

Supplementary Material for “Localizing Strictly Proper Scoring Rules”

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

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

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

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

October 10, 2025

Abstract

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

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

Contents

A	Proofs	4
A.1	Proof of Theorem 2	4
A.2	Proof of Theorem 3	8
B	Additional Proofs and Derivations	13
B.1	The censored density	13
B.2	Proof of Corollary 1	14
B.3	Proof of Corollary 2	15
C	Z, Q -Randomization	16
D	Censoring of Distance-Sensitive Scoring Rules	19
E	Derivations for Specific Scoring Rules	23
E.1	Logarithmic Score (LogS)	24
E.2	Power Scores (PowS_α)	25
E.3	PseudoSpherical Scores (PsSphS_α)	27
E.4	Kernel Scores (S_ρ)	30
F	Additional Example	35
G	Monte Carlo Study	36
G.1	Size	38
G.2	Power	39
H	Details on Model Specifications	51
H.1	Financial risk management	51
H.2	Macroeconomics	52
H.3	Climate	55
I	Additional Tables	56

I.1	MCS table at confidence level 0.75	56
I.2	Financial risk management	56
I.3	Macroeconomics	71
I.4	Climate	76

A Proofs

A.1 Proof of Theorem 2

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

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

Then

$$\int_{\mathcal{Y}} S_{w,H}^b(F, y)P(dy) = \int_{\mathcal{Y}} S(F_{w,H}^b, y)P_{w,H}^b(dy),$$

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

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

$$\begin{aligned} \int_{\mathcal{Y}} S_{w,H}^b(F, y)P(dy) &= \int_{\mathcal{Y}} \left(w(y)S(F_{w,H}^b, y) + (1 - w(y)) \int_{\mathcal{Y}} S(F_{w,H}^b, q)H(dq) \right) P(dy) \\ &= \int_{\mathcal{Y}} w(y)S(F_{w,H}^b, y)P(dy) + \int_{\mathcal{Y}} S(F_{w,H}^b, q) \int_{\mathcal{Y}} (1 - w(y))P(dy)H(dq) \\ &= \int_{\mathcal{Y}} S(F_{w,H}^b, y)P_w(dy) + \int_{\mathcal{Y}} S(F_{w,H}^b, y)\bar{P}_wH(dy) \\ &= \int_{\mathcal{Y}} S(F_{w,H}^b, y)(P_w(dy) + \bar{P}_wH(dy)) \\ &= \int_{\mathcal{Y}} S(F_{w,H}^b, y)P_{w,H}^b(dy). \end{aligned}$$

□

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

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

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

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

$$\begin{aligned} P_{w,H}^b(E) = F_{w,H}^b(E) &\implies P_{w,H}^b(E \cap \tilde{E}) = F_{w,H}^b(E \cap \tilde{E}) \\ &\implies \int_{\mathcal{Y}} (1-w) dP \cdot H(E \cap \tilde{E}) = \int_{\mathcal{Y}} (1-w) dF \cdot H(E \cap \tilde{E}) \\ &\implies \int_{\mathcal{Y}} (1-w) dP \cdot H(\tilde{E}) = \int_{\mathcal{Y}} (1-w) dF \cdot H(\tilde{E}) \\ &\implies \int_{\mathcal{Y}} (1-w) dP = \int_{\mathcal{Y}} (1-w) dF, \end{aligned}$$

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

$$\begin{aligned} P_{w,H}^b(E) = F_{w,H}^b(E) &\iff \int_{\mathcal{Y}} w(y) \mathbf{1}_{y \in E} P(dy) + \bar{P}_w H(E) = \int_{\mathcal{Y}} w(y) \mathbf{1}_{y \in E} F(dy) + \bar{F}_w H(E) \\ &\implies \int_{\mathcal{Y}} w(y) \mathbf{1}_{y \in E} P(dy) = \int_{\mathcal{Y}} w(y) \mathbf{1}_{y \in E} F(dy) \\ &\implies P(E \cap \{w > 0\}) = F(E \cap \{w > 0\}). \end{aligned}$$

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

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

and, consequently, $\mathbb{P}_{w, \mathbb{H}}^b(E) = \mathbb{F}_{w, \mathbb{H}}^b(E)$, $\forall E \in \mathcal{G}, \forall \mathbb{H} \in \mathcal{H}$. \square

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

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

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

(i) $S_{w, \mathbb{H}}^b(\mathbb{P}, y)$ is localizing relative to \mathcal{W} , $\forall \mathbb{H} \in \mathcal{H}$, by Corollary A3.

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

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

By definition of the class $\mathcal{P}_{w, \mathbb{H}}^b \equiv \{[\mathbb{P}]_{w, \mathbb{H}}^b, \mathbb{P} \in \mathcal{P}\}$, where $[\cdot]_{w, \mathbb{H}}^b : \mathcal{P} \rightarrow \mathcal{P}_{w, \mathbb{H}}^b$, denotes the

map given by the censored measure in Definition 6, this is equivalent to

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

and hence, by Lemma A1, also to

$$\int_{\mathcal{Y}} S_{w,\mathcal{H}}^b(\mathbf{P}, y)\mathbf{P}(dy) \geq \int_{\mathcal{Y}} S_{w,\mathcal{H}}^b(\mathbf{F}, y)\mathbf{P}(dy), \quad \forall \mathbf{P}, \mathbf{F} \in \mathcal{P}. \quad (\text{A.3})$$

Therefore, $S_{w,\mathcal{H}}^b(\mathbf{P}, y)$ is proper relative to \mathcal{P} by Definition 1.

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

$$\int_{\mathcal{Y}} S(\mathbf{P}_{w,\mathcal{H}}^b, y)\mathbf{P}_{w,\mathcal{H}}^b(dy) = \int_{\mathcal{Y}} S(\mathbf{F}_{w,\mathcal{H}}^b, y)\mathbf{P}_{w,\mathcal{H}}^b(dy) \iff \mathbf{P}_{w,\mathcal{H}}^b(E) = \mathbf{F}_{w,\mathcal{H}}^b(E),$$

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

$$\int_{\mathcal{Y}} S(\mathbf{P}_{w,\mathcal{H}}^b, y)\mathbf{P}_w^b(dy) = \int_{\mathcal{Y}} S(\mathbf{F}_{w,\mathcal{H}}^b, y)\mathbf{P}_{w,\mathcal{H}}^b(dy) \iff \mathbf{P}(E \cap \{w > 0\}) = \mathbf{F}(E \cap \{w > 0\}),$$

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

$$\int_{\mathcal{Y}} S_{w,\mathcal{H}}^b(\mathbf{P}, y)\mathbf{P}(dy) = \int_{\mathcal{Y}} S_{w,\mathcal{H}}^b(\mathbf{F}, y)\mathbf{P}(dy) \iff \mathbf{P}(E \cap \{w > 0\}) = \mathbf{F}(E \cap \{w > 0\}),$$

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

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

The score divergence $\mathbb{D}_{S_{w,\mathcal{H}}^b}$ is a local divergence because $S_{w,\mathcal{H}}^b$ is strictly locally proper.

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

$$\begin{aligned} \mathbb{D}_{S_{w,H}^b}(P\|F) &= \int_{\mathcal{Y}} S_{w,H}^b(F, y)P(dy) - \int_{\mathcal{Y}} S_{w,H}^b(F, y)P(dy) \\ &= \int_{\mathcal{Y}} S(P_{w,H}^b, y)P_{w,H}^b(dy) - \int_{\mathcal{Y}} S(F_{w,H}^b, y)P_{w,H}^b(dy), \\ &= \mathbb{D}_S(P_{w,H}^b\|F_{w,H}^b). \end{aligned}$$

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

A.2 Proof of Theorem 3

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

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

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

the equivalent hypotheses

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

which further simplifies to

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

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

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

B Additional Proofs and Derivations

B.1 The censored density

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

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

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

is the censored $(\mu + \delta_*)$ -density of F_w^b .

