Probability forecasts with observation error: is the Brier score proper?

Ian Jolliffe

University of Exeter, UK

i.t.jolliffe@ex.ac.uk

1. Observations with error – examples
2. Propriety of usual Brier score
3. Is it still proper when errors are present?
4. Examples revisited

Observations with error

- Event of interest is occurrence of ‘severe weather’ somewhere in a large geographical area, but observations are sparse – occurrences of the event may be missed.
- Event of interest is related to a threshold of a continuous variable, for example temperature below 0°C, but measuring instrument has errors. Mistakes could be made in either direction.
  - [Candille &Talagrand (2008), QJRMS, 134, 959-971, consider properties of the Brier score, though not its propriety, for a related scenario.]

The Brier score

If (f_i, o_i) i = 1,2, …, n are a set of n forecast probabilities of an event and the corresponding observations, then the Brier score is

$$B = \frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2$$

f_i can take any value between 0 and 1; o_i is 1 or 0, depending on whether or not the event occurred.

The Brier score is proper. It cannot be hedged. Its expected value cannot be improved by issuing a forecast probability other than that which is believed to be correct.

Children should early be taught the lesson of Propriety and Good Manners.

Propriety of the Brier score

- Suppose that *f* is a probability forecast and *o* is the corresponding 0-1 observation. The forecaster believes the probability of the event is *p*, but wants to know if the expected Brier score can be improved by forecasting *f*, which may differ from *p*.

$$E[(o-f)^2] = f^2(1-p) + (1-f)^2 p.$$

Differentiating this expected score with respect to *f*, shows that it is uniquely minimised when *f* = *p*, so there is no virtue in hedging. The score is strictly proper.

But what if observations are made with error?

Two questions

- What are we forecasting?

If observations are made with error, we could forecast:

  - p, the probability of the event;
  - q, the probability that our observation says the event has occurred.

- What are the observations?

We could take these as:

  - 0 or 1 depending on whether the observations say the event has occurred;
  - The probability that the event has occurred given the observation.

Is the Brier score proper?

	Forecast *p*	Forecast *q*
Observe 0,1	No	Yes
Observe probability	Yes	No

The simplest case is top-right above. Simply replace *p* by *q* in the earlier proof of propriety. If *q* is forecast and the observations are taken as 0 or 1, the Brier score is proper.

Is the Brier score proper?

	Forecast p	Forecast q
Observe 0,1	No	Yes
Observe probability	Yes	No

Notation: denote the occurrence of the event as E and the observation indicating that the event has occurred as I. Then Pr(E) = p; Pr(I) = q. Let c_1 = Pr(I|E); c_0 = Pr(I|$\bar{E}$). Then $q = pc_1 + (1-p)c_0 = c_0 + p(c_1-c_0)$.

The equations for the expected Brier score and its derivatives will be the same as in the last case i.e choose $f = q \neq p$, unless $p = c_0/(1-c_1+c_0)$. Values of p should be hedged upwards (downwards) if $p < c_0/(1-c_1+c_0)$ ($p > c_0/(1-c_1+c_0)$)

Thus the Brier score is not proper. p should be hedged to the probability, q, of the only thing we can actually observe.

Is the Brier score proper?

	Forecast p	Forecast q
Observe 0,1	No	Yes
Observe probability	Yes	No

More notation: Let $d_1 = \Pr(E|I)$; $d_0 = \Pr(E|\bar{I})$. These are now the 'observations' and the expected Brier score becomes:

$$E[(o-f)^2] = (d_0 - f)^2(1-q) + (d_1 - f)^2 q$$

Differentiating with respect to f and equating to zero gives
$(1-q)(d_0-f) + q(d_1-f) = 0$ and

$$f = \frac{(1-q)d_0 + qd_1}{(1-q)+q} = \Pr(\bar{I})P(E|\bar{I}) + \Pr(I)P(E|I) = \Pr(E) = p$$

So if *p* is forecast, hedging has no advantage, and the Brier score is proper when the observations are appropriate probabilities.
But if *q* is forecast with such observations, hedging should take place to *p.*

Missed events

- Suppose that severe weather is forecast somewhere in a large geographical area with sparse observations. Assume that a severe weather event may be missed with probability ($1-c_1$), but 'false occurrences' are virtually unknown, so $c_0 \approx 0$. Forecasts of p should then be hedged to $q = pc_1$.
- For example with $p = 0.05$ and $c_1 = 0.8$, we have $q = 0.04$.

