Supplemental Appendix to:

"Comparing Possibly Misspecified Forecasts"

Andrew J. Patton

Duke University

22 December 2016

The proof below uses 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 $U > L$.

$$X \sim Unif(L,U) \qquad Z \sim Tri(L,U) \tag{1}$$

$$F_x(x) = \begin{cases} 0, & z < L \\ \frac{x-L}{U-L}, & z \in [L,U] \\ 1, & z > U \end{cases} \qquad F_z(z) = \begin{cases} 0, & z < L \\ \frac{(U-L)^2-(U-z)^2}{(U-L)^2}, & z \in [L,U] \\ 1, & z > U \end{cases} \tag{2}$$

$$f_x(x) = \begin{cases} \frac{1}{U-L}, & z \in [L,U] \\ 0 & else \end{cases} \qquad f_z(z) = \begin{cases} \frac{2(U-x)}{(U-L)^2}, & z \in [L,U] \\ 0 & else \end{cases} \tag{3}$$

$$F_x^{-1}(\alpha) = L + \alpha(U-L), \text{ for } \alpha \in [0,1] \qquad F_z^{-1}(\alpha) = U - (U-L)\sqrt{1-\alpha}, \text{ for } \alpha \in [0,1] \tag{4}$$

$$\mathbb{E}[X] = \frac{1}{2}(U+L) \qquad \mathbb{E}[Z] = \frac{1}{3}(2L+U) \tag{5}$$

$$\mathbb{E}\left[X^2\right] = \frac{1}{3}\left(L^2+U^2+LU\right) \qquad \mathbb{E}\left[Z^2\right] = \frac{1}{6}\left(3L^2+2LU+U^2\right) \tag{6}$$

$$\mathbb{E}\left[X^3\right] = \frac{1}{4}\left(L^3+U^3+L^2U+LU^2\right) \qquad \mathbb{E}\left[Z^3\right] = \frac{1}{10}\left(4L^3+3L^2U+2LU^2+U^3\right) \tag{7}$$

$$M_x \equiv Median[X] = \frac{1}{2}(U+L) \qquad M_z \equiv Median[Z] = U - \frac{U-L}{\sqrt{2}} \tag{8}$$

$$\mathbb{E}[\mathbf{1}\{X<b\}X] = \frac{b^2-L^2}{2(U-L)}, \text{ for } b \in [L,U] \qquad \mathbb{E}[\mathbf{1}\{Z \le b\}Z] = \frac{3b^2U-2b^3-L^2(3U-2L)}{3(U-L)^2}, \text{ for } b \in [L,U] \tag{9}$$

$$\mathbb{E}\left[\mathbf{1}\{X<b\}X^2\right] = \frac{b^3-L^3}{3(U-L)} \qquad \mathbb{E}\left[\mathbf{1}\{Z \le b\}Z^2\right] = \frac{4b^3U-3b^4-L^3(4U-3L)}{6(U-L)^2} \tag{10}$$

$$\mathbb{E}\left[\mathbf{1}\{X<b\}X^3\right] = \frac{b^4-L^4}{4(U-L)} \qquad \mathbb{E}\left[\mathbf{1}\{Z \le b\}Z^3\right] = \frac{5b^4U-4b^5-L^4(5U-4L)}{10(U-L)^2} \tag{11}$$

$$\mathbb{E}[\mathbf{1}\{X<M_x\}X] = \frac{3L+U}{8} \qquad \mathbb{E}[\mathbf{1}\{Z \le M_z\}Z] = \frac{U\left(\sqrt{2}-1\right)+L\left(4-\sqrt{2}\right)}{6} \tag{12}$$

$$\mathbb{E}\left[\mathbf{1}\{X<M_x\}X^2\right] = \frac{7L^2+4LU+U^2}{24} \qquad \mathbb{E}\left[\mathbf{1}\{Z \le M_z\}Z^2\right] = \frac{9L^2+2LU\left(7-4\sqrt{2}\right)+U^2\left(8\sqrt{2}-11\right)}{24} \tag{13}$$

$$\mathbb{E}\left[\mathbf{1}\{X<M_x\}X^3\right] = \frac{(3L+U)\left(5L^2+2LU+U^2\right)}{64} \qquad \mathbb{E}\left[\mathbf{1}\{Z \le M_z\}Z^3\right] = \frac{2L^3\left(8-\sqrt{2}\right)+3L^2U\left(2\sqrt{2}-1\right)}{40} + \frac{2LU^2\left(19-13\sqrt{2}\right)+U^3\left(22\sqrt{2}-31\right)}{40} \tag{14}$$

