Supplemental Appendix to:

"Comparing Possibly Misspecified Forecasts"

Andrew J. Patton

Duke University

26 January 2019

This supplemental appendix contains two parts. Appendix SA.1 contains proofs of the propositions presented in the main paper. Appendix SA.2 contains derivations used in the analytical results presented in the paper.

Appendix SA.1: Proofs

Proof of Proposition 1(a). We will show that under Assumptions (1)–(3), $MSE_B \geq MSE_A \Rightarrow \mathcal{F}_t^B \subseteq \mathcal{F}_t^A \ \forall t \Rightarrow \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^B\right)\right] \geq E\left[L\left(Y_t, \hat{Y}_t^A\right)\right] \ \forall L \in \mathcal{L}_{Bregman}$.

For the first implication: Assume that $\mathcal{F}_t^A \subseteq \mathcal{F}_t^B \ \forall t$. This implies $\mathbb{E}\left[\left(Y_t - \hat{Y}_t^A\right)^2 | \mathcal{F}_t^B\right] \geq \mathbb{E}\left[\left(Y_t - \hat{Y}_t^B\right)^2 | \mathcal{F}_t^B\right]$ $a.s. \forall t$ since $\hat{Y}_t^A \in \mathcal{F}_t^A \subseteq \mathcal{F}_t^B$. Then $\mathbb{E}\left[\left(Y_t - \hat{Y}_t^A\right)^2\right] \geq \mathbb{E}\left[\left(Y_t - \hat{Y}_t^B\right)^2\right]$ by the law of iterated expectations (LIE). The only way that this can also satisfy the first assumption that $MSE_B \geq MSE_A$ is under equality: $MSE_B = MSE_A$. Since $\mathbb{E}\left[\left(Y_t - \hat{Y}_t^A\right)^2 | \mathcal{F}_t^B\right] \geq \mathbb{E}\left[\left(Y_t - \hat{Y}_t^B\right)^2 | \mathcal{F}_t^B\right]$ $a.s. \forall t$, equality of (unconditional) MSEs can only obtain under equality of conditional MSEs at each point in time, i.e. $\mathbb{E}\left[\left(Y_t - \hat{Y}_t^A\right)^2 | \mathcal{F}_t^B\right] = \mathbb{E}\left[\left(Y_t - \hat{Y}_t^B\right)^2 | \mathcal{F}_t^B\right]$ $a.s. \forall t$, which in turn can only hold if $\hat{Y}_t^A = \hat{Y}_t^B$ $a.s. \forall t$, violating the "not identical" part of Assumption (1). Thus we have a contradiction, and so under Assumptions (1)-(3), $MSE_B \geq MSE_A \Rightarrow \mathcal{F}_t^B \subseteq \mathcal{F}_t^A \ \forall t$.

Now consider the second implication: Let

$$Y_t = \hat{Y}_t^A + \eta_t = \hat{Y}_t^B + \eta_t + \varepsilon_t \tag{15}$$

Then

$$\begin{aligned} \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^A\right) - L\left(Y_t, \hat{Y}_t^B\right)\right] &= \mathbb{E}\left[-\phi\left(\hat{Y}_t^A\right) - \phi'\left(\hat{Y}_t^A\right)\eta_t + \phi\left(\hat{Y}_t^B\right) + \phi'\left(\hat{Y}_t^B\right)(\eta_t + \varepsilon_t)\right] \\ &= \mathbb{E}\left[\phi\left(\hat{Y}_t^B\right) - \phi\left(\hat{Y}_t^A\right)\right] \end{aligned} \tag{16}$$

since $\mathbb{E}\left[\phi'\left(\hat{Y}_t^A\right)\eta_t\right] = \mathbb{E}\left[\phi'\left(\hat{Y}_t^A\right)\mathbb{E}\left[\eta_t|\mathcal{F}_t^A\right]\right]$ by the LIE and $\mathbb{E}\left[\eta_t|\mathcal{F}_t^A\right] = 0$, by Assumptions (2)-(3). Similarly for $\mathbb{E}\left[\phi'\left(\hat{Y}_t^B\right)(\eta_t + \varepsilon_t)\right]$. Next, consider the second-order mean-value expansion:

$$\phi\left(\hat{Y}_t^A\right) = \phi\left(\hat{Y}_t^B\right) - \phi'\left(\hat{Y}_t^B\right)\varepsilon_t + \phi''\left(\ddot{Y}_t^A\right)\varepsilon_t^2 \tag{17}$$

where $\ddot{Y}_t^A = \lambda_t\hat{Y}_t^A + (1-\lambda_t)\hat{Y}_t^B$, for $\lambda_t \in [0,1]$. Thus

$$\mathbb{E}\left[L\left(Y_t, \hat{Y}_t^A\right) - L\left(Y_t, \hat{Y}_t^B\right)\right] = \mathbb{E}\left[\phi'\left(\hat{Y}_t^B\right)\varepsilon_t\right] - \mathbb{E}\left[\phi''\left(\ddot{Y}_t^A\right)\varepsilon_t^2\right] \leq 0 \tag{18}$$

since $\mathbb{E}\left[\phi'\left(\hat{Y}_t^B\right)\varepsilon_t\right] = 0$ and ϕ is convex. And so $\mathcal{F}_t^B \subseteq \mathcal{F}_t^A\ \forall t \Rightarrow \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^B\right)\right] \geq E\left[L\left(Y_t, \hat{Y}_t^A\right)\right]\ \forall L \in \mathcal{L}_{Bregman}$. ■

Proof of Proposition 2. First we note that

$$\begin{aligned}
\mathbb{E}\left[L\left(Y_t, \hat{Y}_t^i; a\right)\right] &= \frac{2}{a^2}\left(\mathbb{E}\left[\exp\{aY_t\}\right] - \mathbb{E}\left[\exp\left\{a\hat{Y}_t^i\right\}\right]\right) \\
&= \frac{2}{a^2}\left(\exp\left\{\frac{a}{2}\left(a\sigma^2 + 2\mu\right)\right\} - \exp\left\{\frac{a}{2}\left(a\omega_i^2 + 2\mu\right)\right\}\right) \\
&\rightarrow \sigma^2 - \omega_i^2 \quad \text{as } a \rightarrow 0.
\end{aligned}$$

where the first equality holds under mean-unbiasedness (assumption (ii)) and the second follows from normality of the target variable and the forecast (assumption (i)). The last line implies that whichever forecast is based on the richest information set, leading to the greatest (optimal) variability in the forecast (ω_i^2), will have the lowest MSE loss. Then note that for non-MSE exponential Bregman loss (i.e., for $a \neq 0$), that if $\mathbb{E}\left[L\left(Y_t, \hat{Y}_t^A; a\right)\right] \geq \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^B; a\right)\right]$, then $\exp\left\{\frac{a}{2}\left(a\omega_A^2 + 2\mu\right)\right\} \leq \exp\left\{\frac{a}{2}\left(a\omega_B^2 + 2\mu\right)\right\}$ and so $\omega_A^2 \leq \omega_B^2$ and thus $MSE_A \geq MSE_B$. The converse holds using the same derivations, proving the proposition. ■

Proof of Proposition 3(a). The first-order condition for the optimization is:

$$\begin{aligned}
0 &= \left.\frac{\partial}{\partial\theta}\mathbb{E}\left[L\left(Y_t, m\left(X_t;\theta\right);\phi\right)\right]\right|_{\theta=\hat{\theta}_\phi^*} \\
&= \mathbb{E}\left[\phi''\left(m\left(X_t;\hat{\theta}_\phi^*\right)\right)\left(Y_t - m\left(X_t;\hat{\theta}_\phi^*\right)\right)\frac{\partial m\left(X_t;\hat{\theta}_\phi^*\right)}{\partial\theta}\right] \\
&= \mathbb{E}\left[\phi''\left(m\left(X_t;\hat{\theta}_\phi^*\right)\right)\left(\mathbb{E}\left[Y_t|\mathcal{F}_t\right] - m\left(X_t;\hat{\theta}_\phi^*\right)\right)\frac{\partial m\left(X_t;\hat{\theta}_\phi^*\right)}{\partial\theta}\right]
\end{aligned}$$

where the last equality holds by the LIE. Note that the first-order condition is satisfied when $\hat{\theta}^*_\phi = \theta_0$ by assumption (i), and the solution is unique since ϕ is strictly convex and $\partial m/\partial\theta \neq 0$ a.s. by assumption (ii). ■

