SCORING MEASURES FOR PREDICTION PROBLEMS

Prof. Tom Fomby
Department of Economics
Southern Methodist University
Dallas, TX 75275
6/08/06 and 2/4/08

Let $S_j,\ j = 1,\cdots,V$ denote the V validation data prediction scores derived for some prediction method (e.g. multiple regression or artificial neural network) built using the training data set. Further, let $A_j,\ j = 1,\cdots,V$ denote the V actual values of the target variable contained in the validation data set. Last, let $E_j = A_j - S_j,\ j = 1,\cdots,V$ denote the V scoring (prediction) errors (over the validation data set) associated with the chosen prediction method. The following are commonly used scoring measures used to evaluate prediction methods vis-à-vis the validation data set.

MEAN ABSOLUTE ERROR: MAE

$$MAE = \frac{\sum_{j=1}^{V} | E_j |}{V}$$

MEAN SQUARE ERROR: MSE

$$MSE = \frac{\sum_{j=1}^{V} E_j^2}{V}$$

ROOT MEAN SQUARE ERROR: RMSE

$$RMSE = \sqrt{MSE}$$

MEAN ABSOLUTE PERCENTAGE ERROR: MAPE

$$MAPE = \frac{\sum_{j=1}^{V} \left| \frac{E_j}{A_j} \right|}{V}$$

MEAN SQUARE PERCENTAGE ERROR: MSPE

$$MSPE = \frac{\sum_{j=1}^{V} \left(\frac{E_j}{A_j} \right)^2}{V}$$

ROOT MEAN SQUARE PERCENTAGE ERROR: RMSPE

$$RMSPE = \sqrt{MSPE}$$

When considering the "percentage error" associated with a particular method as revealed by the validation data, several methods can be used. One can compute the ratio of the MAE or RMSE to the mean of the target variable to get a percentage error for the validation data. Alternatively, one can directly compute the MAPE or RMSPE over the validation data and get the percentage error that way. In both cases one has to be sure that the respective measures are multiplied by 100 in order to convert them to percentages.

Of course, these are not the only available ways to score a prediction method. If one has in hand a non-negative loss, say, $L(E_j)$. Then for this specific loss function the **Mean Loss Function** is

$$\overline{L} = \frac{\sum_{j=1}^{V} L(E_j)}{V}.$$

An example of such a loss function might be

$$L(E_j) = \left\{ \begin{matrix} 12E_j, when E_j > 0 \\ 6 \mid E_j \mid, when E_j \le 0 \end{matrix} \right\}.$$

In all of these cases, the lower the measure, the better the prediction method has scored.

**Testing for a Significant Difference
In the Forecasting Accuracies
Of Two Competing Prediction Methods
(2-4-08)**

In addition to observing the superiority of one prediction method over another prediction method using a comparison of mean losses incurred in the validation data set, say $\bar{L}_1 < \bar{L}_2$, one might wish to know if the mean loss of one method is significantly less than the mean loss of another method. If one is willing to assume that, observation by observation, the loss scores of the two competing prediction methods come from two independent populations, then we are interested in testing the equality of the population means of the losses of the two methods versus the alternative hypothesis that the population mean of the losses arising from method 1 is less than the population mean of the losses arising from method 2. That is we are interested in testing

$$H_0 : \mu_{L_1} = \mu_{L_2}$$

versus

$$H_1 : \mu_{L_1} < \mu_{L_2} \quad .$$

This test can be carried out by using a standard t-test of the difference of population means that assumes that the population variances of the two losses are equal to each other or, alternatively, using a Fisher-Behrens t-test that allows the population variances of the losses to be unequal. The raw data for these tests are, of course, the validation data set losses for method 1, $L_1(E_j),\ j = 1,2,\cdots,V$, as one independent data set and the validation data set losses for method 2, $L_2(E_j),\ j = 1,2,\cdots,V$, as the other independent data set.

Of course, if one is interested in testing the simultaneous equality of the losses of more than two competing prediction methods, one would need to pursue a multivariate test of equality of means as in the Hotelling T^2 statistic.