The results in part (ii) below use 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:

$$F_y(y) = \begin{cases} 0, & y < L \\ \frac{(L-y)^2}{-2LU} & y \in [L, 0] \\ \frac{2y-L}{2U} & y \in [0, L+U] \\ \frac{-2LU-(U-y)^2}{-2LU} & y \in [L+U, U] \\ 1, & y > U \end{cases} \quad \text{and} \quad f_z(z) = \begin{cases} \frac{y-L}{-LU}, & y \in [L, 0] \\ \frac{1}{U}, & y \in [0, L+U] \\ \frac{U-y}{-LU} & y \in [L+U, U] \\ 0, & else \end{cases} \tag{15}$$

$$F_z^{-1}(\alpha) = \begin{cases} L + \sqrt{2\alpha |L| U}, & \alpha \in \left[0, \frac{-L}{2U}\right] \\ \frac{1}{2}L + \alpha U, & \alpha \in \left[\frac{-L}{2U}, \frac{L+2U}{2U}\right] \\ U - \sqrt{2(1-\alpha)|L| U}, & \alpha \in \left[\frac{L+2U}{2U}, 1\right] \end{cases} \tag{16}$$

And then

$$\mathbb{E}[Y] = \frac{1}{2}(L+U) \tag{17}$$

$$\mathbb{E}\left[Y^2\right] = \frac{1}{6}\left(2L^2 + 3LU + 2U^2\right) \tag{18}$$

$$\mathbb{E}\left[Y^3\right] = \frac{1}{4}\left(L^3 + 2L^2U + 2LU^2 + U^3\right)$$

$$Median[Y] = \frac{1}{2}(L+U) \tag{19}$$

$$\mathbb{E}[\mathbf{1}\{Y \leq M_y\} Y] = \frac{1}{24}\left(6L - \frac{L^2}{U} + 3U\right) \tag{20}$$

$$\mathbb{E}\left[\mathbf{1}\{Y \leq M_y\} Y^2\right] = \frac{(L+U)^3 - 2L^3}{24U} \tag{21}$$

$$\mathbb{E}\left[\mathbf{1}\{Y \leq M_y\} Y^3\right] = \frac{5(L+U)^4 - 16L^4}{320U} \tag{22}$$

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

$$\mathbb{E}\left[\mathbf{1}\left\{Y \leq \hat{Y}\right\} | \hat{Y}\right] = \alpha \tag{23}$$

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

$$\begin{aligned} \mathbb{E}\left[L\left(Y, \hat{Y}; g\right)\right] &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}\right\} - \alpha\right)\left(g\left(\hat{Y}\right) - g(Y)\right)\right] \\ &= \mathbb{E}\left[g\left(\hat{Y}\right)\left(\mathbf{1}\left\{Y \leq \hat{Y}\right\} - \alpha\right)\right] - \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}\right\} - \alpha\right) g(Y)\right] \\ &= \mathbb{E}\left[g\left(\hat{Y}\right)\left(\mathbb{E}\left[\mathbf{1}\left\{Y \leq \hat{Y}\right\} | \hat{Y}\right] - \alpha\right)\right] - \mathbb{E}\left[\mathbf{1}\left\{Y \leq \hat{Y}\right\} g(Y)\right] + \alpha \mathbb{E}[g(Y)] \\ &= \alpha \mathbb{E}[g(Y)] - \mathbb{E}\left[\mathbf{1}\left\{Y \leq \hat{Y}\right\} g(Y)\right] \end{aligned} \tag{24}$$

Proof of Proposition 4(b). (i) We first consider the case of non-nested information sets (violating Assumption 1). Consider the following simple example:

$$Y = X + Z \tag{25}$$
$$\text{where } X \sim Unif\,(0,10)\,, \quad Z \sim Tri\,(0,12)\,, \quad X \perp\!\!\!\perp Z$$