Proof of Proposition 5. (a) The first-order condition for an optimal forecast based on a convex combination of Bregman and GPL$^{1/2}$ loss is:

$$0 = -\lambda\phi''\left(\hat{Y}_t^*\right)\left(\mathbb{E}_{t-1}\left[Y_t\right] - \hat{Y}_t^*\right) + (1-\lambda)\left(\mathbb{E}_{t-1}\left[\mathbf{1}\left\{Y_t \leq \hat{Y}_t^*\right\}\right] - 1/2\right) g'\left(\hat{Y}_t^*\right)$$

using the assumption that F_t^* is continuous. Then note that $\mathbb{E}_{t-1}\left[\mathbf{1}\left\{Y_t \leq \hat{y}\right\}\right] \equiv F_t^*\left(\hat{y}\right)$, and recall that F_t^* is symmetric, which implies that $\mathbb{E}_{t-1}\left[Y_t\right] = Median_{t-1}\left[Y_t\right]$ and that $F_t^*\left(\mathbb{E}_{t-1}\left[Y_t\right]\right) = 1/2$. Thus $\hat{Y}_t^* = \mathbb{E}_{t-1}\left[Y_t\right]$ is a solution to the optimization problem, and this solution is unique as ϕ is strictly convex and g is strictly increasing.

(b) From the proofs of Propositions 1(a) and 4(a), we know that under Assumptions (1)–(3) we have $MSE_B \geq MSE_A \Rightarrow \mathcal{F}_t^B \subseteq \mathcal{F}_t^A\ \forall t$ and $MAE_B \geq MAE_A \Rightarrow \mathcal{F}_t^B \subseteq \mathcal{F}_t^A\ \forall t$, and that $\mathcal{F}_t^B \subseteq \mathcal{F}_t^A\ \forall t \Rightarrow \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^B\right)\right] \geq \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^A\right)\right]\ \forall\ L \in \left\{\mathcal{L}_{Bregman}\ ,\ \mathcal{L}_{GPL}^{1/2}\right\}$. This immediately yields $\mathcal{F}_t^B \subseteq \mathcal{F}_t^A\ \forall t \Rightarrow \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^B\right)\right] \geq \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^A\right)\right]$ for any $L \in \mathcal{L}_{Breg\times GPL}$ since $\lambda \in [0,1]$.

(c) The proof of this negative result requires only an example. This can be constructed using methods similar to those for Propositions 1(b) and 4(b), and is omitted in the interest of brevity. ■

The proofs of Propositions 4 and 6 below use the following results on uniform random variable, and a "triangular" random variable with a mode at L and a PDF that declines linearly to zero at $U > L$.

Uniform random variable	**"Triangular" random variable**	
$X \sim Unif\,(L, U)$	$Z \sim Tri\,(L, U)$	(1)
$F_x\,(x) = \begin{cases} 0, & z < L \\ \frac{x-L}{U-L}, & z \in [L, U] \\ 1, & z > U \end{cases}$	$F_z\,(z) = \begin{cases} 0, & z < L \\ \frac{(U-L)^2-(U-z)^2}{(U-L)^2}, & z \in [L, U] \\ 1, & z > U \end{cases}$	(2)
$f_x\,(x) = \begin{cases} \frac{1}{U-L}, & z \in [L, U] \\ 0 & else \end{cases}$	$f_z\,(z) = \begin{cases} \frac{2(U-x)}{(U-L)^2}, & z \in [L, U] \\ 0 & else \end{cases}$	(3)
$F_x^{-1}\,(\alpha) = L + \alpha\,(U - L)\,,$ for $\alpha \in [0, 1]$	$F_z^{-1}\,(\alpha) = U - (U - L)\sqrt{1-\alpha}$, for $\alpha \in [0, 1]$	(4)
$\mathbb{E}\,[X] = \frac{1}{2}\,(U + L)$	$\mathbb{E}\,[Z] = \frac{1}{3}\,(2L + U)$	(5)
$\mathbb{E}\left[X^2\right] = \frac{1}{3}\left(L^2 + U^2 + LU\right)$	$\mathbb{E}\left[Z^2\right] = \frac{1}{6}\left(3L^2 + 2LU + U^2\right)$	(6)
$\mathbb{E}\left[X^3\right] = \frac{1}{4}\left(L^3 + U^3 + L^2U + LU^2\right)$	$\mathbb{E}\left[Z^3\right] = \frac{1}{10}\left(4L^3 + 3L^2U + 2LU^2 + U^3\right)$	(7)
$M_x \equiv Median\,[X] = \frac{1}{2}\,(U + L)$	$M_z \equiv Median\,[Z] = U - \frac{U-L}{\sqrt{2}}$	(8)
$\mathbb{E}\,[\mathbf{1}\,\{X < b\}\,X] = \frac{b^2-L^2}{2(U-L)}$, for $b \in [L, U]$	$\mathbb{E}\,[\mathbf{1}\,\{Z \leq b\}\,Z] = \frac{3b^2U-2b^3-L^2(3U-2L)}{3(U-L)^2}$, for $b \in [L, U]$	(9)
$\mathbb{E}\left[\mathbf{1}\,\{X < b\}\,X^2\right] = \frac{b^3-L^3}{3(U-L)}$	$\mathbb{E}\left[\mathbf{1}\,\{Z \leq b\}\,Z^2\right] = \frac{4b^3U-3b^4-L^3(4U-3L)}{6(U-L)^2}$	(10)
$\mathbb{E}\left[\mathbf{1}\,\{X < b\}\,X^3\right] = \frac{b^4-L^4}{4(U-L)}$	$\mathbb{E}\left[\mathbf{1}\,\{Z \leq b\}\,Z^3\right] = \frac{5b^4U-4b^5-L^4(5U-4L)}{10(U-L)^2}$	(11)
$\mathbb{E}\,[\mathbf{1}\,\{X < M_x\}\,X] = \frac{3L+U}{8}$	$\mathbb{E}\,[\mathbf{1}\,\{Z \leq M_z\}\,Z] = \frac{U\left(\sqrt{2}-1\right)+L\left(4-\sqrt{2}\right)}{6}$	(12)
$\mathbb{E}\left[\mathbf{1}\,\{X < M_x\}\,X^2\right] = \frac{7L^2+4LU+U^2}{24}$	$\mathbb{E}\left[\mathbf{1}\,\{Z \leq M_z\}\,Z^2\right] = \frac{9L^2+2LU\left(7-4\sqrt{2}\right)+U^2\left(8\sqrt{2}-11\right)}{24}$	(13)
$\mathbb{E}\left[\mathbf{1}\,\{X < M_x\}\,X^3\right] = \frac{(3L+U)\left(5L^2+2LU+U^2\right)}{64}$	$\mathbb{E}\left[\mathbf{1}\,\{Z \leq M_z\}\,Z^3\right] = \frac{2L^3\left(8-\sqrt{2}\right)+3L^2U\left(2\sqrt{2}-1\right)}{40} + \frac{2LU^2\left(19-13\sqrt{2}\right)+U^3\left(22\sqrt{2}-31\right)}{40}$	(14)

