.12**	**0.00**	**0.02**	**0.00**	**0.09**	**0.00**	**0.05**	**0.00**
		GARCH-t	**0.00**	**0.10**	**0.00**	**0.02**	**0.00**	**0.05**	**0.00**	**0.05**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.02**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**
		QGARCH-I-t	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**
		QGARCH-II-$\mathcal{N}$	0.36	0.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		QGARCH-II-t	1.00	1.00	0.77	**0.20**	0.64	0.66	0.90	0.48	**0.17**
	2	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.02**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.05**	**0.00**	**0.01**	**0.00**	**0.04**	**0.00**	**0.01**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	**0.02**	**0.02**	1.00	0.79	1.00	**0.04**	1.00	**0.22**	1.00
		QGARCH-II-t	1.00	1.00	**0.00**	1.00	**0.00**	1.00	**0.00**	1.00	**0.00**
	3	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.01**	1.00	0.58	1.00	**0.00**	1.00	**0.16**	1.00
		QGARCH-II-t	1.00	1.00	**0.00**	1.00	**0.00**	1.00	**0.00**	1.00	**0.01**

Table F.3.b (Continued): Model Confidence Sets for climate data examples

			LogS		QS		SphS		CRPS		
q	h	Method	$\flat$	$\sharp$	$\flat$	$\sharp$	$\flat$	$\sharp$	$\flat$	$\sharp$	tw
4	1	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	**0.01**	**0.02**	1.00	0.56	1.00	**0.01**	**0.24**	0.66	1.00
		QGARCH-II-t	1.00	1.00	0.57	1.00	0.33	1.00	1.00	1.00	0.29
	2	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	0.66	1.00	**0.00**	1.00	0.85	1.00
		QGARCH-II-t	1.00	1.00	**0.00**	1.00	**0.00**	1.00	**0.00**	1.00	**0.00**
	3	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		GARCH-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	**0.24**	1.00	**0.00**	1.00	0.65	1.00
		QGARCH-II-t	1.00	1.00	**0.00**	1.00	**0.00**	1.00	**0.00**	1.00	**0.00**
		Tails									
0.99	1	GARCH-$\mathcal{N}$	0.71	0.37	**0.08**	0.99	0.31	0.94	0.55	0.46	
		GARCH-t	0.95	**0.11**	**0.08**	0.99	0.31	1.00	1.00	0.46	
		QGARCH-I-$\mathcal{N}$	0.51	1.00	**0.08**	1.00	**0.01**	0.94	**0.10**	**0.01**	
		QGARCH-I-t	0.32	**0.13**	**0.08**	0.99	**0.02**	0.94	**0.07**	**0.01**	
		QGARCH-II-$\mathcal{N}$	0.37	**0.07**	1.00	**0.12**	0.98	**0.12**	**0.10**	0.46	
		QGARCH-II-t	1.00	**0.07**	0.97	**0.12**	1.00	**0.12**	**0.07**	1.00	
	2	GARCH-$\mathcal{N}$	**0.01**	**0.10**	**0.00**	0.68	**0.00**	0.26	**0.18**	**0.00**	
		GARCH-t	1.00	1.00	**0.00**	1.00	**0.00**	1.00	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.01**	**0.03**	**0.00**	**0.04**	**0.00**	**0.03**	**0.18**	**0.00**	
		QGARCH-I-t	**0.01**	**0.03**	**0.00**	**0.04**	**0.00**	**0.03**	**0.18**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.01**	**0.02**	0.45	**0.04**	**0.11**	**0.03**	**0.18**	**0.25**	
		QGARCH-II-t	0.63	**0.03**	1.00	**0.04**	1.00	**0.03**	**0.18**	1.00	
	3	GARCH-$\mathcal{N}$	**0.00**	**0.03**	**0.00**	**0.24**	**0.00**	**0.02**	**0.10**	**0.00**	
		GARCH-t	1.00	1.00	**0.00**	1.00	**0.00**	1.00	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.03**	**0.00**	**0.04**	**0.00**	**0.01**	**0.10**	**0.00**	
		QGARCH-I-t	**0.00**	**0.03**	**0.00**	**0.04**	**0.00**	**0.01**	**0.10**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.02**	1.00	**0.03**	1.00	**0.01**	**0.07**	0.94	
		QGARCH-II-t	0.92	**0.03**	0.27	**0.23**	0.99	**0.01**	**0.07**	1.00	