We can simplify this density as follows. Understanding that

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

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

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

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

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

By the same token, we conclude from

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

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

Put together, we arrive at

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

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

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

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

B.2 Proof of Corollary 1

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

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

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

B.3 Proof of Corollary 2

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

$$\begin{aligned}\lambda_A^\sharp(y) &:= e^{\bar{\lambda}_A^\sharp(y)} = \frac{F_0(A)}{F_1(A)} \frac{f_1(y)}{f_0(y)} \mathbb{1}_A(y) + \frac{1}{1} \mathbb{1}_{A^c}(y) \\ \lambda_A^b(y) &= e^{\bar{\lambda}_A^b(y)} = \frac{f_1(y)}{f_0(y)} \mathbb{1}_A(y) + \frac{F_1(A^c)}{F_0(A^c)} \mathbb{1}_{A^c}(y).\end{aligned}$$

There exist many examples for which the power of the censored test is strictly larger than the power of the conditional test. For instance, suppose that f_0 and f_1 have identical support in \mathbb{R} , and on A are strictly positive, proportional and unequal. Then, by proportionality, $\frac{F_0(A)}{F_1(A)} \frac{f_1(y)}{f_0(y)} = 1$ for $y \in A$ and hence $\lambda_A^\sharp(y) = 1$, so that a test of level $\alpha \in (0, 1)$ based on λ_A^\sharp will reject with probability α regardless of the outcome y , and hence only have trivial power α under f_1 . In that same situation, a level α test based on the censored likelihood ratio λ_A^b will have power strictly larger than α , because $\lambda_A^b(y)$ can take two different values, depending on whether the outcome y falls in A or A^c . By rejecting \mathbb{H}_0 with higher probability for those realizations y for which the censored likelihood ratio $\lambda_A^b(y)$ takes the larger of these values, it will achieve power strictly larger than α . Consequently, the conditional test ϕ_A^\sharp is not UMP.

As a concrete example, consider the above proportional situation with $F_0(A) = \eta_0 \in (0, 1)$ and $F_1(A) = \eta_1 \in (0, 1)$, where $\eta_1 > \eta_0$ and hence $\frac{\eta_1}{\eta_0} > 1 > \frac{1-\eta_1}{1-\eta_0}$, so the likelihood ratio in favor of \mathbb{H}_1 is larger than 1 for $y \in A$ and smaller than 1 for $y \in A^c$. Under these

conditions it can be verified that the level α test based on λ_A^b for $\alpha \in (0, \eta_0)$ rejects \mathbb{H}_0 in favor of \mathbb{H}_1 with probability $\gamma_A = \frac{\alpha}{\eta_0}$ for $y \in A$ and with probability $\gamma_{A^c} = 0$ for $y \in A^c$. The power of this test is $\frac{\alpha\eta_1}{\eta_0} > \alpha$. For $\alpha \in [\eta_0, 1)$, the level α test based on λ_A^b rejects with probability $\gamma_A = 1$ for $y \in A$ and with probability $\gamma_{A^c} = \frac{\alpha - \eta_0}{1 - \eta_0}$ for $y \in A^c$. The power of the test is $\eta_1 + \frac{\alpha - \eta_0}{1 - \eta_0}(1 - \eta_1)$, which is also strictly larger than α , as can be seen by inspecting the complementary probability: $1 - \left(\eta_1 + \frac{\alpha - \eta_0}{1 - \eta_0}(1 - \eta_1)\right) = (1 - \eta_1) \left(1 - \frac{\alpha - \eta_0}{1 - \eta_0}\right) = \frac{1 - \eta_1}{1 - \eta_0}(1 - \alpha) < 1 - \alpha$. \square

C Z, Q -Randomization

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

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

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

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

map reduces to the censored random variable in Equation (1).

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

$$S_w^b(F, y) = \mathbb{E}_{B_{w(y)}} S(F_w^b, Y_w^b(y, Z)). \quad (\text{C.2})$$

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

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

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

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

$$S_{w,H}^b(F, y) = \mathbb{E}_{B_{w(y),H}} S(F_{w,H}^b, Y_{w,H}^b(y, Z, Q)). \quad (\text{C.4})$$

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

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

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

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

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

D Censoring of Distance-Sensitive Scoring Rules

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

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

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

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

same for all distributions in \mathcal{P} , the censored scoring rule based on F_{A,r_1,r_2}^b remains localizing.

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

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

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

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

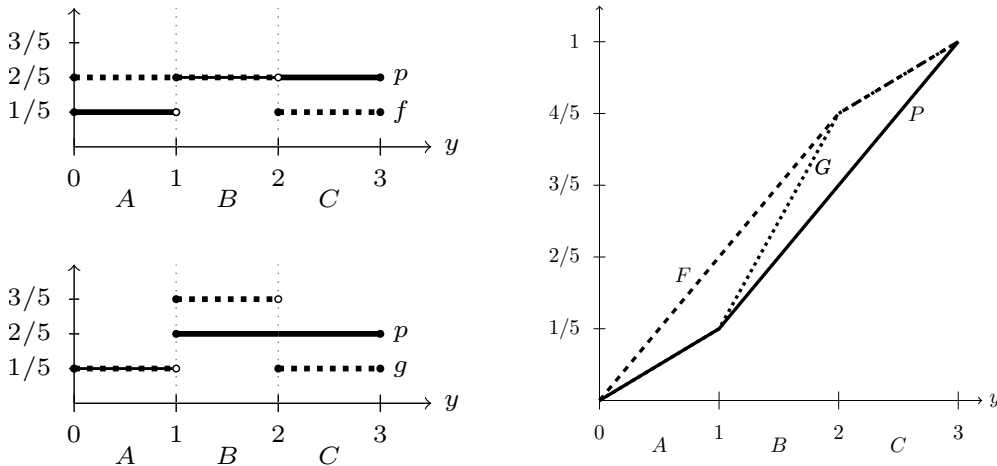


Figure D.1: Densities (left) and distribution functions (right) of distributions F , G and true distribution P , all piecewise uniformly distributed on $[0, 3]$ but with different probabilities $\boldsymbol{\pi} := (\pi_A, \pi_B, \pi_C)'$. Specifically, $\boldsymbol{\pi}_p = (1/5, 2/5, 2/5)'$, $\boldsymbol{\pi}_f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

The localization bias towards G in Example D.3 is undesirable if the researcher is solely concerned with outcomes in B . In practice, however, interest might extend beyond region B alone, encompassing outcomes located specifically in either region A or C , rather than their combined complement $B^c = A \cup C$. Put differently, a researcher might have explicit

interest in events within the collection $\tilde{\mathcal{B}}_A$, which cannot be adequately represented by a weight function defined on the outcome space alone. Consequently, in those situations non-localizing scoring rules, such as the threshold-weighted continuous ranked probability score (twCRPS), may be better suited for evaluating predictive accuracy when the researcher’s focus does not correspond precisely to a predefined weight function w . However, in this paper, consistent with the frameworks presented by Diks et al. (2011), Holzmann and Klar (2017a), and Allen et al. (2023), we maintain the assumption that the researcher’s interests are fully captured by a given weight function defined on the outcome space.