Part of the proof of Proposition 4(b)(ii) below uses the distribution of $Y = X + Z$, where $X \sim Unif\,(L, 0)$ and $Z \sim Unif\,(0, U)$, where $L < 0 < |L| < U$. This variable has the following properties:

$$F_y(y) = \begin{cases} 0, & y < L \\ \frac{(L-y)^2}{-2LU} & y \in [L, 0] \\ \frac{2y-L}{2U} & y \in [0, L+U] \\ \frac{-2LU-(U-y)^2}{-2LU} & y \in [L+U, U] \\ 1, & y > U \end{cases} \quad \text{and} \quad f_z(z) = \begin{cases} \frac{y-L}{-LU}, & y \in [L, 0] \\ \frac{1}{U}, & y \in [0, L+U] \\ \frac{U-y}{-LU} & y \in [L+U, U] \\ 0, & else \end{cases} \tag{19}$$

$$F_z^{-1}(\alpha) = \begin{cases} L + \sqrt{2\alpha |L| U}, & \alpha \in \left[0, \frac{-L}{2U}\right] \\ \frac{1}{2}L + \alpha U, & \alpha \in \left[\frac{-L}{2U}, \frac{L+2U}{2U}\right] \\ U - \sqrt{2(1-\alpha)|L| U}, & \alpha \in \left[\frac{L+2U}{2U}, 1\right] \end{cases} \tag{20}$$

And then

$$\begin{aligned}
\mathbb{E}[Y] &= \frac{1}{2}(L+U) && (21) \\
\mathbb{E}\left[Y^2\right] &= \frac{1}{6}\left(2L^2 + 3LU + 2U^2\right) && (22) \\
\mathbb{E}\left[Y^3\right] &= \frac{1}{4}\left(L^3 + 2L^2U + 2LU^2 + U^3\right) \\
M_y &\equiv Median\,[Y] = \frac{1}{2}(L+U) && (23) \\
\mathbb{E}\left[\mathbf{1}\{Y \le M_y\} Y\right] &= \frac{1}{24}\left(6L - \frac{L^2}{U} + 3U\right) && (24) \\
\mathbb{E}\left[\mathbf{1}\{Y \le M_y\} Y^2\right] &= \frac{(L+U)^3 - 2L^3}{24U} && (25) \\
\mathbb{E}\left[\mathbf{1}\{Y \le M_y\} Y^3\right] &= \frac{5(L+U)^4 - 16L^4}{320U} && (26)
\end{aligned}$$

Analogous to the mean case, define an "α-quantile unbiased" forecast as one which satisfies:

$$\mathbb{E}\left[\mathbf{1}\left\{Y \le \hat{Y}\right\} | \hat{Y}\right] = \alpha \tag{27}$$

Note that for an α-quantile unbiased forecast we have:

$$\begin{aligned}
\mathbb{E}\left[L\left(Y, \hat{Y}; g\right)\right] &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \le \hat{Y}\right\} - \alpha\right)\left(g\left(\hat{Y}\right) - g(Y)\right)\right] && (28) \\
&= \mathbb{E}\left[g\left(\hat{Y}\right)\left(\mathbf{1}\left\{Y \le \hat{Y}\right\} - \alpha\right)\right] - \mathbb{E}\left[\left(\mathbf{1}\left\{Y \le \hat{Y}\right\} - \alpha\right) g(Y)\right] \\
&= \mathbb{E}\left[g\left(\hat{Y}\right)\left(\mathbb{E}\left[\mathbf{1}\left\{Y \le \hat{Y}\right\} | \hat{Y}\right] - \alpha\right)\right] - \mathbb{E}\left[\mathbf{1}\left\{Y \le \hat{Y}\right\} g(Y)\right] + \alpha \mathbb{E}[g(Y)] \\
&= \alpha \mathbb{E}[g(Y)] - \mathbb{E}\left[\mathbf{1}\left\{Y \le \hat{Y}\right\} g(Y)\right]
\end{aligned}$$

Holzmann and Eulert (2014) present a different proof of part (a) of Proposition 4 below. I present the following for comparability with the conditional mean case presented in Proposition 1.

Proof of Proposition 4. **(a)** We will show that under Assumptions (1)–(3), $LinLin_B^{\alpha} \geq LinLin_A^{\alpha} \Rightarrow \mathcal{F}_t^B \subseteq \mathcal{F}_t^A \ \forall t \Rightarrow \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^B\right)\right] \geq \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^A\right)\right] \ \forall L \in \mathcal{L}_{GPL}^{\alpha}$, where $LinLin_j^{\alpha} \equiv \mathbb{E}\left[LinLin\left(Y_t, \hat{Y}_t^j\right)\right]$ for $j \in \{A, B\}$ and $LinLin$ is the "Lin-Lin" loss function in equation (27).

First: we are given that $LinLin_B^{\alpha} \geq LinLin_A^{\alpha}$, and assume that $\mathcal{F}_t^A \subseteq \mathcal{F}_t^B \ \forall t$. This implies $\mathbb{E}\left[LinLin\left(Y_t, \hat{Y}_t^A\right)|\mathcal{F}_t^B\right] \geq \mathbb{E}\left[LinLin\left(Y_t, \hat{Y}_t^B\right)|\mathcal{F}_t^B\right]$ $a.s.$ $\forall t$, since $\hat{Y}_t^A \in \mathcal{F}_t^A \subseteq \mathcal{F}_t^B$, and $\mathbb{E}\left[LinLin\left(Y_t, \hat{Y}_t^A\right)\right] \geq \mathbb{E}\left[LinLin\left(Y_t, \hat{Y}_t^B\right)\right]$ by the LIE. The only way that this also satisfy the assumption that $LinLin_B^{\alpha} \geq LinLin_A^{\alpha}$ is if $\mathbb{E}\left[LinLin\left(Y_t, \hat{Y}_t^A\right)|\mathcal{F}_t^B\right] = \mathbb{E}\left[LinLin\left(Y_t, \hat{Y}_t^B\right)|\mathcal{F}_t^B\right]$ $a.s.$ $\forall t$. Let $\widehat{\mathcal{Y}_t^i} = \left\{\hat{y} : \alpha = F_t^i\left(\hat{Y}_t^i\right)\right\}$, for $i \in \{A, B\}$. This accommodates the fact that we do not assume that F_t^i, for $i \in \{A, B\}$, is strictly increasing, and so the α-quantile is not necessarily unique. The necessity and sufficiency of GPL loss (which includes LinLin loss) for quantile estimation, implies that this set can alternatively be defined as $\widehat{\mathcal{Y}_t^i} = \arg\min_{\hat{y} \in \widehat{\mathcal{Y}}} \mathbb{E}\left[LinLin\left(Y_t, \hat{y}\right)|\mathcal{F}_t^i\right]$. Thus $\mathbb{E}\left[LinLin\left(Y_t, \hat{Y}_t^A\right)|\mathcal{F}_t^B\right] = \mathbb{E}\left[LinLin\left(Y_t, \hat{Y}_t^B\right)|\mathcal{F}_t^B\right]$ $a.s.$ $\forall t$ implies that $\hat{Y}_t^A \in \widehat{\mathcal{Y}_t^B} \ \forall t$ and so $\widehat{\mathcal{Y}_t^A} \cap \widehat{\mathcal{Y}_t^B} \neq \varnothing \ \forall t$. This violates Assumption 1, leading to a contradiction. Thus $LinLin_B^{\alpha} \geq LinLin_A^{\alpha} \Rightarrow \mathcal{F}_t^B \subseteq \mathcal{F}_t^A \ \forall t$. Next: Let

$$\bar{L}^j \equiv \mathbb{E}\left[L_{GPL}\left(Y_t, \hat{Y}_t^j; \alpha, g\right)\right], \quad j \in \{A, B\}$$

where $L_{GPL}(\cdot, \cdot; \alpha, g)$ is a GPL loss function defined by g, a nondecreasing function. Under Assumptions (2)-(3) we know that $\hat{Y}_t^j$ is the solution to $\min_{\hat{y}} \mathbb{E}\left[L_{GPL}\left(Y_t, \hat{Y}_t^j; \alpha, g\right)|\mathcal{F}_t^j\right]$. It is straightforward to show that $\hat{Y}_t^j$ then satisfies $\alpha = \mathbb{E}\left[\mathbf{1}\left\{Y_t \leq \hat{Y}_t^j\right\}|\mathcal{F}_t^j\right]$. This holds for all possible (conditional) distributions of Y_t, and from Saerens (2000) and Gneiting (2011b) we know that this implies (by the necessity of GPL loss for optimal quantile forecasts) that $\hat{Y}_t^j$ then moreover satisfies

$$\hat{Y}_t^j = \arg\min_{\hat{y}} \mathbb{E}\left[\left(\mathbf{1}\left\{Y_t \leq \hat{y}\right\} - \alpha\right)\left(g\left(\hat{y}\right) - g\left(Y_t\right)\right)|\mathcal{F}_t^j\right]$$

for any nondecreasing function g. If $\mathcal{F}_t^B \subseteq \mathcal{F}_t^A \ \forall t$ then by the LIE we have $\bar{L}^B(g) \geq \bar{L}^A(g)$ for any nondecreasing function g.