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

$$\hat{Y}^a = X + Median\,[Z] = X + 0.45, \text{ since } Median\,[Z] = 12 - 6\sqrt{2} \approx 3.51 \tag{26}$$
$$\hat{Y}^b = Z + Median\,[X] = Z + 2.5, \text{ since } Median\,[X] = 5 \tag{27}$$

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.

$$\bar{L}_A(g_1) \equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} - 1/2\right)\left(\hat{Y}^a - Y\right)\right] \tag{28}$$
$$= \frac{1}{2}\mathbb{E}[Y] - \mathbb{E}\left[\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} Y\right]$$
$$\text{where } \mathbb{E}[Y] = \mathbb{E}[X] + \mathbb{E}[Z] \tag{29}$$
$$\text{and } \mathbb{E}\left[\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} Y\right] = \mathbb{E}\left[\mathbf{1}\left\{X + Z \leq X + M_z\right\}(X+Z)\right] \tag{30}$$
$$= \mathbb{E}\left[\mathbf{1}\left\{Z \leq M_z\right\}\right]\mathbb{E}[X] + \mathbb{E}\left[\mathbf{1}\left\{Z \leq M_z\right\} Z\right], \text{ since } X \perp\!\!\!\perp Z$$
$$= \frac{1}{2}\mathbb{E}[X] + \mathbb{E}\left[\mathbf{1}\left\{Z \leq M_z\right\} Z\right], \text{ since } \mathbb{E}\left[\mathbf{1}\left\{Z \leq M_z\right\}\right] = 1/2$$

We find an analogous expression for the other forecaster:

$$\bar{L}_B(g_1) = \frac{1}{2}\mathbb{E}[Y] - \mathbb{E}\left[\mathbf{1}\left\{Y \leq \hat{Y}^b\right\} Y\right] \tag{31}$$
$$= \frac{1}{2}\mathbb{E}[Y] - \left(\frac{1}{2}\mathbb{E}[Z] + \mathbb{E}\left[\mathbf{1}\left\{X \leq M_x\right\} X\right]\right)$$

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

$$\bar{L}_A(g_2) \equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} - 1/2\right)\left(\left(\hat{Y}^a\right)^3 - Y^3\right)\right] \tag{32}$$
$$= \frac{1}{2}\mathbb{E}\left[Y^3\right] - \mathbb{E}\left[\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} Y^3\right]$$
$$\mathbb{E}\left[Y^3\right] = \mathbb{E}\left[(X+Z)^3\right] = \mathbb{E}\left[X^3\right] + 3\mathbb{E}\left[X^2\right]\mathbb{E}[Z] + 3\mathbb{E}[X]\,\mathbb{E}\left[Z^2\right] + \mathbb{E}\left[Z^3\right] \tag{33}$$
$$\mathbb{E}\left[\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} Y^3\right] = \mathbb{E}\left[\mathbf{1}\left\{Z \leq M_z\right\}\right]\mathbb{E}\left[X^3\right] + 3\mathbb{E}\left[\mathbf{1}\left\{Z \leq M_z\right\} Z\right]\mathbb{E}\left[X^2\right] \tag{34}$$
$$+3\mathbb{E}\left[\mathbf{1}\left\{Z \leq M_z\right\} Z^2\right]\mathbb{E}[X] + \mathbb{E}\left[\mathbf{1}\left\{Z \leq M_z\right\} Z^3\right]$$

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

$$\bar{L}_A(g_1) = 1.17 < 1.25 = \bar{L}_B(g_1) \quad (35)$$

$$\bar{L}_A(g_2) = 350.45 > 349.38 = \bar{L}_B(g_2) \quad (36)$$

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. Let us simplify the DGP and assume that

$$\begin{aligned} Y &= X + Z \quad (37) \\ X &\sim Unif(-10, 0), \;\; Z \sim Unif(0, 12), \;\; X \perp\!\!\!\perp Z \end{aligned}$$

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

$$\hat{Y}^a = Median[Y] = 1 \quad (38)$$

Forecaster B uses information on Z, but to exploit it must estimate $Median[X]$. 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

$$\hat{Y}^b = \tilde{X} + Z \quad (39)$$

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

$$\begin{aligned} \bar{L}_A(g_1) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} - 1/2\right)\left(\hat{Y}^a - Y\right)\right] \quad (40) \\ &= \mathbb{E}\left[\left(\mathbf{1}\{Y \leq M_y\} - 1/2\right)(M_y - Y)\right] \\ &= M_y \mathbb{E}\left[\mathbf{1}\{Y \leq M_y\}\right] - \mathbb{E}\left[\mathbf{1}\{Y \leq M_y\} Y\right] - 1/2\left(M_y - \mathbb{E}[Y]\right) \end{aligned}$$