Missed events II

$$d_1 = Pr(E|I) = 1$$

$$d_0 = \Pr(E \mid \bar{I}) = \frac{\Pr(\bar{I} \mid E)P(E)}{\Pr(\bar{I} \mid E)P(E) + \Pr(\bar{I} \mid \bar{E})P(\bar{E})} = \frac{(1-c_1)p}{(1-c_1)p + (1-p)}$$

These quantities are required in order to use 'probabilities' as 'observations' in order to evaluate the Brier score. This is turn needs knowledge of *p*.

Example with $p = 0.05$, $c_1 = 0.8$: $d_0 = 0.01/0.96 = 0.0104$.

But is this case ever relevant? Why not forecast *q* and verify against the 0-1 observation?

Temperature below/above threshold

- Let T be the observed temperature and τ be the true temperature. Suppose that our event E is $\tau < 0°C$. Rather than have I as T < 0°C, at first sight it seems more appropriate here to condition on the actual observed and true temperatures.
- Suppose that $T|\ \tau \sim N(\tau, \sigma_T^2)$; $\tau \sim N(\mu, \sigma_\tau^2)$. Then

$$\tau \mid T \sim N\left(\left(\frac{\mu}{\sigma_\tau^2} + \frac{T}{\sigma_T^2}\right) / \left(\frac{1}{\sigma_\tau^2} + \frac{1}{\sigma_T^2}\right), \left(\frac{1}{\sigma_\tau^2} + \frac{1}{\sigma_T^2}\right)^{-1}\right)$$

Temperature above/below a threshold II

- The parameters μ, σ_T^2, σ_τ^2 need to be specified. Once they are, we can calculate:
  - *p* from the distribution of τ;
  - Pr(I|τ) instead of c_0 and c_1;
  - Pr(E|T) instead of d_0 and d_1.
- To get c_0, c_1 and *q* it is necessary to integrate over τ.
- To get d_0, d_1, it is necessary to integrate over T.

Temperature above/below a threshold

We can also find the joint (bivariate Gaussian) distribution for T and τ which has the general form:

$$E[T] = E[\tau] = \mu; \mathrm{var}(\tau) = \sigma_\tau^2; \mathrm{var}(T) = \sigma_T^2 + \sigma_\tau^2; \mathrm{cov}(T, \tau) = \sigma_\tau^2$$

From this joint distribution we can calculate *p, q,* c_0, c_1, d_0, d_1 (do the appropriate integrations) using R package *mvtnorm.*

Temperature below a threshold – example

- Let $\sigma_T^2 = 0.25; \sigma_\tau^2 = 9$

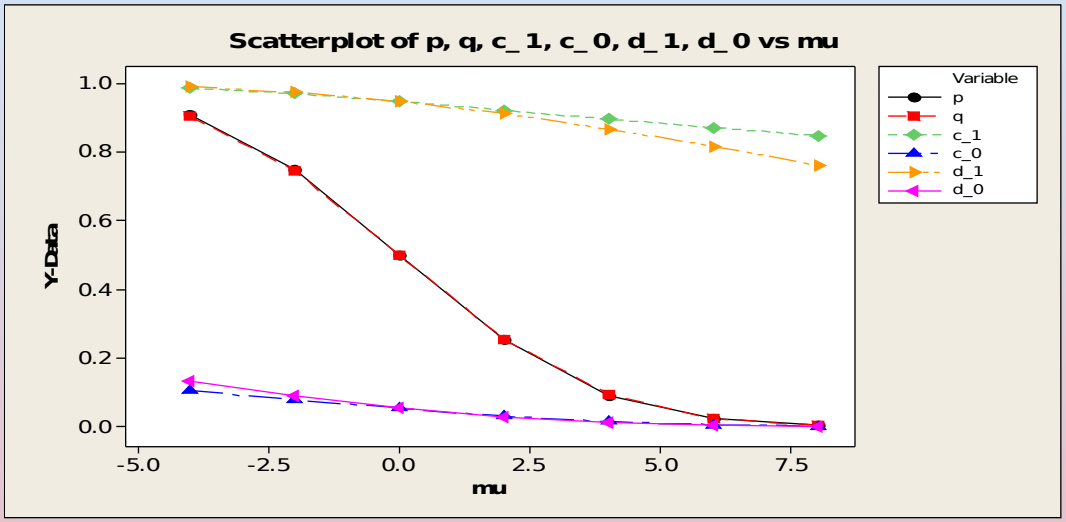
  This implies that 'true' temperatures are mostly within a range of about 12 degrees and measurement error is usually no more than ± 1 degree.
  With E = {$\tau < 0$} and I = {$T < 0$}, we calculate p, q, c_0, c_1, d_0, d_1 for various values of μ.