(b)(i) We first consider the case of non-nested information sets (violating Assumption 1). Consider the following simple example:

$$\begin{aligned} Y &= X + Z \qquad (29) \\ \text{where } X &\sim Unif(0, 10), \quad Z \sim Tri(0, 12), \quad X \perp\!\!\!\perp Z \end{aligned}$$

Let $\alpha = \frac{1}{2}$, and assume that forecast A conditions on X and forecast B conditions on Z. Then:

$$\hat{Y}^a = X + Median[Z] = X + 0.45, \text{ since } Median[Z] = 12 - 6\sqrt{2} \approx 3.51 \tag{30}$$

$$\hat{Y}^b = Z + Median[X] = Z + 2.5, \text{ since } Median[X] = 5 \tag{31}$$

Next consider the GPL loss functions generated by $g_1(y) = y$ and $g_2(y) = y^3$. Notice that both $\hat{Y}^a$ and $\hat{Y}^b$ are median-unbiased forecasts, which simplifies the calculation of their expected loss.

$$\begin{aligned}
\bar{L}_A(g_1) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \le \hat{Y}^a\right\} - 1/2\right)\left(\hat{Y}^a - Y\right)\right] \qquad (32)\\
&= \frac{1}{2}\mathbb{E}[Y] - \mathbb{E}\left[\mathbf{1}\left\{Y \le \hat{Y}^a\right\}Y\right]\\
\text{where } \mathbb{E}[Y] &= \mathbb{E}[X] + \mathbb{E}[Z] \qquad (33)\\
\text{and } \mathbb{E}\left[\mathbf{1}\left\{Y \le \hat{Y}^a\right\}Y\right] &= \mathbb{E}\left[\mathbf{1}\{X + Z \le X + M_z\}(X + Z)\right] \qquad (34)\\
&= \mathbb{E}\left[\mathbf{1}\{Z \le M_z\}\right]\mathbb{E}[X] + \mathbb{E}\left[\mathbf{1}\{Z \le M_z\}Z\right], \text{ since } X \perp\!\!\!\perp Z\\
&= \frac{1}{2}\mathbb{E}[X] + \mathbb{E}\left[\mathbf{1}\{Z \le M_z\}Z\right], \text{ since } \mathbb{E}\left[\mathbf{1}\{Z \le M_z\}\right] = 1/2
\end{aligned}$$

We find an analogous expression for the other forecaster:

$$\begin{aligned}
\bar{L}_B(g_1) &= \frac{1}{2}\mathbb{E}[Y] - \mathbb{E}\left[\mathbf{1}\left\{Y \le \hat{Y}^b\right\}Y\right] \qquad (35)\\
&= \frac{1}{2}\mathbb{E}[Y] - \left(\frac{1}{2}\mathbb{E}[Z] + \mathbb{E}\left[\mathbf{1}\{X \le M_x\}X\right]\right)
\end{aligned}$$

Next consider the loss GPL function obtained when $g_2(y) = y^3$.

$$\begin{aligned}
\bar{L}_A(g_2) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \le \hat{Y}^a\right\} - 1/2\right)\left(\left(\hat{Y}^a\right)^3 - Y^3\right)\right] \qquad (36)\\
&= \frac{1}{2}\mathbb{E}\left[Y^3\right] - \mathbb{E}\left[\mathbf{1}\left\{Y \le \hat{Y}^a\right\}Y^3\right]\\
\mathbb{E}\left[Y^3\right] &= \mathbb{E}\left[(X+Z)^3\right] = \mathbb{E}\left[X^3\right] + 3\mathbb{E}\left[X^2\right]\mathbb{E}[Z] + 3\mathbb{E}[X]\mathbb{E}\left[Z^2\right] + \mathbb{E}\left[Z^3\right] \qquad (37)\\
\mathbb{E}\left[\mathbf{1}\left\{Y \le \hat{Y}^a\right\}Y^3\right] &= \mathbb{E}\left[\mathbf{1}\{Z \le M_z\}\right]\mathbb{E}\left[X^3\right] + 3\mathbb{E}\left[\mathbf{1}\{Z \le M_z\}Z\right]\mathbb{E}\left[X^2\right] \qquad (38)\\
&\quad + 3\mathbb{E}\left[\mathbf{1}\{Z \le M_z\}Z^2\right]\mathbb{E}[X] + \mathbb{E}\left[\mathbf{1}\{Z \le M_z\}Z^3\right]
\end{aligned}$$

Pulling these terms together and using the expressions for these moments given above, we find:

$$\bar{L}_A(g_1) = 1.17 < 1.25 = \bar{L}_B(g_1) \tag{39}$$

$$\bar{L}_A(g_2) = 350.45 > 349.38 = \bar{L}_B(g_2) \tag{40}$$

Thus the ranking is reversed depending on the choice of function g. Note that while the differences in these values may appear small, these are analytical population values, and so there is no sampling or simulation variability.

(ii) Next we consider the case that both forecasters use correctly specified models, given their (nested) information sets, but they are subject to estimation error. Assume that

$$\begin{aligned} Y &= X + Z \\ X &\sim Unif\left(-10,0\right), \;\; Z \sim Unif\left(0,12\right), \;\; X \perp\!\!\!\perp Z \end{aligned} \tag{41}$$

Assume that forecaster A uses no conditioning information, and so reports her optimal forecast as:

$$\hat{Y}^a = Median\left[Y\right] = 1 \tag{42}$$

Forecaster B uses information on Z, but to exploit it must estimate $Median\left[X\right]$. He treats that as an unknown parameter and assume that he estimates it using $n = 1$ observation of X. Forecaster B's prediction will then be

$$\hat{Y}^b = \tilde{X} + Z \tag{43}$$

where $\tilde{X}$ is a realization from a $Unif\left(L,0\right)$ distribution, independent of $\left(X,Z\right)$. This design allows for a signal/noise trade-off. In this design we find that:

$$\begin{aligned} \bar{L}_A\left(g_1\right) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} - 1/2\right)\left(\hat{Y}^a - Y\right)\right] \\ &= \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq M_y\right\} - 1/2\right)\left(M_y - Y\right)\right] \\ &= M_y \mathbb{E}\left[\mathbf{1}\left\{Y \leq M_y\right\}\right] - \mathbb{E}\left[\mathbf{1}\left\{Y \leq M_y\right\} Y\right] - 1/2\left(M_y - \mathbb{E}\left[Y\right]\right) \end{aligned} \tag{44}$$