Table F.3.b (Continued): Model Confidence Sets for climate data examples

			LogS		QS		SphS		CRPS		
q	h	Method	♭	♯	♭	♯	♭	♯	♭	♯	tw
0.95	1	GARCH-$\mathcal{N}$	**0.06**	0.89	**0.00**	1.00	**0.00**	1.00	**0.18**	**0.00**	
		GARCH-t	**0.16**	1.00	**0.00**	**0.13**	**0.00**	**0.07**	1.00	**0.01**	
		QGARCH-I-$\mathcal{N}$	**0.16**	0.89	**0.00**	0.52	**0.00**	0.95	**0.08**	**0.00**	
		QGARCH-I-t	**0.06**	0.69	**0.00**	**0.09**	**0.00**	**0.07**	**0.18**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.06**	**0.17**	1.00	0.26	0.42	**0.07**	**0.18**	0.52	
		QGARCH-II-t	1.00	**0.19**	0.49	0.26	1.00	**0.07**	**0.18**	1.00	
	2	GARCH-$\mathcal{N}$	**0.00**	**0.01**	**0.00**	1.00	**0.00**	0.59	**0.05**	**0.00**	
		GARCH-t	**0.05**	1.00	**0.00**	0.72	**0.00**	1.00	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.05**	**0.00**	
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.05**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	**0.02**	1.00	**0.00**	**0.05**	1.00	
		QGARCH-II-t	1.00	**0.01**	0.29	0.72	0.56	**0.01**	**0.05**	0.35	
	3	GARCH-$\mathcal{N}$	**0.00**	**0.01**	**0.00**	0.73	**0.00**	0.37	**0.07**	**0.00**	
		GARCH-t	**0.00**	1.00	**0.00**	0.73	**0.00**	1.00	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.07**	**0.00**	
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.07**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	**0.01**	1.00	**0.01**	**0.07**	1.00	
		QGARCH-II-t	1.00	**0.01**	0.97	1.00	0.86	**0.03**	**0.07**	0.53	
0.9	1	GARCH-$\mathcal{N}$	**0.02**	0.62	**0.00**	0.80	**0.00**	1.00	**0.07**	**0.00**	
		GARCH-t	**0.03**	1.00	**0.00**	0.80	**0.00**	0.83	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.02**	0.62	**0.00**	0.48	**0.00**	0.80	**0.10**	**0.00**	
		QGARCH-I-t	**0.02**	0.62	**0.00**	0.66	**0.00**	0.79	**0.05**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.03**	**0.15**	1.00	0.80	1.00	0.58	**0.05**	0.38	
		QGARCH-II-t	1.00	0.62	**0.03**	1.00	0.52	0.83	**0.05**	1.00	
	2	GARCH-$\mathcal{N}$	**0.00**	**0.01**	**0.00**	**0.14**	**0.00**	0.50	**0.05**	**0.00**	
		GARCH-t	**0.00**	1.00	**0.00**	**0.14**	**0.00**	1.00	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.01**	**0.05**	**0.00**	
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.05**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	**0.17**	1.00	**0.06**	**0.05**	1.00	
		QGARCH-II-t	1.00	**0.03**	**0.01**	1.00	**0.15**	0.50	**0.05**	0.31	
	3	GARCH-$\mathcal{N}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.00**	
		GARCH-t	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.08**	**0.00**	
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.08**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	0.31	1.00	**0.05**	**0.08**	1.00	
		QGARCH-II-t	1.00	**0.02**	**0.03**	1.00	**0.05**	0.44	**0.08**	0.27	