E Derivations for Specific Scoring Rules

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

E.1 Logarithmic Score (LogS)

Following the order in which the assertions pertaining to the Logarithmic scoring rule

$\text{LogS}(f, y) = \log f(y)$ appear in Table 1, they can be easily verified as follows:

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

$$\begin{aligned}
\text{LogS}_w^\flat(f, y) &= w(y) \log (w(y)f(y) + \bar{F}_w \mathbb{1}_{\{*\}}(y)) + (1 - w(y)) \log \bar{F}_w \\
&\stackrel{\text{a.s.}}{=} w(y) \log (w(y)f(y)) + (1 - w(y)) \log \bar{F}_w \\
&\stackrel{\text{eqv.}}{=} w(y) \log (f(y)) + (1 - w(y)) \log \bar{F}_w \\
&= S_w^{\text{CSL}}(f, y),
\end{aligned}$$

$$\begin{aligned}
\text{LogS}_{w,h}^\flat(f, y) &= w(y) \log f_{w,h}^\flat(y) + (1 - w(y)) \int_{\mathcal{Y}} \log f_{w,h}^\flat(q) h(q) \mu(\text{d}q) \\
&= w(y) \left(\log (f_w(y)) \mathbb{1}_{A_w}(y) + \log (\bar{F}_w h(y)) \mathbb{1}_{A_w^c}(y) \right) \\
&\quad + (1 - w(y)) \int_{A_w^c} \left(\log (f_w(q)) \mathbb{1}_{A_w}(y) + \log (\bar{F}_w h(q)) \mathbb{1}_{A_w^c}(y) \right) h(q) \mu(\text{d}q) \\
&= w(y) \log f_w(y) + (1 - w(y)) \int_{A_w^c} \log (\bar{F}_w h(q)) h(q) \mu(\text{d}q) \\
&\stackrel{\text{eqv.}}{=} w(y) \log f(y) + (1 - w(y)) \log \bar{F}_w \\
&= S_w^{\text{CSL}}(f, y).
\end{aligned}$$

E.2 Power Scores (PowS $_{\alpha}$)

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

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

for which we rely on the result expressed in

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

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

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

where in the before last step l'Hôpital's rule was used. Furthermore, the conditional version of the PowS $_{\alpha}$ family displayed in Table 1 is nothing but a direct application of the conditioning procedure.

By the linearity of limits, we obtain

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

Turning to the censored focusing method, we note that

$$\|f_w^{\flat}\|_{\alpha}^{\alpha} = \int_{\mathcal{Y}} (f_w(y) + \bar{F}_w \mathbf{1}_{\{*\}}(y))^{\alpha} (\mu + \delta_*)(dy) = \|f_w(y)\|_{\alpha}^{\alpha} + \bar{F}_w^{\alpha}. \quad (\text{E.1})$$

Using this result, we obtain

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

which bears the following limit

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

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

E.3 PseudoSpherical Scores (PsSphS_α)

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

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

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

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

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

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

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

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

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

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

For the conditional PsSphS $_\alpha$ family, we find

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

By the linearity of limits, we find

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

Moreover, for the censored PsSphS_α family, we recall Equation (E.1) and find that

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

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

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

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

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

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

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

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

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

E.4 Kernel Scores (S_ρ)

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

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

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

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

$\mathbb{E}_F \rho(X, X')$ introduced above. The corresponding score divergence reads

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

Assumption 1 is satisfied for all weight functions and distributions $F \in \mathcal{P}_r$ since $F_w(r) = 0$.

Therefore, the score divergence is a localized divergence if S_{ρ} is strictly proper relative to \mathcal{P}_r^b , which follows from the conditions under which S_{ρ} is strictly proper with respect to \mathcal{P}_r .

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

$$\text{ES}_{\beta}(F, y) := \frac{1}{2} \mathbb{E}_F \|\mathbf{Y} - \mathbf{Y}'\|_2^{\beta} - \mathbb{E}_F \|\mathbf{Y} - \mathbf{y}\|_2^{\beta},$$

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

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

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

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

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

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

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

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

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

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

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

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

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

Accordingly, the corresponding score divergence is given by

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

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

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

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

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

F Additional Example

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

G Monte Carlo Study

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

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

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

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

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

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

To put our findings for the censoring approach into perspective, we also evaluate DM test statistics based on focused scoring rules resulting from alternative localization procedures. In particular, we include conditional scoring rules and composite rules based on the proposal of Holzmann and Klar (2017a) to augment a conditional scoring rule with the auxiliary rule `sbar` or `slog`, as discussed in Section 3.5. For these composite scoring rules, we also assess to what extent their discriminative ability is attributable

to the original and auxiliary scoring rules, by decomposing the associated standardized score divergences, as follows. We denote the score differences of the original conditional rule and of the auxiliary rule by $D_S^\sharp \equiv D_S^\sharp(Y; F, G) := S_w^\sharp(F, Y) - S_w^\sharp(G, Y)$ and $D_s \equiv D_s(Y; F, G) := s(b_{\bar{F}_w}, Y) - s(b_{\bar{G}_w}, Y)$, respectively. The score divergence for the composite rule can then be written as $\mathbb{D}_{S^\sharp+s}(F\|G) \equiv \mathbb{E}_F(D_S^\sharp + D_s) = \mathbb{E}_F(D_S^\sharp) + \mathbb{E}_F(D_s)$, with associated variance $\mathbb{V}_F(D_S^\sharp + D_s) = \mathbb{V}_F(D_S^\sharp) + \mathbb{V}_F(D_s) + 2\text{Cov}_F(D_S^\sharp, D_s)$. Define the standardized divergence as $\tilde{\mathbb{D}}_S(F\|G) := \mathbb{D}_S(F\|G)/\sqrt{\mathbb{V}_F(D_S)}$. To evaluate the contribution of the auxiliary scoring rule s to $\tilde{\mathbb{D}}_{S^\sharp+s}$, we may compare $\tilde{\mathbb{D}}_{S^\sharp}(F\|G) = \mathbb{E}_F(D_{S^\sharp})/\sqrt{\mathbb{V}_F(D_{S^\sharp})}$ versus $\tilde{\mathbb{D}}_{S^\sharp+s}(F\|G) = \mathbb{E}_F(D_{S^\sharp} + D_s)/\sqrt{\mathbb{V}_F(D_{S^\sharp} + D_s)}$. In the power experiments below, we therefore consider their ratio $\xi_{S,s} := \tilde{\mathbb{D}}_{S^\sharp}(F\|G)/\tilde{\mathbb{D}}_{S^\sharp+s}(F\|G)$. Note that this ratio is non-negative by construction, but not necessarily less than unity. Since $\mathbb{E}_F(D_S) \geq 0$ for any proper scoring rule S , it follows that $\mathbb{E}_F(D_{S^\sharp} + D_s) \geq \mathbb{E}_F(D_{S^\sharp})$. For the variance, however, it is possible that $\mathbb{V}_F(D_{S^\sharp}) \geq \mathbb{V}_F(D_{S^\sharp} + D_s)$, namely whenever $\mathbb{V}_F(s) \leq -2\text{Cov}_F(D_{S^\sharp}, D_s)$. Hence, it might be that the standardized score divergence $\tilde{\mathbb{D}}_{S^\sharp+s}(F\|G)$ based on the composite scoring rule actually is smaller than the corresponding $\tilde{\mathbb{D}}_{S^\sharp}(F\|G)$ based on the original conditional rule.

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

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

G.1 Size

Following Diks et al. (2011), we assess the size properties of the DM test statistic using an i.i.d. standard normal density as DGP. The two candidate density forecasts f and g also are normal with unit variance, but with means $\mu_f = -0.2$ and $\mu_g = 0.2$. The evaluation is focused on the central region around the true mean 0 by using $I_C(y; r)$ as weight function. Due to the symmetric design, the norms and \bar{F}_w -probabilities of the candidates f and g are equivalent, such that the test statistics based on QS and SphS scoring rules coincide. The equal norms and discrete probabilities also imply that the censoring and conditioning rules are equivalent within a semi-local scoring rule family, because in this case observations outside the region of interest obtain the same scores under both candidates. By the same token, the auxiliary scoring rules sbar and slog of the composite scoring rule of Holzmann and Klar (2017a) become equivalent to the conditional rules of both semi-local rules and the CRPS. Given these observations, the 16 weighted scoring rules arising from the combination of the unweighted scoring rules {LogS, SphS, QS, CRPS} with the conditioning, censoring, conditioning appended with sbar and conditioning appended with slog localization techniques, can be represented by the censored LogS, SphS, CRPS and conditional CRPS. The twCRPS is added because it will be included as a benchmark in the power studies based on weight functions for which the censored CRPS variants do

not reduce to the twCRPS.