For forecaster B we find:

$$\begin{aligned} \bar{L}_B\left(g_1\right) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^b\right\} - 1/2\right)\left(\hat{Y}^b - Y\right)\right] \\ &= \mathbb{E}\left[\left(\mathbf{1}\left\{X + Z \leq \tilde{X} + Z\right\} - 1/2\right)\left(\tilde{X} + Z - X - Z\right)\right] \\ &= \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\left(\tilde{X} - X\right)\right] - 1/2\left(\mathbb{E}\left[\tilde{X} + Z - Y\right]\right), \text{ note } \mathbb{E}\left[\tilde{X} + Z - Y\right] = 0 \\ &= \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\tilde{X}\right] - \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}X\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}|\tilde{X}\right]\tilde{X}\right] - \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}|X\right]X\right] \\ &= \mathbb{E}\left[F_x\left(\tilde{X}\right)\tilde{X}\right] - \mathbb{E}\left[\left(1 - F_x\left(X\right)\right)X\right] \\ &= 2\mathbb{E}\left[F_x\left(X\right)X\right] - \mathbb{E}\left[X\right], \text{ since } \tilde{X} \overset{d}{=} X \end{aligned} \tag{45}$$

And for the second loss function we obtain:

$$\begin{aligned} \bar{L}_A(g_2) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} - 1/2\right)\left(\left(\hat{Y}^a\right)^3 - Y^3\right)\right] \\ &= \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq M_y\right\} - 1/2\right)\left(M_y^3 - Y^3\right)\right] \\ &= M_y^3 \mathbb{E}\left[\mathbf{1}\left\{Y \leq M_y\right\}\right] - \mathbb{E}\left[\mathbf{1}\left\{Y \leq M_y\right\} Y^3\right] - 1/2\left(M_y^3 - \mathbb{E}\left[Y^3\right]\right) \end{aligned} \tag{46}$$

and

$$\begin{aligned} \bar{L}_B(g_2) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^b\right\} - 1/2\right)\left(\left(\hat{Y}^b\right)^3 - Y^3\right)\right] \\ &= \mathbb{E}\left[\left(\mathbf{1}\left\{X + Z \leq \tilde{X} + Z\right\} - 1/2\right)\left(\left(\tilde{X} + Z\right)^3 - (X+Z)^3\right)\right] \\ &= \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\left(\left(\tilde{X} + Z\right)^3 - (X+Z)^3\right)\right] - 1/2\left(\mathbb{E}\left[\left(\tilde{X} + Z\right)^3\right] - \mathbb{E}\left[(X+Z)^3\right]\right) \\ &= \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\left(\tilde{X}^3 + 3\tilde{X}^2 Z + 3\tilde{X} Z^2 - X^3 - 3X^2 Z - 3XZ^2\right)\right] \\ &= \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\tilde{X}^3\right] - \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\} X^3\right] \\ &\quad + 3\mathbb{E}[Z]\left(\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\tilde{X}^2\right] - \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\} X^2\right]\right) \\ &\quad + 3\mathbb{E}\left[Z^2\right]\left(\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\tilde{X}\right] - \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\} X\right]\right) \end{aligned} \tag{47}$$

Then we use, for $p = 1, 2, 3$:

$$\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\tilde{X}^p\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}|\tilde{X}\right]\tilde{X}^p\right] = \mathbb{E}\left[F_x(X) X^p\right], \text{ since } \tilde{X} \overset{d}{=} X \tag{48}$$

$$\begin{aligned} \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\} X^p\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}|X\right] X^p\right] = \mathbb{E}\left[\left(1 - F_x(X)\right) X^p\right] \\ &= \mathbb{E}\left[X^p\right] - \mathbb{E}\left[F_x(X) X^p\right] \end{aligned} \tag{49}$$

And so

$$\begin{aligned} \bar{L}_B(g_2) &= 2\mathbb{E}\left[F_x(X) X^3\right] - \mathbb{E}\left[X^3\right] \\ &\quad + 3\mathbb{E}[Z]\left(2\mathbb{E}\left[F_x(X) X^2\right] - \mathbb{E}\left[X^2\right]\right) \\ &\quad + 3\mathbb{E}\left[Z^2\right]\left(2\mathbb{E}\left[F_x(X) X\right] - \mathbb{E}[X]\right) \end{aligned} \tag{50}$$

For $X \sim Unif(L, U)$ we have:

$$\mathbb{E}\left[F_x(X) X\right] = \frac{L + 2U}{6} \tag{51}$$

$$\mathbb{E}\left[F_x(X) X^2\right] = \frac{L^2 + 2LU + 3U^2}{12} \tag{52}$$

$$\mathbb{E}\left[F_x(X) X^3\right] = \frac{L^3 + 2L^2 U + 3LU^2 + 4U^3}{20} \tag{53}$$

Pulling these terms together, we find that

$$\bar{L}_A(g_1) = 1.85 > 1.67 = \bar{L}_B(g_1) \tag{54}$$

$$\bar{L}_A(g_2) = 72.65 < 90 = \bar{L}_B(g_2) \tag{55}$$

Thus the ranking is reversed depending on the choice of function g.

(iii) Finally, we consider a violation assumption 3, and consider models that are misspecified. We will simplify the DGP, and assume that

$$Y = X \sim Unif\,(0, 10) \tag{56}$$

We will assume that the two forecasters use misspecified models, in that they use a linear model with parameters that differ from $(0, 1)$:

$$\hat{Y}^a = \beta_0 + \beta_1 X \tag{57}$$

$$\hat{Y}^b = \gamma_0 + \gamma_1 X \tag{58}$$

Of course here we cannot use the simplifcation that holds when the forecasts are median unbiased. In this example, if we set $(\beta_0, \beta_1) = (0.33, 0.67)$ and $(\gamma_0, \gamma_1) = (-0.25, 1.25)$ then both forecasts use the same information set, neither has estimation error, but both are based on misspecified models. In this case we find:

$$\begin{aligned}
\bar{L}_A(g_1) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \le \hat{Y}^a\right\} - 1/2\right)\left(\hat{Y}^a - Y\right)\right] \\
&= \mathbb{E}\left[\left(\mathbf{1}\left\{X \le \beta_0 + \beta_1 X\right\} - 1/2\right)\left(\beta_0 + \beta_1 X - X\right)\right] \\
&= \beta_0 \mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \le \beta_0\right\}\right] - \frac{\beta_0}{2} - \frac{\beta_1 - 1}{2}\mathbb{E}[X] \\
&\quad + (\beta_1 - 1)\,\mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \le \beta_0\right\} X\right]
\end{aligned} \tag{59}$$

$$\mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \le \beta_0\right\}\right] = \begin{cases} F_x\left(\frac{\beta_0}{1-\beta_1}\right), & \beta_1 < 1 \\ 1 - F_x\left(\frac{\beta_0}{1-\beta_1}\right), & \beta_1 > 1 \\ \mathbf{1}\left\{\beta_0 \ge 0\right\}, & \beta_1 = 1 \end{cases} \tag{60}$$

$$\mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \le \beta_0\right\} X^p\right] = \begin{cases} \mathbb{E}\left[\mathbf{1}\left\{X \le \frac{\beta_0}{1-\beta_1}\right\} X^p\right], & \beta_1 < 1 \\ \mathbb{E}[X^p] - \mathbb{E}\left[\mathbf{1}\left\{X \le \frac{\beta_0}{1-\beta_1}\right\} X^p\right], & \beta_1 > 1 \\ \mathbf{1}\left\{\beta_0 \ge 0\right\}\mathbb{E}[X^p], & \beta_1 = 1 \end{cases} \tag{61}$$

The same expressions can be used for $\bar{L}_B(g_1)$ plugging in (γ_0, γ_1) for (β_0, β_1). We use $p = 1$ for the first GPL loss function above, and $p = 1, 2, 3$ for the second, below.