For forecaster B we find:

$$\begin{aligned}
\bar{L}_B(g_1) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^b\right\} - 1/2\right)\left(\hat{Y}^b - Y\right)\right] \\
&= \mathbb{E}\left[\left(\mathbf{1}\left\{X + Z \leq \tilde{X} + Z\right\} - 1/2\right)\left(\tilde{X} + Z - X - Z\right)\right] \\
&= \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\left(\tilde{X} - X\right)\right] - 1/2\left(\mathbb{E}\left[\tilde{X} + Z - Y\right]\right), \text{ note } \mathbb{E}\left[\tilde{X} + Z - Y\right] = 0 \\
&= \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\tilde{X}\right] - \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}X\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}|\tilde{X}\right]\tilde{X}\right] - \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}|X\right]X\right] \\
&= \mathbb{E}\left[F_x\left(\tilde{X}\right)\tilde{X}\right] - \mathbb{E}\left[(1 - F_x(X))X\right] \\
&= 2\mathbb{E}\left[F_x(X)X\right] - \mathbb{E}[X], \text{ since } \tilde{X} \overset{d}{=} X
\end{aligned} \tag{41}$$

And for the second loss function we obtain:

$$\begin{aligned}
\bar{L}_A(g_2) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} - 1/2\right)\left(\left(\hat{Y}^a\right)^3 - Y^3\right)\right] \\
&= \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq M_y\right\} - 1/2\right)\left(M_y^3 - Y^3\right)\right] \\
&= M_y^3\mathbb{E}\left[\mathbf{1}\left\{Y \leq M_y\right\}\right] - \mathbb{E}\left[\mathbf{1}\left\{Y \leq M_y\right\}Y^3\right] - 1/2\left(M_y^3 - \mathbb{E}\left[Y^3\right]\right)
\end{aligned} \tag{42}$$

and

$$\begin{aligned}
\bar{L}_B(g_2) &\equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^b\right\} - 1/2\right)\left(\left(\hat{Y}^b\right)^3 - Y^3\right)\right] \\
&= \mathbb{E}\left[\left(\mathbf{1}\left\{X + Z \leq \tilde{X} + Z\right\} - 1/2\right)\left(\left(\tilde{X} + Z\right)^3 - (X + Z)^3\right)\right] \\
&= \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\left(\left(\tilde{X} + Z\right)^3 - (X + Z)^3\right)\right] - 1/2\left(\mathbb{E}\left[\left(\tilde{X} + Z\right)^3\right] - \mathbb{E}\left[(X + Z)^3\right]\right) \\
&= \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\left(\tilde{X}^3 + 3\tilde{X}^2 Z + 3\tilde{X}Z^2 - X^3 - 3X^2 Z - 3XZ^2\right)\right] \\
&= \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\tilde{X}^3\right] - \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}X^3\right] \\
&\quad + 3\mathbb{E}[Z]\left(\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\tilde{X}^2\right] - \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}X^2\right]\right) \\
&\quad + 3\mathbb{E}\left[Z^2\right]\left(\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\tilde{X}\right] - \mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}X\right]\right)
\end{aligned} \tag{43}$$

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

$$\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}\tilde{X}^p\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}|\tilde{X}\right]\tilde{X}^p\right] = \mathbb{E}\left[F_x(X)X^p\right], \text{ since } \tilde{X} \overset{d}{=} X \tag{44}$$

$$\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}X^p\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{X \leq \tilde{X}\right\}|X\right]X^p\right] = \mathbb{E}\left[(1 - F_x(X))X^p\right] = \mathbb{E}[X^p] - \mathbb{E}\left[F_x(X)X^p\right] \tag{45}$$

And so

$$\bar{L}_B(g_2) = 2\mathbb{E}\left[F_x(X)X^3\right] - \mathbb{E}\left[X^3\right] \tag{46}$$
$$+3\mathbb{E}[Z]\left(2\mathbb{E}\left[F_x(X)X^2\right] - \mathbb{E}\left[X^2\right]\right)$$
$$+3\mathbb{E}\left[Z^2\right](2\mathbb{E}[F_x(X)X] - \mathbb{E}[X])$$