Figure G.1: Size properties of the DM test

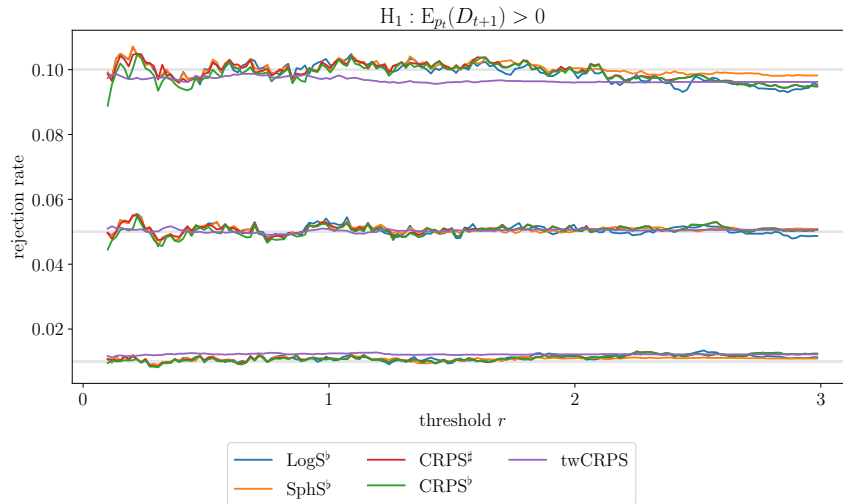


Figure G.1 displays the rejection rates of the null of equal predictive ability against the one-sided alternative that candidate f is statistically closer to p than g as a function of the threshold r for a fixed sample size $T = 500$. The rejection rates are given at nominal significance levels 0.01,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

G.2 Power

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

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

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

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

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

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

Normal versus Student- t : Center. In our second power experiment, we again consider standard normal and Student- t_5 densities as the competing forecasts, but now focus on

the central region by using $I_C(y; r) = \mathbb{1}_{(-r, r)}(y)$ as weight function. We also increase T , so that the expected number of observations in the region of interest is equal to $c = 200$. Figure G.4 displays the rejection rates for the same selection of regular scoring rules and focusing procedures as in the first power experiment.

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

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

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

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

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

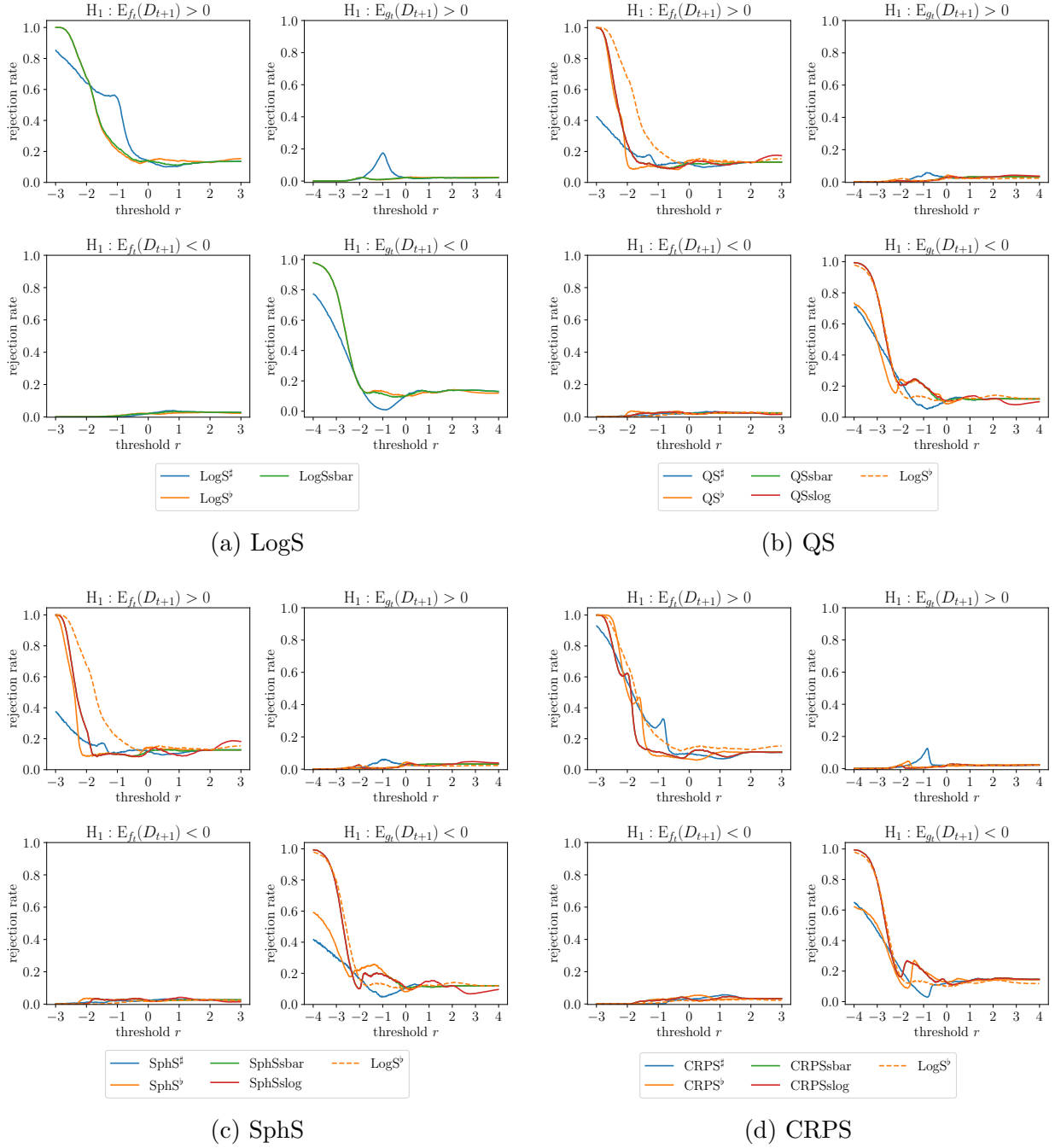
Laplace tails In our third and final power experiment we study the consequences of the inability of the conditional scoring rule to distinguish two proportional tails when using the left-tail indicator $I_L(y; r) = \mathbb{1}_{(-\infty, r)}(y)$ as weight function; see Example F.1. In particular, we consider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

power against the distributions being statistically equally far away from the actual density on A_w through the lens of the regular score divergence and, therefore, still a lack of local discriminative ability. More fundamentally, the test statistic degenerates in this case, as the score differences are exactly zero.