Next consider

$$\begin{aligned}
\bar{L}_A(g_2) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} - 1/2\right)\left(\left(\hat{Y}^a\right)^3 - Y^3\right)\right] \\
&= \mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1)X \leq \beta_0\right\}\left((\beta_0+\beta_1 X)^3 - X^3\right)\right] - 1/2\mathbb{E}\left[(\beta_0+\beta_1 X)^3 - X^3\right] \\
&= \beta_0^3\mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1)X \leq \beta_0\right\}\right] + 3\beta_0^2\beta_1\mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1)X \leq \beta_0\right\}X\right] \\
&\quad +3\beta_0\beta_1^2\mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1)X \leq \beta_0\right\}X^2\right] + \left(\beta_1^3-1\right)\mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1)X \leq \beta_0\right\}X^3\right] \\
&\quad -1/2\left(\beta_0^3 + 3\beta_0^2\beta_1\mathbb{E}[X] + 3\beta_0\beta_1^2\mathbb{E}\left[X^2\right] + \left(\beta_1^3-1\right)\mathbb{E}\left[X^3\right]\right)
\end{aligned} \tag{62}$$

The same expressions can be used for $\bar{L}_B(g_2)$ plugging in (γ_0, γ_1) for (β_0, β_1). Pulling these terms together, we find that

$$\bar{L}_A(g_1) = 0.68 > 0.51 = \bar{L}_B(g_1) \tag{63}$$

$$\bar{L}_A(g_2) = 79.44 < 100.19 = \bar{L}_B(g_2) \tag{64}$$

Thus the ranking is reversed depending on the choice of function g.

We have thus demonstrated analytically that the presence of *any* of non-nested information sets, estimation error, or model misspecification can lead to sensitivity in the ranking of two quantile forecasts to the choice of consistent (GPL) loss function. ■

Proof of Proposition 6. **(a)** We again prove this result by showing that $\mathbb{E}\left[L\left(F_t^A, Y_t\right)\right] \leq \mathbb{E}\left[L\left(F_t^B, Y_t\right)\right]$ for some $L \in \mathcal{L}_{\text{Proper}} \Rightarrow \mathcal{F}_t^B \subseteq \mathcal{F}_t^A\ \forall t \Rightarrow \mathbb{E}\left[L\left(F_t^A, Y_t\right)\right] \leq \mathbb{E}\left[L\left(F_t^B, Y_t\right)\right]\ \forall\ L \in \mathcal{L}_{\text{Proper}}$. First: we are given that $\mathbb{E}\left[L\left(F_t^A, Y_t\right)\right] \leq \mathbb{E}\left[L\left(F_t^B, Y_t\right)\right]$, and assume that $\mathcal{F}_t^A \subseteq \mathcal{F}_t^B\ \forall t$. Under Assumptions (2)-(3), this implies that we can take F_t^B as the data generating process for Y_t. Then $\mathbb{E}\left[L\left(F_t^B, Y_t\right)|\mathcal{F}_t^B\right] = \mathbb{E}_{F_t^B}\left[L\left(F_t^B, Y_t\right)|\mathcal{F}_t^B\right] \leq \mathbb{E}_{F_t^B}\left[L\left(F_t^A, Y_t\right)|\mathcal{F}_t^B\right]\ \forall t$ the propriety of L. By the LIE this implies $\mathbb{E}\left[L\left(F_t^B, Y_t\right)\right] \leq \mathbb{E}\left[L\left(F_t^A, Y_t\right)\right]$, which can only hold if $\mathbb{E}\left[L\left(F_t^B, Y_t\right)|\mathcal{F}_t^B\right] = \mathbb{E}\left[L\left(F_t^A, Y_t\right)|\mathcal{F}_t^B\right]$ *a.s.* $\forall t$, but since L is a strictly proper scoring rule this implies $F_t^A = F_t^B$ *a.s.*$\forall t$ which violates Assumption 1, leading to a contradiction. Thus $\mathbb{E}\left[L\left(F_t^A, Y_t\right)\right] \leq \mathbb{E}\left[L\left(F_t^B, Y_t\right)\right]$ for some $L \in \mathcal{L}_{\text{Proper}} \Rightarrow \mathcal{F}_t^B \subseteq \mathcal{F}_t^A\ \forall t$. Next, using similar logic to above, given $\mathcal{F}_t^B \subseteq \mathcal{F}_t^A$ we have that $\mathbb{E}\left[L\left(F_t^A, Y_t\right)\right] \leq \mathbb{E}\left[L\left(F_t^B, Y_t\right)\right]$ for any $L \in \mathcal{L}_{\text{Proper}}$, completing the proof.

(b)(i) We first consider the case of non-nested information sets (violating Assumption 1). Consider the following example:

$$\begin{aligned} Y &= -\beta_2 A - \beta_1 (1-A) + \beta_1 B + \beta_2 (1-B) \\ A &\sim Bernoulli\,(p) \\ B &\sim Bernoulli\,(q)\,, \quad B \perp\!\!\!\perp A \\ \beta_2 &> \beta_1 > 0 \end{aligned} \tag{65}$$

The indicator, A reveals whether the left "tail" will be long or short, and B reveals whether the right tail will be long or short. Forecaster A observes the signal A and forecaster B observes signal B, i.e., each forecaster only gets information about a single tail (left or right). Then we find:

$$\begin{aligned} \mathbb{E}\left[wCRPS\left(F_A, Y, \omega\right)\right] &= pq\,(1-q) \int_{\beta_1-\beta_2}^{0} \omega\,(z)\,dz + q\,(1-p)\,(1-q) \int_{0}^{\beta_2-\beta_1} \omega\,(z)\,dz \\ \mathbb{E}\left[wCRPS\left(F_B, Y, \omega\right)\right] &= pq\,(1-p) \int_{\beta_1-\beta_2}^{0} \omega\,(z)\,dz + p\,(1-p)\,(1-q) \int_{0}^{\beta_2-\beta_1} \omega\,(z)\,dz \end{aligned} \tag{66}$$

The two proper scoring rules we consider (equation 33) place different weights on the left vs. right tails using the logistic function:

$$\omega\,(z; a) = \frac{1}{1+\exp\{-az\}} \tag{67}$$

When $a > 0$ more weight is placed on the right tail, and when $a < 0$ more weight is placed on the left tail. We then compute the integrals, setting $\omega_R\,(z) = \omega\,(z; +1)$ and $\omega_L\,(z) = \omega\,(z; -1)$

$$\begin{aligned} \int_{\beta_1-\beta_2}^{0} \omega_R\,(z)\,dz &= \int_{0}^{\beta_2-\beta_1} \omega_L\,(z)\,dz = \beta_2 + \log 2 - \log\left(\exp\{\beta_2\} + \exp\{\beta_1\}\right) \\ \int_{0}^{\beta_2-\beta_1} \omega_R\,(z)\,dz &= \int_{\beta_1-\beta_2}^{0} \omega_L\,(z)\,dz = \log\left(\frac{1}{2}\left(1+\exp\{\beta_2-\beta_1\}\right)\right) \end{aligned} \tag{68}$$

With these in hand, if we set $(p, q, \beta_1, \beta_2) = (0.25, 0.75, 1, 5)$ we find:

$$\mathbb{E}\left[wCRPS\left(F_A, Y; \omega_R\right)\right] = 0.50 > 0.25 = \mathbb{E}\left[wCRPS\left(F_B, Y; \omega_R\right)\right] \tag{69}$$

$$\mathbb{E}\left[wCRPS\left(F_A, Y; \omega_L\right)\right] = 0.25 < 0.50 = \mathbb{E}\left[wCRPS\left(F_B, Y; \omega_L\right)\right] \tag{70}$$

And so the ranking of these two distribution forecasts can be reversed depending on the choice of (proper) scoring rule.

(ii) Next, we consider a violation assumption 3, and consider models that are misspecified. In this case, consider the case where forecaster A uses the unconditional distribution of the target variable, while forecaster B continues to use her signal, but based on $\tilde{p} \neq p$. If we set $(p, q, \beta_1, \beta_2, \tilde{p}) = (0.25, 0.75, 1, 5, 0.5)$ we find

$$\mathbb{E}\left[wCRPS\left(\bar{F}_A, Y; \omega_R\right)\right] = 0.61 > 0.33 = \mathbb{E}\left[wCRPS\left(\tilde{F}_B, Y; \omega_R\right)\right] \tag{71}$$

$$\mathbb{E}\left[wCRPS\left(\bar{F}_A, Y; \omega_L\right)\right] = 0.61 < 0.67 = \mathbb{E}\left[wCRPS\left(\tilde{F}_B, Y; \omega_L\right)\right] \tag{72}$$

And so the ranking of these two distribution forecasts can be reversed depending on the choice of (proper) scoring rule. (Note that $\mathbb{E}\left[wCRPS\left(\bar{F}_A, Y; \omega_R\right)\right] = \mathbb{E}\left[wCRPS\left(\bar{F}_A, Y; \omega_L\right)\right]$ as the distribution forecast $\left(\bar{F}_A\right)$ is symmetric around zero, and the weighting functions satisfy $\omega_R(z) = \omega_L(-z)$.)