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

$$\mathbb{E}[F_x(X)X] = \frac{L+2U}{6} \tag{47}$$

$$\mathbb{E}\left[F_x(X)X^2\right] = \frac{L^2+2LU+3U^2}{12} \tag{48}$$

$$\mathbb{E}\left[F_x(X)X^3\right] = \frac{L^3+2L^2U+3LU^2+4U^3}{20} \tag{49}$$

Pulling these terms together, we find that

$$\bar{L}_A(g_1) = 1.85 > 1.67 = \bar{L}_B(g_1) \tag{50}$$

$$\bar{L}_A(g_2) = 72.65 < 90 = \bar{L}_B(g_2) \tag{51}$$

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

$$Y = X \sim Unif(0, 10) \tag{52}$$

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

$$\hat{Y}^a = \beta_0 + \beta_1 X \tag{53}$$

$$\hat{Y}^b = \gamma_0 + \gamma_1 X \tag{54}$$

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:

$$\bar{L}_A(g_1) \equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} - 1/2\right)\left(\hat{Y}^a - Y\right)\right] \tag{55}$$

$$= \mathbb{E}\left[\left(\mathbf{1}\left\{X \leq \beta_0 + \beta_1 X\right\} - 1/2\right)\left(\beta_0 + \beta_1 X - X\right)\right]$$

$$= \beta_0 \mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \leq \beta_0\right\}\right] - \frac{\beta_0}{2} - \frac{\beta_1 - 1}{2}\mathbb{E}[X]$$

$$+ (\beta_1 - 1)\,\mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \leq \beta_0\right\} X\right]$$

$$\mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \leq \beta_0\right\}\right] = \begin{cases} F_x\left(\frac{\beta_0}{1-\beta_1}\right), & \beta_1 < 1 \\ 1 - F_x\left(\frac{\beta_0}{1-\beta_1}\right), & \beta_1 > 1 \\ \mathbf{1}\left\{\beta_0 \geq 0\right\}, & \beta_1 = 1 \end{cases} \tag{56}$$

$$\mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \leq \beta_0\right\} X^p\right] = \begin{cases} \mathbb{E}\left[\mathbf{1}\left\{X \leq \frac{\beta_0}{1-\beta_1}\right\} X^p\right], & \beta_1 < 1 \\ \mathbb{E}[X^p] - \mathbb{E}\left[\mathbf{1}\left\{X \leq \frac{\beta_0}{1-\beta_1}\right\} X^p\right], & \beta_1 > 1 \\ \mathbf{1}\left\{\beta_0 \geq 0\right\}\mathbb{E}[X^p], & \beta_1 = 1 \end{cases} \tag{57}$$

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

$$\bar{L}_A(g_2) \equiv \mathbb{E}\left[\left(\mathbf{1}\left\{Y \leq \hat{Y}^a\right\} - 1/2\right)\left(\left(\hat{Y}^a\right)^3 - Y^3\right)\right] \tag{58}$$

$$= \mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \leq \beta_0\right\}\left(\left(\beta_0 + \beta_1 X\right)^3 - X^3\right)\right] - 1/2\mathbb{E}\left[\left(\beta_0 + \beta_1 X\right)^3 - X^3\right]$$

$$= \beta_0^3 \mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \leq \beta_0\right\}\right] + 3\beta_0^2\beta_1 \mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \leq \beta_0\right\} X\right]$$

$$+ 3\beta_0\beta_1^2 \mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \leq \beta_0\right\} X^2\right] + \left(\beta_1^3 - 1\right)\mathbb{E}\left[\mathbf{1}\left\{(1-\beta_1) X \leq \beta_0\right\} X^3\right]$$

$$- 1/2\left(\beta_0^3 + 3\beta_0^2\beta_1 \mathbb{E}[X] + 3\beta_0\beta_1^2 \mathbb{E}\left[X^2\right] + \left(\beta_1^3 - 1\right)\mathbb{E}\left[X^3\right]\right)$$