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

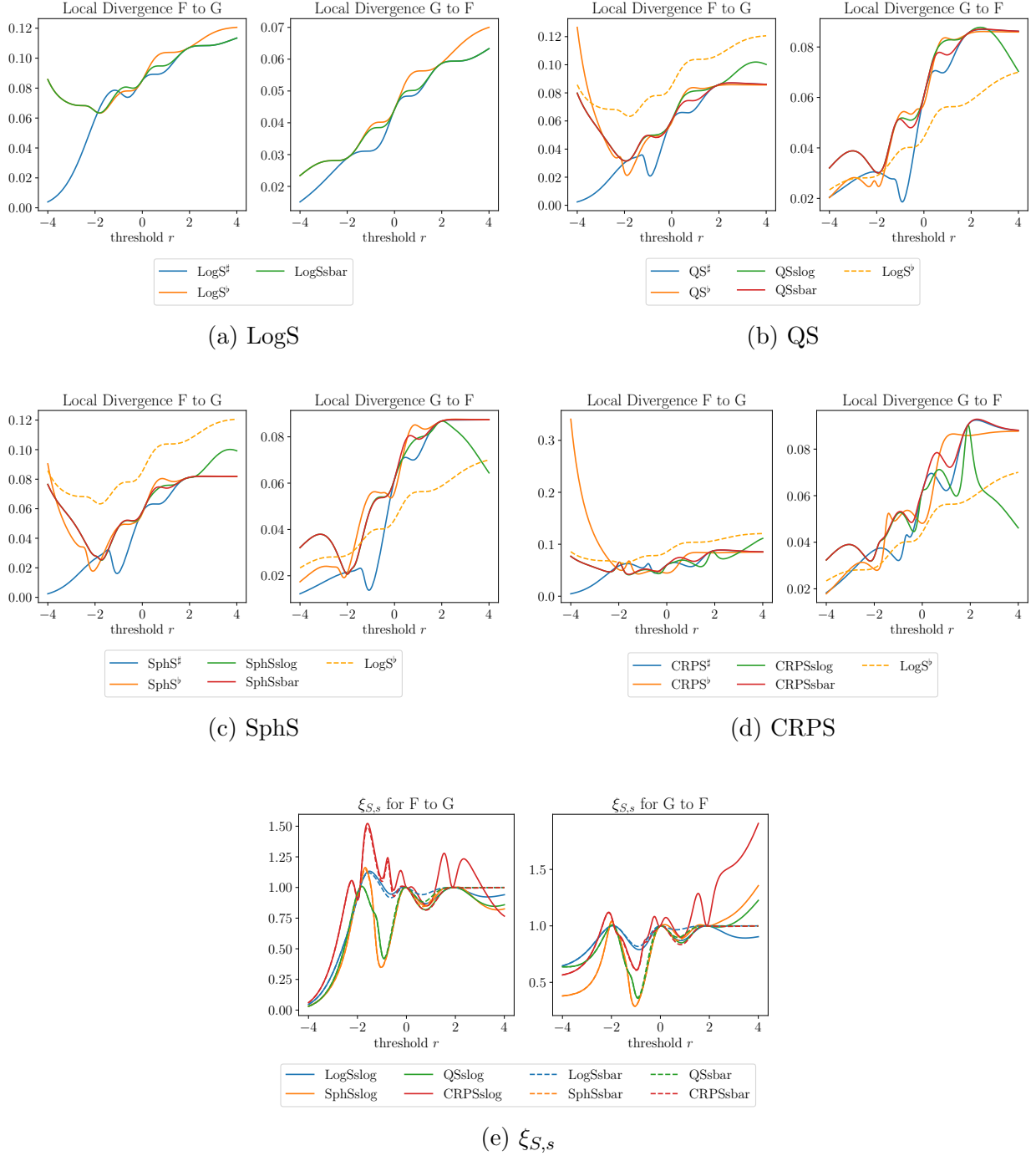
Four observations are clearly apparent. First, the increase in true power when moving from the conditional operator to the censoring operator is immense for all four scoring rules and thresholds $r < \mu_f$. The difference decreases over the interval $r \in (\mu_f, \mu_g)$, after which both conditioning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

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



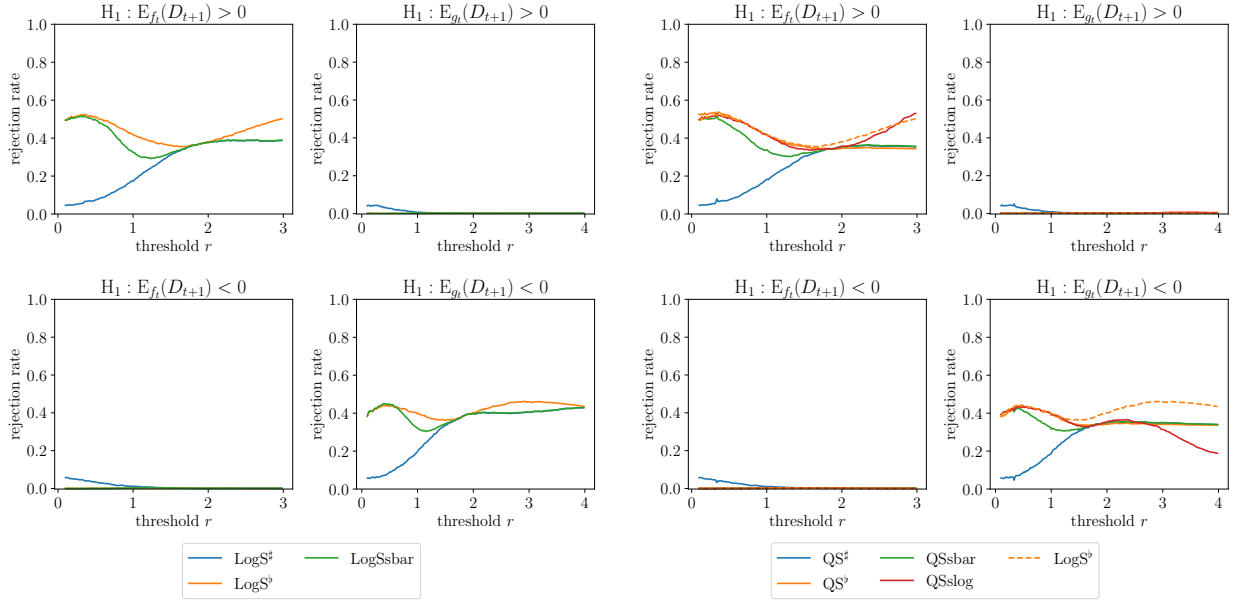
One-sided rejection rates of the DM test of equal predictive ability of the candidates f_t (standard normal) and g_t (Student- t_5) at a nominal significance level of 0.05 based on 10,000 simulations. The DGP is either f_t (left-hand side) or g_t (right-hand side). Moreover, rejections in the top panels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

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



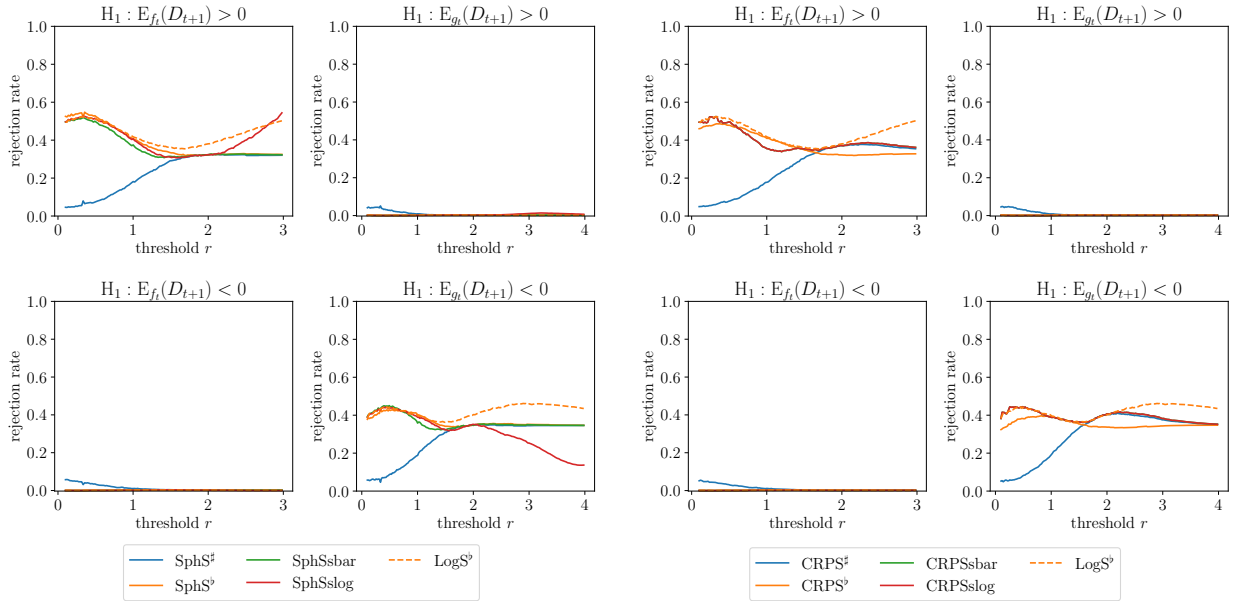
Standardized divergences ($\tilde{\mathbb{D}}_S(f_t||g_t)$) and ratios ($\xi_{S,s}$) of the candidates f_t ($\mathcal{N}(0, 1)$) and g_t (Student- t_5), for weighted versions of unweighted scoring rules $S \in \{\text{LogS}, \text{QS}, \text{SphS}, \text{CRPS}\}$. Each plot includes the conditional and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