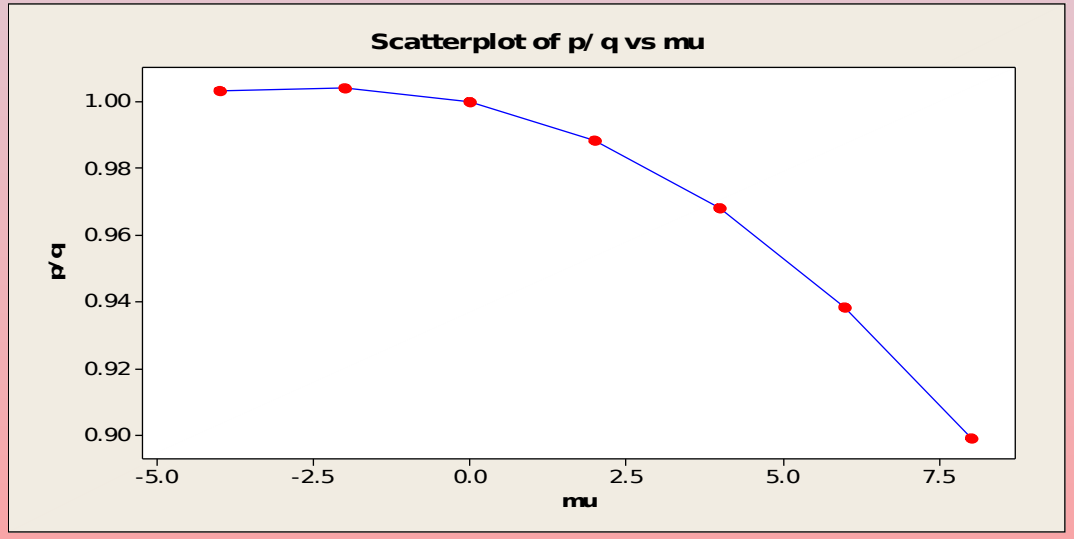
$$T \mid \tau \sim N(\tau, 0.25)$$
$$\tau \mid T \sim N(0.973T + 0.027\mu, 0.24)$$

Example – the numbers

	p	q	c_1	c_0	d_1	d_0
μ=-4	0.909	0.906	0.986	0.104	0.990	0.133
μ=-2	0.748	0.745	0.970	0.078	0.974	0.088
μ=0	0.500	0.500	0.947	0.053	0.947	0.053
μ=2	0.252	0.255	0.922	0.030	0.912	0.026
μ=4	0.091	0.094	0.896	0.014	0.867	0.011
μ=6	0.023	0.024	0.871	0.005	0.817	0.003
μ=8	0.004	0.004	0.848	0.001	0.761	0.001