Table F.3.b (Continued): Model Confidence Sets for climate data examples

			LogS		QS		SphS		CRPS		
q	h	Method	♭	♯	♭	♯	♭	♯	♭	♯	tw
0.85	1	GARCH-$\mathcal{N}$	**0.00**	0.31	**0.00**	**0.14**	**0.00**	**0.24**	0.55	**0.00**	
		GARCH-t	**0.01**	0.63	**0.00**	**0.14**	**0.00**	1.00	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	0.31	**0.00**	**0.14**	**0.00**	**0.16**	**0.05**	**0.00**	
		QGARCH-I-t	**0.00**	**0.18**	**0.00**	**0.14**	**0.00**	**0.13**	**0.05**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.01**	**0.08**	1.00	0.27	1.00	**0.13**	**0.05**	0.63	
		QGARCH-II-t	1.00	1.00	**0.02**	1.00	**0.13**	1.00	**0.05**	1.00	
	2	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.10**	**0.05**	**0.00**	
		GARCH-t	**0.00**	1.00	**0.00**	**0.00**	**0.00**	1.00	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.05**	**0.00**	
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.05**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	**0.03**	1.00	**0.01**	**0.05**	1.00	
		QGARCH-II-t	1.00	0.42	**0.01**	1.00	**0.00**	0.98	**0.05**	**0.15**	
	3	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.10**	**0.10**	**0.00**	
		GARCH-t	**0.00**	1.00	**0.00**	**0.00**	**0.00**	1.00	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.10**	**0.00**	
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.10**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	**0.01**	1.00	**0.00**	**0.10**	1.00	
		QGARCH-II-t	1.00	0.58	**0.00**	1.00	**0.01**	0.95	**0.06**	**0.09**	
0.8	1	GARCH-$\mathcal{N}$	**0.00**	**0.16**	**0.00**	**0.06**	**0.00**	0.91	0.46	**0.00**	
		GARCH-t	**0.00**	**0.22**	**0.00**	**0.06**	**0.00**	0.91	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.22**	**0.00**	**0.06**	**0.00**	0.91	**0.09**	**0.00**	
		QGARCH-I-t	**0.00**	**0.16**	**0.00**	**0.06**	**0.00**	0.77	0.46	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.08**	1.00	1.00	1.00	0.91	0.28	1.00	
		QGARCH-II-t	1.00	1.00	**0.10**	0.51	**0.17**	1.00	0.27	0.93	
	2	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.30	**0.04**	**0.00**	
		GARCH-t	**0.00**	0.61	**0.00**	**0.00**	**0.00**	0.56	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.04**	**0.00**	
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.04**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	0.45	1.00	**0.04**	**0.04**	1.00	
		QGARCH-II-t	1.00	1.00	**0.00**	1.00	**0.00**	1.00	**0.04**	**0.02**	
	3	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.41	**0.09**	**0.00**	
		GARCH-t	**0.00**	0.31	**0.00**	**0.00**	**0.00**	0.41	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.09**	**0.00**	
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.09**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	0.26	1.00	**0.03**	**0.07**	1.00	
		QGARCH-II-t	1.00	1.00	**0.00**	1.00	**0.00**	1.00	**0.07**	**0.03**	

Table F.3.b (Continued): Model Confidence Sets for climate data examples

q	h	Method	LogS ♭	LogS ♯	QS ♭	QS ♯	SphS ♭	SphS ♯	CRPS ♭	CRPS ♯	tw
0.75	1	GARCH-$\mathcal{N}$	**0.00**	**0.09**	**0.00**	**0.00**	**0.00**	0.35	0.25	**0.00**	
		GARCH-t	**0.00**	**0.09**	**0.00**	**0.00**	**0.00**	0.35	0.49	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.09**	**0.00**	**0.00**	**0.00**	0.35	**0.07**	**0.00**	
		QGARCH-I-t	**0.00**	**0.09**	**0.00**	**0.00**	**0.00**	0.35	1.00	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.09**	1.00	1.00	1.00	0.65	**0.19**	1.00	
		QGARCH-II-t	1.00	1.00	0.29	0.63	**0.16**	1.00	**0.18**	0.63	
	2	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.15**	**0.02**	**0.00**	
		GARCH-t	**0.00**	**0.05**	**0.00**	**0.00**	**0.00**	**0.15**	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.06**	**0.00**	
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.11**	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	1.00	1.00	**0.15**	**0.11**	1.00	
		QGARCH-II-t	1.00	1.00	**0.00**	0.61	**0.00**	1.00	**0.10**	**0.00**	
	3	GARCH-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.03**	**0.08**	**0.00**	
		GARCH-t	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.03**	1.00	**0.00**	
		QGARCH-I-$\mathcal{N}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.05**	**0.00**	
		QGARCH-I-t	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	1.00	**0.00**	
		QGARCH-II-$\mathcal{N}$	**0.00**	**0.00**	1.00	0.73	1.00	**0.03**	**0.10**	1.00	
		QGARCH-II-t	1.00	1.00	**0.00**	1.00	**0.00**	1.00	**0.09**	**0.01**	

NOTE: This table mimics the setup of Table F.2.b. The emphasis is on the centre or tails, incorporated by the weight function $w_t(y_t) = \mathbb{1}_{[18-q,18+q]}(y_t)$ and $w_t(y_t) = \mathbb{1}_{(r_t^q,\infty)}(y_t)$, respectively. For the latter weight function, the twCRPS is equivalent to the CRPS$^\flat$ by construction, and is therefore excluded from the table. All other settings are consistent with those outlined in the caption of Table F.2.b, except for the block length $k = 200$.

References

Diks, C., V. Panchenko, and D. van Dijk (2011). Likelihood-based scoring rules for comparing density forecasts in tails. Journal of Econometrics 163, 215–230.

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

Holzmann, H. and B. Klar (2016). Weighted scoring rules and hypothesis testing. Arxiv, ArXiv E-Prints.

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, 289–337.