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



(a) LogS

(b) QS

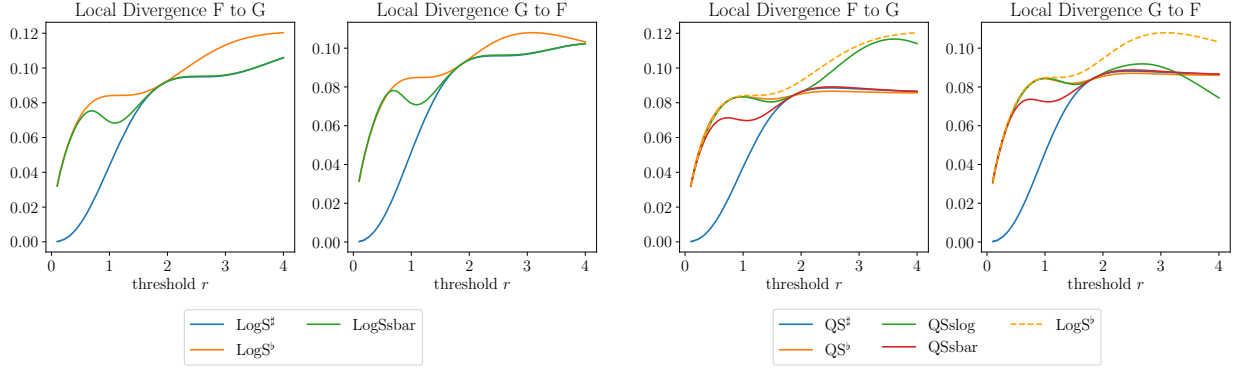


(c) SphS

(d) CRPS

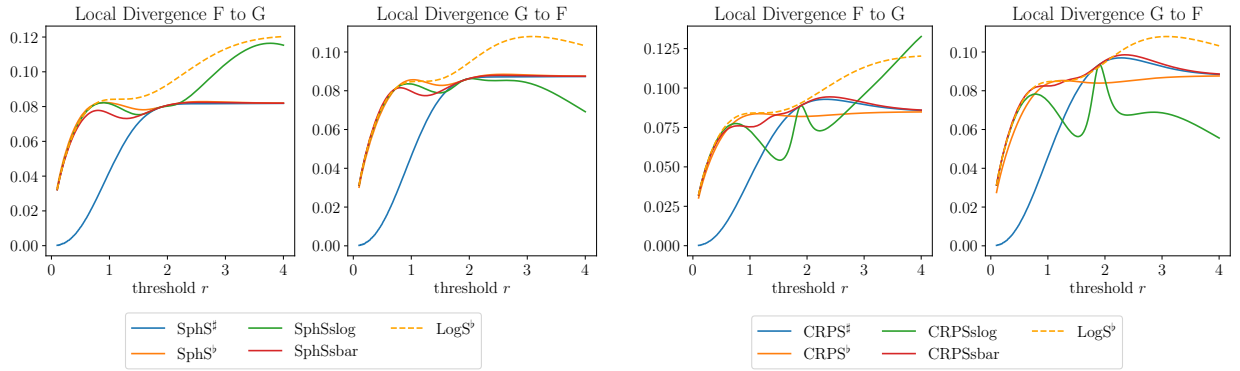
One-sided rejection rates of the DM test of equal predictive ability of the candidates f_t (standard normal) and g_t (Student- t_5) at a nominal significance level of 0.05 based on 10,000 simulations. The DGP is either f_t (left-hand side) or g_t (right-hand side). Moreover, rejections in the top panels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