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

$$\bar{L}_A(g_1) = 0.68 > 0.51 = \bar{L}_B(g_1) \tag{59}$$

$$\bar{L}_A(g_2) = 79.44 < 100.19 = \bar{L}_B(g_2) \tag{60}$$

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

$$\begin{aligned} Y &= -\beta_2 A - \beta_1 (1-A) + \beta_1 B + \beta_2 (1-B) \\ A &\sim Bernoulli\,(p) \\ B &\sim Bernoulli\,(q)\,, \;\; B \perp\!\!\!\perp A \\ \beta_2 &> \beta_1 > 0 \end{aligned} \tag{61}$$

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:

$$\begin{aligned} \mathbb{E}\left[wCRPS\left(F_A, Y, \omega\right)\right] &= pq\,(1-q)\int_{\beta_1-\beta_2}^{0} \omega(z)\,dz + q\,(1-p)\,(1-q)\int_{0}^{\beta_2-\beta_1} \omega(z)\,dz \\ \mathbb{E}\left[wCRPS\left(F_B, Y, \omega\right)\right] &= pq\,(1-p)\int_{\beta_1-\beta_2}^{0} \omega(z)\,dz + p\,(1-p)\,(1-q)\int_{0}^{\beta_2-\beta_1} \omega(z)\,dz \end{aligned} \tag{62}$$

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

$$\omega(z;a) = \frac{1}{1+\exp\{-az\}} \tag{63}$$

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 $\omega_L(z) = \omega(z;-1)$

$$\begin{aligned} \int_{\beta_1-\beta_2}^{0} \omega_R(z)\,dz &= \int_{0}^{\beta_2-\beta_1} \omega_L(z)\,dz = \beta_2 + \log 2 - \log\left(\exp\{\beta_2\} + \exp\{\beta_1\}\right) \\ \int_{0}^{\beta_2-\beta_1} \omega_R(z)\,dz &= \int_{\beta_1-\beta_2}^{0} \omega_L(z)\,dz = \log\left(\frac{1}{2}\left(1+\exp\{\beta_2-\beta_1\}\right)\right) \end{aligned} \tag{64}$$

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

$$\mathbb{E}\left[wCRPS\left(F_A, Y; \omega_R\right)\right] = 0.50 > 0.25 = \mathbb{E}\left[wCRPS\left(F_B, Y; \omega_R\right)\right] \tag{65}$$

$$\mathbb{E}\left[wCRPS\left(F_A, Y; \omega_L\right)\right] = 0.25 < 0.50 = \mathbb{E}\left[wCRPS\left(F_B, Y; \omega_L\right)\right] \tag{66}$$

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

$$\mathbb{E}\left[wCRPS\left(\bar{F}_A, Y; \omega_R\right)\right] = 0.61 > 0.33 = \mathbb{E}\left[wCRPS\left(\tilde{F}_B, Y; \omega_R\right)\right] \quad (67)$$

$$\mathbb{E}\left[wCRPS\left(\bar{F}_A, Y; \omega_L\right)\right] = 0.61 < 0.67 = \mathbb{E}\left[wCRPS\left(\tilde{F}_B, Y; \omega_L\right)\right] \quad (68)$$

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 $\omega_R(z) = \omega_L(-z)$.)

(ii) 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

$$n\hat{p} = \sum_{i=1}^{n} A_i \sim Binomial(n, p) \quad (69)$$

In this case, we have:

$$\mathbb{E}\left[wCRPS\left(\hat{F}_B(\hat{p}), Y; \omega\right)\right] = \sum_{\tilde{p}} \mathbb{E}\left[wCRPS\left(\tilde{F}_B(\tilde{p}), Y; \omega\right)\right] \Pr\left[\hat{p} = \tilde{p}\right] \quad (70)$$

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

$$\mathbb{E}\left[wCRPS\left(\bar{F}_A, Y; \omega_R\right)\right] = 0.61 > 0.31 = \mathbb{E}\left[wCRPS\left(\hat{F}_B(\hat{p}), Y; \omega_R\right)\right] \quad (71)$$

$$\mathbb{E}\left[wCRPS\left(\bar{F}_A, Y; \omega_L\right)\right] = 0.61 < 0.62 = \mathbb{E}\left[wCRPS\left(\hat{F}_B(\hat{p}), Y; \omega_L\right)\right] \quad (72)$$

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