(iii) Finally, we consider the case that both forecasters use correctly specified models, given their (nested) information sets, but are subject to estimation error. Consider the case that forecaster A again uses the unconditional distribution of the target variable, while forecaster B uses her signal, but to do so must estimate the parameter p. Assume she does so based on n observations of the signal A. (Note that since forecaster B observes the signal B, the value for A can be backed out, *ex post*, from the realized value of the target variable.) Then

$$n\hat{p} = \sum_{i=1}^{n} A_i \sim Binomial\left(n, p\right) \tag{73}$$

In this case, we have:

$$\mathbb{E}\left[wCRPS\left(\hat{F}_B\left(\hat{p}\right), Y; \omega\right)\right] = \sum_{\tilde{p}} \mathbb{E}\left[wCRPS\left(\tilde{F}_B\left(\tilde{p}\right), Y; \omega\right)\right] \Pr\left[\hat{p} = \tilde{p}\right] \tag{74}$$

and we can use the expressions from part (ii) to help solve this problem. Consider the case that $n = 4$, and so $\hat{p}$ can take one of five values $\{0, 1/4, 1/2, 3/4, 1\}$. In this case we find

$$\mathbb{E}\left[wCRPS\left(\bar{F}_A, Y; \omega_R\right)\right] = 0.61 > 0.31 = \mathbb{E}\left[wCRPS\left(\hat{F}_B\left(\hat{p}\right), Y; \omega_R\right)\right] \tag{75}$$

$$\mathbb{E}\left[wCRPS\left(\bar{F}_A, Y; \omega_L\right)\right] = 0.61 < 0.62 = \mathbb{E}\left[wCRPS\left(\hat{F}_B\left(\hat{p}\right), Y; \omega_L\right)\right] \tag{76}$$

And so the ranking of these two distribution forecasts can be reversed depending on the choice of (proper) scoring rule. ■

Appendix SA.2: Derivations

The results below draw on the following lemma, which summarizes some useful results on moments that arise when the data are Gaussian and the loss function is exponential Bregman.

Lemma 1 *If* $X \sim N\left(\mu, \sigma^2\right)$ *and* $(a,b) \in \mathbb{R}^2$, *then*

$$
\begin{array}{lrcl}
\textit{(i)} & \mathbb{E}\left[\exp\{a+bX\}\right] & = & \exp\left\{a+b\mu+\frac{1}{2}b^2\sigma^2\right\} \\
\textit{(ii)} & \mathbb{E}\left[\exp\{a+bX\}X\right] & = & \exp\left\{a+b\mu+\frac{1}{2}b^2\sigma^2\right\}\left(\mu+b\sigma^2\right) \\
\textit{(iii)} & \mathbb{E}\left[\exp\{a+bX\}X^2\right] & = & \exp\left\{a+b\mu+\frac{1}{2}b^2\sigma^2\right\}\left(\sigma^2+\left(\mu+b\sigma^2\right)^2\right) \\
\textit{(iv)} & \mathbb{E}\left[\exp\{a+bX\}X^3\right] & = & \exp\left\{a+b\mu+\frac{1}{2}b^2\sigma^2\right\}\left(\mu+b\sigma^2\right)\left(3\sigma^2+\left(\mu+b\sigma^2\right)^2\right)
\end{array}
$$

Some results below are simplified if we consider the following definition:

Definition 1 *A forecast* $\hat{Y}_t^i$ *is "mean unbiased" if* $\hat{Y}_t^i = \mathbb{E}\left[Y_t|\mathcal{F}_t^i\right]$ *a.s.*

By the law of iterated expectations, this implies that $\mathbb{E}\left[Y_t|\hat{Y}_t^i\right] = \hat{Y}_t^i$ *a.s.* Note that this does *not* require that $\mathcal{F}_t^i$ contains all relevant information for forecasting Y_t, only that that $\hat{Y}_t^i$ optimally uses all information available in $\mathcal{F}_t^i$.

Appendix SA.2.1: Derivations for the AR(p) models in Section 2.2

The Gaussian AR(5) specification in equation (12) implies:

$$\left[\begin{array}{ccccc} Y_t & Y_{t-1} & Y_{t-2} & Y_{t-3} & Y_{t-4} \end{array}\right]' \sim N\left(\mu\boldsymbol{\iota}_5, \Sigma\right) \tag{77}$$

where $\boldsymbol{\iota}_5$ is a (5×1) vector of ones, μ is the mean of Y_t and Σ is the covariance matrix of the left-hand side vector. These can be obtained using standard methods from time series analysis, see Hamilton (1994) for example).

$$
\begin{array}{lrcl}
\text{Let} & \boldsymbol{\phi} & \equiv & [\phi_1, \phi_2, ..., \phi_5]' \qquad (78) \\
\text{then} & \mu & = & \dfrac{\phi_0}{1-\boldsymbol{\phi}'\boldsymbol{\iota}_5}, \quad \text{and} \quad vec\left(\Sigma\right) = \left(I_{25} - \left(F\otimes F\right)\right)^{-1} vec\left(Q\right) \\
\text{where} & F & = & \left[\begin{array}{c|c} \multicolumn{2}{c}{\boldsymbol{\phi}'} \\ \hline I_4 & \mathbf{0}_{4\times1} \end{array}\right], \quad Q = \mathbf{e}_1\mathbf{e}_1', \quad \mathbf{e}_1 \equiv [1,0,0,0,0]'
\end{array}
$$

Then note that the joint distribution of (Y_t, Y_{t-1}) is

$$\begin{aligned} [Y_t, Y_{t-1}]' &\sim N(\mu\boldsymbol{\iota}_2, \Sigma_2) \\ \text{where } \Sigma_2 &\equiv \begin{bmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{bmatrix}, \text{ and } \gamma_j \equiv Cov[Y_t, Y_{t-j}], \text{ for } j = 0, 1, 2, ... \end{aligned} \tag{79}$$

Denote $\rho_j \equiv \gamma_j/\gamma_0$. Then the conditional distribution of $Y_t|Y_{t-1}$ is:

$$Y_t|Y_{t-1} \sim N\left(\mu(1-\rho_1) + \rho_1 Y_{t-1}, \gamma_0\left(1-\rho_1^2\right)\right) \tag{80}$$

and so for the parameters used in the example we find $[\beta_0, \beta_1] = [\mu(1-\rho_1), \rho_1] = [1.52, 0.85]$.

Similar calculations for the AR(2) model yield:

$$[Y_t, Y_{t-1}, Y_{t-2}]' \sim N(\mu\boldsymbol{\iota}_3, \Sigma_3) \tag{81}$$

$$\text{where } \Sigma_3 \equiv \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix} \tag{82}$$

$$\begin{aligned} Y_t|(Y_{t-1}, Y_{t-2}) &\sim N(\delta_0 + \delta_1 Y_{t-1} + \delta_2 Y_{t-2}, V_{AR2}) \\ \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} &= \frac{1}{\gamma_0^2 - \gamma_1^2}\begin{bmatrix} \gamma_1(\gamma_0 - \gamma_2) \\ \gamma_0\gamma_2 - \gamma_1^2 \end{bmatrix}, \quad \delta_0 = \mu(1 - \delta_1 - \delta_2) \end{aligned}$$

$$V_{AR2} = \gamma_0 - \begin{bmatrix} \gamma_1 & \gamma_2 \end{bmatrix}\begin{bmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{bmatrix}^{-1}\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \gamma_0\frac{\rho_1^2 + \rho_2^2 - 2\rho_1^2\rho_2}{1 - \rho_1^2} \tag{83}$$

And for the parameters used in the example we find $[\delta_0, \delta_1, \delta_2] = [1.38, 0.76, 0.10]$.