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



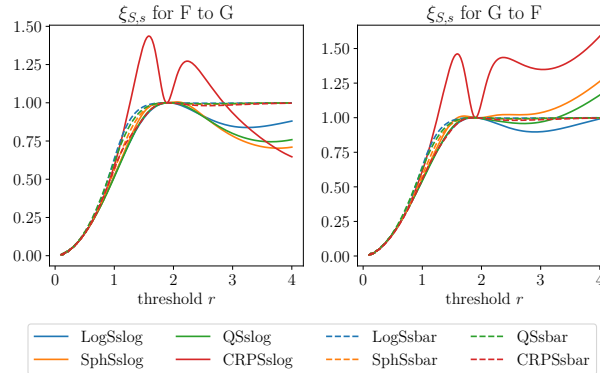
(a) LogS

(b) QS



(c) SphS

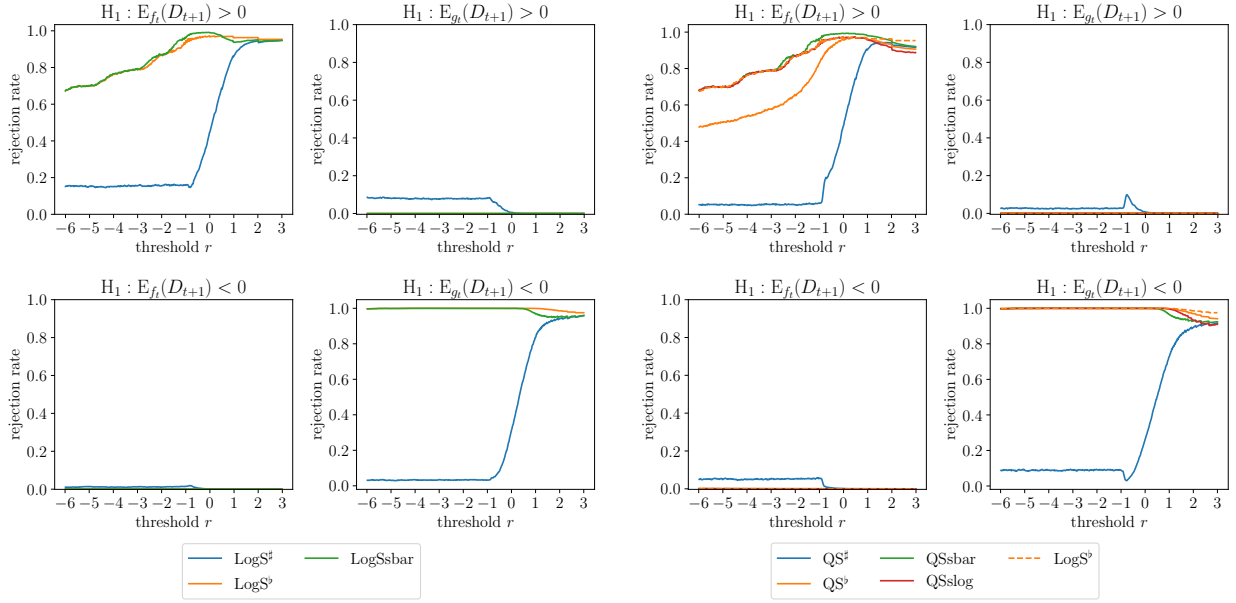
(d) CRPS



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

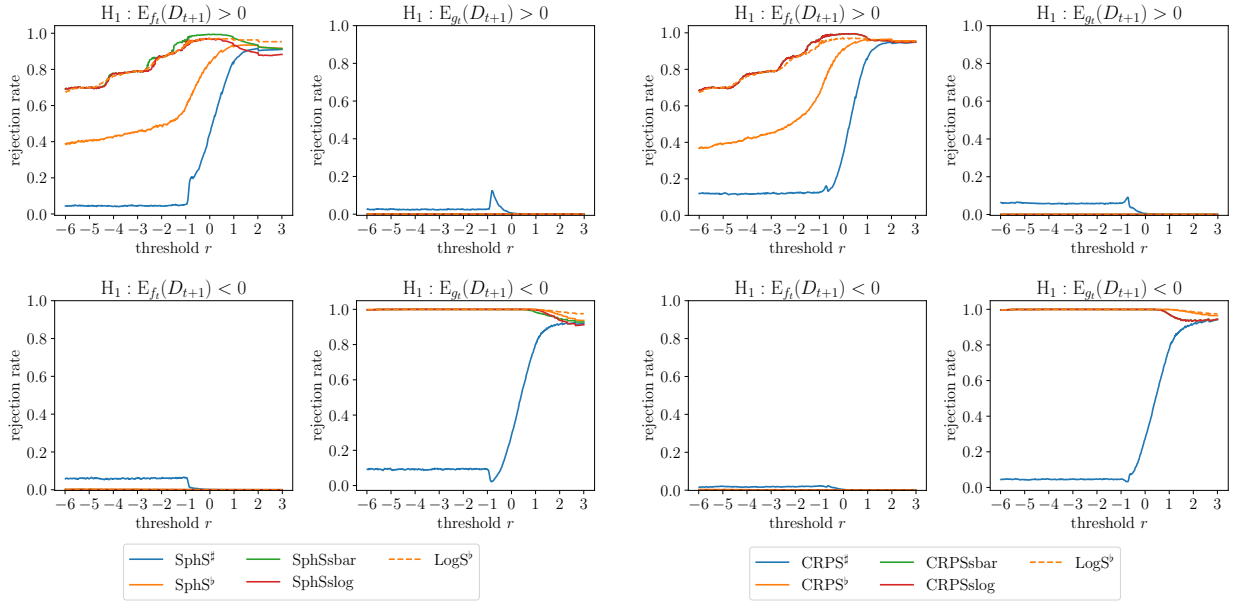
Standardized divergences ($\tilde{\mathbb{D}}_S(f_t||g_t)$) and ratios ($\xi_{S,s}$) of the candidates f_t ($\mathcal{N}(0, 1)$) and g_t (Student- t_5), for weighted versions of unweighted scoring rules $S \in \{\text{LogS}, \text{QS}, \text{SphS}, \text{CRPS}\}$. Each plot includes the conditional and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

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



(a) LogS

(b) QS

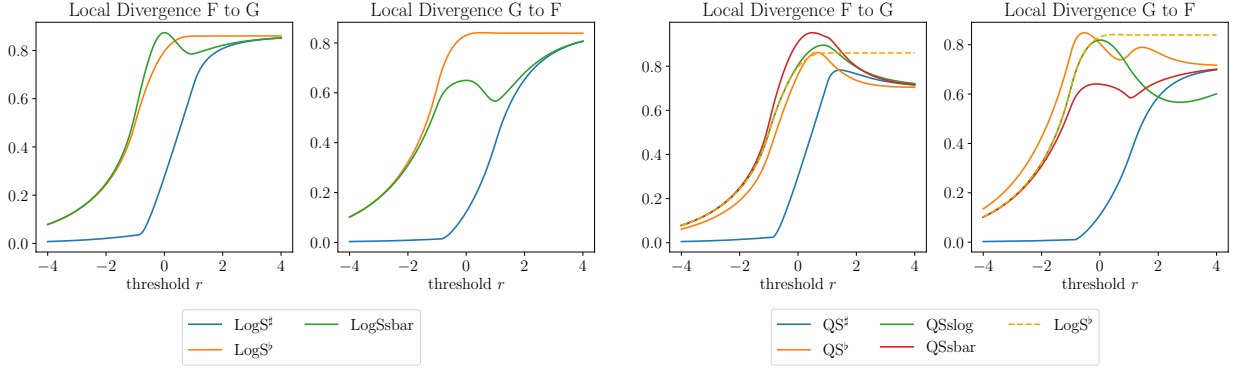


(c) SphS

(d) CRPS

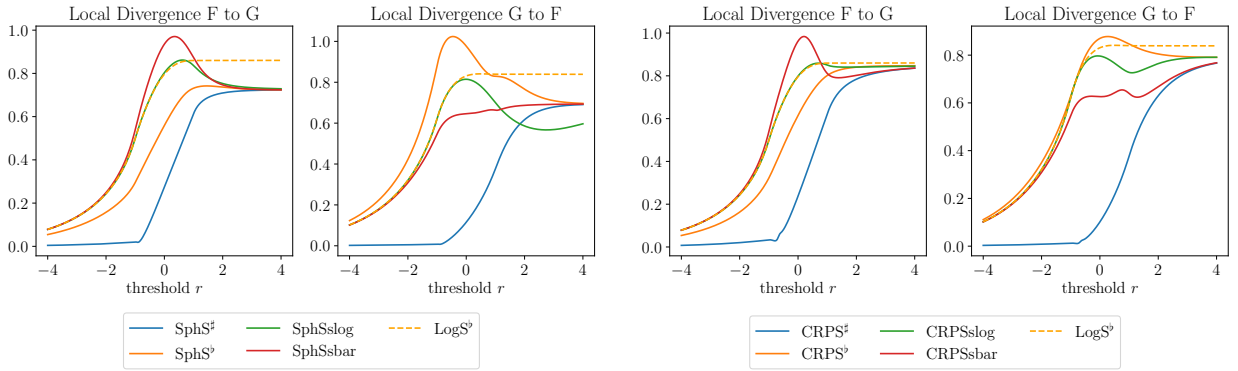
One-sided rejection rates of the DM test of equal predictive ability of the candidates f_t (Laplace(-1, 1)) and g_t (Laplace(1, 1.1)) at a nominal significance level of 0.05 based on 10,000 simulations. The DGP is either f_t (left-hand side) or g_t (right-hand side). Moreover, rejections in the top panels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

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



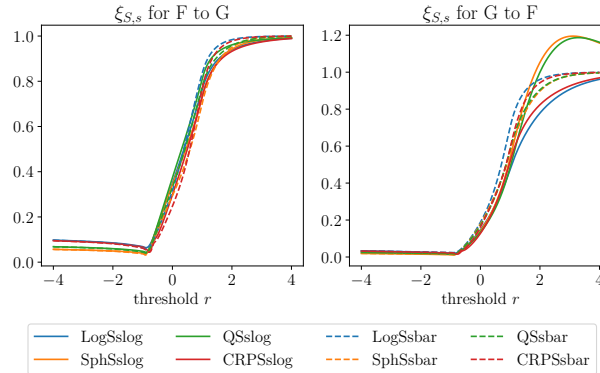
(a) LogS

(b) QS



(c) SphS

(d) CRPS



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

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

H Details on Model Specifications

H.1 Financial risk management

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

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

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

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

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

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

where x_t represents a realized measure¹, $z_t = (y_t - \mu)/\sigma_t$, and u_t denotes a white noise process with variance σ_u^2 . We combine each of the volatility specifications with standard normal and Student- t_ν distributions. Parameter estimates are obtained via maximum-likelihood estimation (MLE).