Example – some pictures



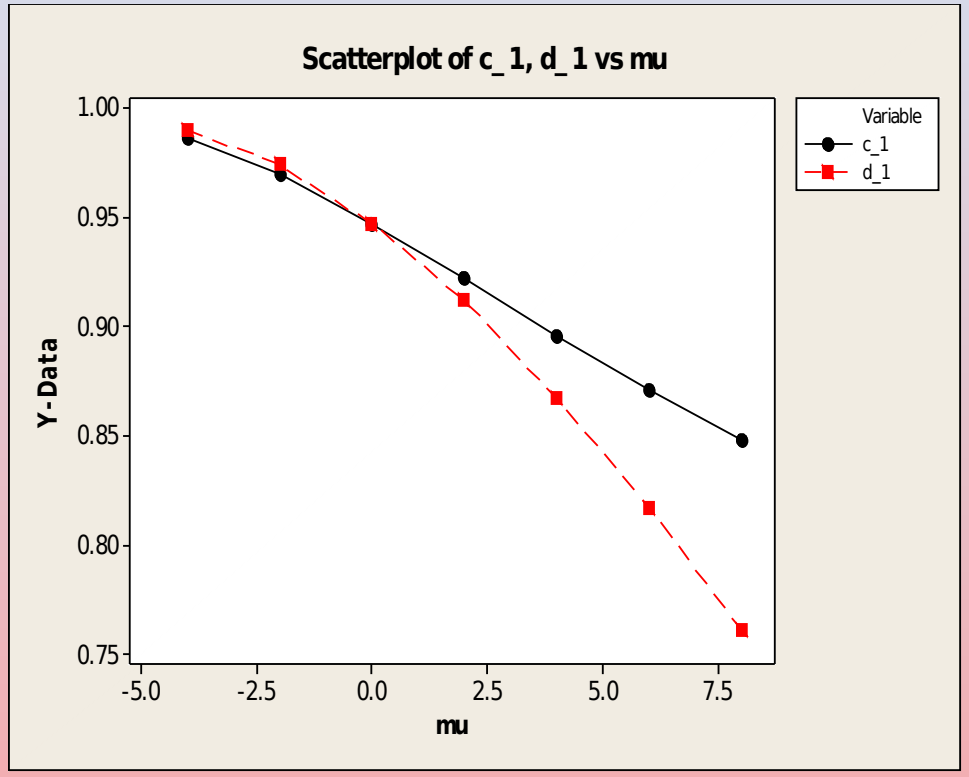


The central pair of lines give p and q. They are very close.

Any hedging will be very small in this example.

But the ratio p/ q decreases fairly rapidly as μ gets a long way from the threshold.

Example – c_1 and d_1

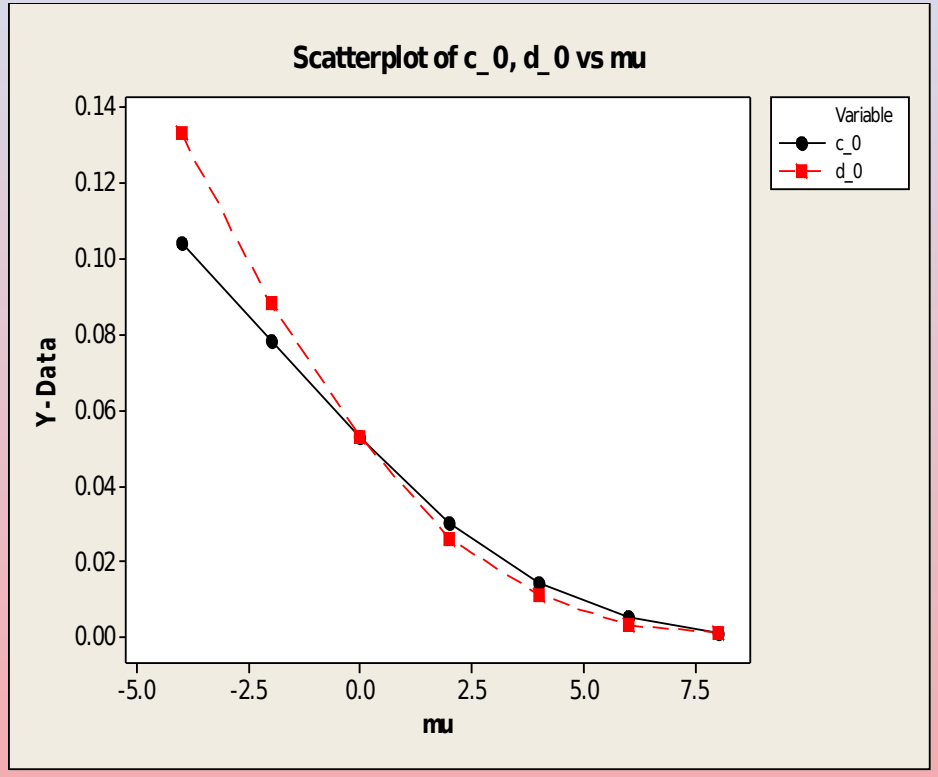


c_1 and d_1 may be needed to calculate how much to hedge, but are of interest in their own right.

Note that for small values of p,q, they are very much larger than q, p. Knowledge of E (I) increases the probability that I (E) occurs.

d_1 drops more rapidly than c_1 as μ gets further above the threshold because of its dependence on μ

Example - c_0 and d_0



Note that for small values of p,q, c_0, d_0 are very much smaller than q, p. Knowledge of $\bar{E}$ ($\bar{I}$) decreases the probability that I (E) occurs.

What have we learned?

- For probability forecasts of binary events where observations are made with error, we should hedge to our belief for the probability that we can actually observe or calculate. Obvious with hindsight? Should it have been obvious beforehand?
- With suitability chosen models for the error mechanism, we can calculate how much to hedge, as well as conditional probabilities of interest.

Questions?

Comments?