Now we derive the expected loss for the AR(1), AR(2) and AR(5) forecasts. First note that since all three of these forecasts are mean-unbiased, the expected loss of any Bregman loss function simplifies to:

$$\mathbb{E}\left[L\left(Y_t, \hat{Y}_t; \phi\right)\right] = \mathbb{E}[\phi(Y_t)] - \mathbb{E}\left[\phi\left(\hat{Y}_t\right)\right] - \mathbb{E}\left[\phi'\left(\hat{Y}_t\right)\left(\mathbb{E}\left[Y_t|\hat{Y}_t\right] - \hat{Y}_t\right)\right] = \mathbb{E}[\phi(Y_t)] - \mathbb{E}\left[\phi\left(\hat{Y}_t\right)\right]. \tag{84}$$

For exponential Bregman loss, where $\phi(Y; a) = 2a^{-2}\exp\{aY\}$, Lemma 1 implies

$$\mathbb{E}[\phi(Y_t)] = 2a^{-2}\mathbb{E}[\exp\{aY_t\}] = 2a^{-2}\exp\left\{\frac{a}{2}(2\mu + a\gamma_0)\right\} \tag{85}$$

To obtain $\mathbb{E}\left[\exp\left\{a\hat{Y}_t\right\}\right]$ for the AR(1), AR(2) and AR(5) forecasts, we exploit the fact that for this Gaussian autoregression, all of these forecasts are unconditionally normally distributed:

$\hat{Y}_t^{\text{ARk}} \sim N(\mu, V_{\text{ARk}})$, where $[V_{\text{AR1}}, V_{\text{AR2}}, V_{\text{AR5}}] = \left[\rho_1^2\gamma_0\ ,\ \delta_1^2\delta_0 + \delta_2^2\delta_0 + 2\delta_1^2\delta_2\ ,\ \gamma_0 - 1\right]$. Thus we find

$$\begin{aligned}\mathbb{E}\left[\phi\left(\hat{Y}_t^{\text{ARk}}\right)\right] &= 2a^{-2}\mathbb{E}\left[\exp\left\{a\hat{Y}_t^{\text{ARk}}\right\}\right] \\ &= 2a^{-2}\exp\left\{\frac{a}{2}\left(2\mu + aV_{\text{ARk}}\right)\right\}\end{aligned}$$

from Lemma 1. The expected loss from an $AR(k)$ forecast is

$$\begin{aligned}\mathbb{E}\left[L\left(Y_t, \hat{Y}_t^{\text{ARk}}; \phi\right)\right] &= 2a^{-2}\left(\exp\left\{\frac{a}{2}\left(2\mu + a\gamma_0\right)\right\} - \exp\left\{\frac{a}{2}\left(2\mu + aV_{\text{ARk}}\right)\right\}\right) \\ &\to \gamma_0 - V_{\text{ARk}} \text{ as } a \to 0\end{aligned}$$

Note that we know that $V_{\text{AR1}} \leq V_{\text{AR2}} \leq V_{\text{AR5}}$ and so we immediately see that the ranking under MSE (i.e., exponential Bregman with $a \to 0$) is $\mathbb{E}\left[L\left(Y_t, \hat{Y}_t^{\text{AR1}}\right)\right] \geq \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^{\text{AR2}}\right)\right] \geq \mathbb{E}\left[L\left(Y_t, \hat{Y}_t^{\text{AR5}}\right)\right]$.

Appendix SA.2.2: Derivations for the Bernoulli forecasters in Section 2.2

Since $\hat{Y}_t^X$ and $\hat{Y}_t^W$ are both optimal with respect to their limited information, they are both "mean unbiased" and so their expected Bregman loss simplifies to $2a^{-2}\left(E\left[\exp\{aY_t\} - \exp\left\{a\hat{Y}_t\right\}\right]\right)$. For this DGP we easily find:

$$\begin{aligned}\mathbb{E}\left[\exp\{aY_t\}\right] &= \exp\left\{\frac{a^2}{2} + a\left(\mu_L + \mu_C\right)\right\}pq + \exp\left\{\frac{a^2}{2} + a\left(\mu_H + \mu_C\right)\right\}(1-p)\,q \qquad (86)\\ &\quad + \exp\left\{\frac{a^2}{2} + a\left(\mu_L + \mu_M\right)\right\}p\,(1-q) + \exp\left\{\frac{a^2}{2} + a\left(\mu_H + \mu_M\right)\right\}(1-p)(1-q) \\ \mathbb{E}\left[\exp\left\{a\hat{Y}_t^X\right\}\right] &= \exp\left\{a\left(\mu_L + q\mu_C + (1-q)\,\mu_M\right)\right\}p + \exp\left\{a\left(\mu_H + q\mu_C + (1-q)\,\mu_M\right)\right\}(1-p) \\ \mathbb{E}\left[\exp\left\{a\hat{Y}_t^W\right\}\right] &= \exp\left\{a\left(p\mu_L + (1-p)\,\mu_H + \mu_C\right)\right\}q + \exp\left\{a\left(p\mu_L + (1-p)\,\mu_H + \mu_M\right)\right\}(1-q)\end{aligned}$$

Figure 2 normalizes the expected loss from forecast X and W by the optimal forecast, using both signals:

$$\hat{Y}_t^{XW} = X_t\mu_L + (1-X_t)\,\mu_H + W_t\mu_C + (1-W_t)\,\mu_M \qquad (87)$$

which leads to

$$\begin{aligned}\mathbb{E}\left[\exp\left\{a\hat{Y}_t^{XW}\right\}\right] &= \exp\left\{a\left(\mu_L + \mu_C\right)\right\}pq + \exp\left\{a\left(\mu_H + \mu_C\right)\right\}(1-p)\,q \qquad (88)\\ &\quad + \exp\left\{a\left(\mu_L + \mu_M\right)\right\}p\,(1-q) + \exp\left\{a\left(\mu_H + \mu_M\right)\right\}(1-p)(1-q)\end{aligned}$$

Appendix SA.2.3: Derivations for the linear model in Section 2.3

The first-order condition for the optimal parameter $\theta \equiv [\alpha, \beta]$ is:

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} \mathbb{E}\left[L\left(Y, m\left(X;\theta\right);\phi\right)\right] \\ &= \mathbb{E}\left[\phi''\left(m\left(X;\theta\right)\right)\left(\mathbb{E}\left[Y|X\right] - m\left(X;\theta\right)\right)\frac{\partial m\left(X;\theta\right)}{\partial \theta}\right] \\ &= 2\mathbb{E}\left[\exp\left\{a\left(\alpha+\beta X\right)\right\}\left(X^2-\alpha-\beta X\right)\left[1,X\right]'\right] \end{aligned} \tag{89}$$

So the two first-order conditions are:

$$0 = \mathbb{E}\left[\exp\left\{a\left(\alpha+\beta X\right)\right\}X^2\right] - \alpha\mathbb{E}\left[\exp\left\{a\left(\alpha+\beta X\right)\right\}\right] - \beta\mathbb{E}\left[\exp\left\{a\left(\alpha+\beta X\right)\right\}X\right] \tag{90}$$

$$0 = \mathbb{E}\left[\exp\left\{a\left(\alpha+\beta X\right)\right\}X^3\right] - \alpha\mathbb{E}\left[\exp\left\{a\left(\alpha+\beta X\right)\right\}X\right] - \beta\mathbb{E}\left[\exp\left\{a\left(\alpha+\beta X\right)\right\}X^2\right] \tag{91}$$

Using Lemma 1 above we have each of the four unique terms above in closed form. Substituting these in and solving for solving for $[\alpha, \beta]$ yields the expressions given in equation (25).