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

¹Obtained from <https://dachxiu.chicagobooth.edu/#risklab>

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

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

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

H.2 Macroeconomics

Following Stock and Watson (2002), among many others, we construct direct forecasts for annualized τ -month inflation rates $y_{t+\tau}^\tau = (1, 200/\tau) \log(P_{t+\tau}/P_t)$, where P_t denotes the U.S. consumer price index (CPI) in month t , for horizons $\tau = 6$ and 24. Each of the forecast methods we consider can be represented as $y_{t+\tau}^\tau = \mu_{t+\tau}^\tau(\mathbf{x}_t) + u_{t+\tau}^\tau$, where \mathbf{x}_t denotes a subset of $n = 122$ variables from the FRED-MD² database, cf. Medeiros et al. (2021). We use

²See <https://research.stlouisfed.org/econ/mccracken/fred-databases/>

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

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

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

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

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

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

As a fourth forecast method, we include Complete Subset Regression (CSR) developed by Elliott et al. (2013, 2015). The key idea of this approach is to average the forecasts obtained from linear regressions for all possible combinations of q predictors selected out of n available candidates, where typically q is set (much) smaller than n . With $n = 122$ as

in our case, however, the number of regressions to be estimated is impractically large. We therefore adopt a pre-testing approach: First estimate a linear regression of $y_{t+\tau}^\tau$ on each of the candidate predictors individually, rank the t -statistics of their coefficients by absolute value, and select the \tilde{n} highest-ranked variables. Then, apply the CSR approach to this subset of variables. We set $\tilde{n} = 20$ and $q = 4$.

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

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

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

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

Finally, for each of the six forecast methods considered for $\mu_{t+\tau}^\tau$, we obtain a density forecast for $y_{t+\tau}^\tau = \mu_{t+\tau}^\tau + u_{t+\tau}^\tau$ by assuming that $u_{t+\tau}^\tau$ follows a two-piece normal distribution,

with density given by

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

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

H.3 Climate

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

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

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

I Additional Tables

I.1 MCS table at confidence level 0.75

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

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

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

I.2 Financial risk management

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

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

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

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

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

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

NOTE: This table continues on the next page.

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

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

NOTE: This table continues on the next page.

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

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

NOTE: This table continues on the next page.

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

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

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

Table I.3: Robustness analysis MCS cardinality

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

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

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

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

NOTE: This table continues on the next page.

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

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

NOTE: This table continues on the next page.

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

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

NOTE: This table continues on the next page.

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

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

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

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

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

NOTE: This table continues on the next page.

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

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

NOTE: This table continues on the next page.

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

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

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

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

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

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

I.3 Macroeconomics

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

Table I.6: MCS p -values for center inflation application.

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

NOTE: This table continues on the next page.

Table I.6 (continued): MCS p -values for center inflation application.

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

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

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

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

NOTE: This table continues on the next page.

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

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

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

I.4 Climate

The MCS p -values for the climate application are reported in Table I.8 (right tail application) and Table I.9 (center application). The MCS p -values in Table I.9 reveal that the MCSs are consistently small in the center application, frequently including one or both of the QGARCH-II methods. This observation suggests that the preference for censoring, as depicted in Table 2, translates into the censored scoring rule's more effective recognition of the QGARCH-II methods' pronounced superiority. Table I.9 further demonstrates that the performance of the CRPS^b and twCRPS is closely matched.

Table I.8: MCS p -values for right tail climate application.

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

NOTE: This table continues on the next page.

Table I.8 (continued): MCS p -values for right tail climate application (Continued).

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

NOTE: This table continues on the next page.

Table I.8 (continued): MCS p -values for right tail climate application (Continued).

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

NOTE: This table continues on the next page.

Table I.8 (continued): MCS p -values for right tail climate application (Continued).

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

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

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

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

NOTE: This table continues on the next page.

Table I.9: MCS p -values for center climate application (Continued).

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

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

References

- Allen, S., D. Ginsbourger, and J. Ziegel (2023), “Evaluating Forecasts for High-Impact Events Using Transformed Kernel Scores”, *SIAM/ASA Journal on Uncertainty Quantification*, 11(3), 906–940.
- Amisano, G. and R. Giacomini (2007), “Comparing Density Forecasts via Weighted Likelihood Ratio Tests”, *Journal of Business & Economic Statistics*, 25(2), 177–190.
- Bernardi, M. and L. Catania (2018), “The Model Confidence Set Package for R”, *International Journal of Computational Economics and Econometrics*, 8(2), 144–158.
- Bollerslev, T. (1986), “Generalized Autoregressive Conditional Heteroskedasticity”, *Journal of Econometrics*, 31(3), 307–327.
- Breiman, L. (2001), “Random forests”, *Machine Learning*, 45(1), 5–32.
- Clements, M. P. (2004), “Evaluating the Bank of England Density Forecasts of Inflation”, *The Economic Journal*, 114(498), 844–866.
- Diebold, F. X. and R. S. Mariano (2002), “Comparing Predictive Accuracy”, *Journal of Business & Economic Statistics*, 20(1), 134–144.
- Diks, C., V. Panchenko, and D. Van Dijk (2011), “Likelihood-based Scoring Rules for Comparing Density Forecasts in Tails”, *Journal of Econometrics*, 163(2), 215–230.
- Elliott, G., A. Gargano, and A. Timmermann (2013), “Complete subset regressions”, *Journal of Econometrics*, 177(2), 357–373.
- Elliott, G., A. Gargano, and A. Timmermann (2015), “Complete subset regressions with large-dimensional sets of predictors”, *Journal of Economic Dynamics and Control*, 54, 86–110.
- Engle, R. (2002), “Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models”, *Journal of Business & Economic Statistics*, 20(3), 339–350.
- Franses, P. H., J. Neele, and D. Van Dijk (2001), “Modeling Asymmetric Volatility in Weekly Dutch Temperature Data”, *Environmental Modelling & Software*, 16(2), 131–137.
- Giacomini, R. and H. White (2006), “Tests of Conditional Predictive Ability”, *Econometrica*, 74(6), 1545–1578.
- Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993), “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks”, *The Journal of Finance*, 48(5), 1779–1801.
- Gneiting, T. and A. E. Raftery (2007), “Strictly Proper Scoring Rules, Prediction, and Estimation”, *Journal of the American Statistical Association*, 102(477), 359–378.
- Gneiting, T. and R. Ranjan (2011), “Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules”, *Journal of Business & Economic Statistics*, 29(3), 411–422.
- Hansen, P. R., Z. Huang, and H. H. Shek (2012), “Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility”, *Journal of Applied Econometrics*, 27(6), 877–906.
- Hansen, P. R., A. Lunde, and J. Nason (2011), “The Model Confidence Set”, *Econometrica*, 79(2), 453–497.

- Holzmann, H. and B. Klar (2017a), “Focusing on Regions of Interest in Forecast Evaluation”, *The Annals of Applied Statistics*, 11(4), 2404–2431.
- Holzmann, H. and B. Klar (2017b). “Weighted Scoring Rules and Hypothesis Testing”. Available at <https://arxiv.org/abs/1611.07345v2>.
- Lerch, S., T. L. Thorarinsdottir, F. Ravazzolo, and T. Gneiting (2017), “Forecaster’s Dilemma: Extreme Events and Forecast Evaluation”, *Statistical Science*, 32(1), 106–127.
- Medeiros, M. C., G. F. R. Vasconcelos, A. Veiga, and E. Zilberman (2021), “Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods”, *Journal of Business & Economic Statistics*, 39(1), 98–119.
- Mitchell, J. and S. G. Hall (2005), “Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR ‘Fan’ Charts of Inflation”, *Oxford Bulletin of Economics and Statistics*, 67(s1), 995–1033.
- Neyman, J. and E. Pearson (1933), “IX. On the Problem of the Most Efficient Tests of Statistical Hypotheses”, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706), 289–337.
- Stock, J. H. and M. W. Watson (2002), “Macroeconomic Forecasting Using Diffusion Indexes”, *Journal of Business & Economic Statistics*, 20(2), 147–162.