Describes a method for combining different forecasts using mean-squared error and standard deviation.

Consensus Forecasting Using Relative Error Weights

Conway Lackman and Charles Brandon

Marketing Intelligence & Planning, Vol. 12 No. 1, 1994, pp. 37-41
© MCB University Press Limited, 0263-4503

Introduction

An effective method of combining different forecasts obtained from two or more models to produce one forecast that is more accurate than any of those included in the combination is one that assigns weights to each model on the basis of the relative size of the errors produced by each in the past.

Models that have produced forecasts with smaller errors are given more weight than those with larger errors. The scheme assumes that each model has some value since each captures certain information that might be missed by the others.

Discarding all of the forecasts in favour of the one with the smallest error – the most popular method of selecting the "best" forecast – results in the loss of much valuable information. No one model, no matter how small an error it produces, is capable of capturing all information contained in the data.

Good Results

A weighting scheme that assigns greater weight to models that produce smaller errors and less to those with larger errors and combining them into one forecast is more likely, on the average, to capture a larger portion of the available information. This approach is used in actual forecasting applications with good results.

Following is a description of how to compute the weights for each of the forecasts that are included in the combined forecast. Note that in the following discussion, a two-model combination is demonstrated; more models can be combined using the same principles.

The most benefit, however, is derived when two models are combined, with decreasing benefit for each model added. (The formula for the *n* model combination appears in Brandon *et al*., 1984.)

MSE and SD

The information required to determine the weights of each model are:

(1) The mean-squared error (MSE) of each model.

(2) The standard deviation (SD) of each model.

(See Tables I and II for methods of computing these two values.)

Table I. *Computation of Mean-Squared Error (MSE)*

Row	Period	Actual[a,b]	Forecast[c]	Error	Error squared
			Model A		
			(Col. 3-Col. 2)	(Col. 4)^2	
	(1)	(2)	(3)	(4)	(5)
1	1	1928	2124	–196	38,416
2	2	1996	2063	– 67	4,489
3	3	2078	1995	83	6,889
4	4	2030	2037	– 7	49
5	Sum of squared error				49,843

$$\text{MSE (Model A)} = \sqrt{\frac{\text{Sum of squared error (Row 5. Col. 5)}}{\text{No. of observations}}} = \sqrt{\frac{49{,}843}{4}} = 12{,}461$$

Row	Period	Actual[a,b]	Forecast[c]	Error	Error squared
			Model B		
			(Col. 6-Col. 2)	(Col. 7)^2	
	(1)	(2)	(6)	(7)	(8)
1	1	1928	2084	–156	24,336
2	2	1996	2118	–122	14,884
3	3	2078	2182	–104	10,816
4	4	2030	2218	–188	35,344
5	Sum of squared error				85,380

$$\text{MSE (Model B)} = \sqrt{\frac{\text{Sum of squared error (Row 5. Col. 8)}}{\text{No. of observations}}} = \sqrt{\frac{85{,}380}{4}} = 21{,}345$$

[a]Refer to column 3 in Table III.
[b]Refer to column 5 in Table III.
[c]Refer to column 4 in Table III.

Table II. *Computation of Standard Deviation*

Period	(SD) Model A		
	Squared error[a]	MSE[b]	Difference squared (Col. 2-Col. 3)^2
(1)	(2)	(3)	(4)
1	38,416	12,461	673,675,135
2	4,489	12,461	63,548,812
3	6,889	12,461	31,044,464
4	49	12,461	154,051,616
5	Sum		922,320,000

$$\text{SD (Model A)} = \sqrt{\frac{\text{Sum of differences squared (Row 5, Col. 4)}}{\text{No. of observations} - 1}} = \sqrt{\frac{922{,}320{,}000}{4-1}}$$

$= 17{,}534$

Period	Model B		
			(Col. 5-Col. 6)^2
(1)	(5)	(6)	(7)
1	24,336	21,345	8,946,018
2	14,884	21,345	41,744,636
3	10,816	21,345	110,859,840
4	35,344	21,345	195,972,128
5	Sum		357,522,720

$$\text{SD (Model B)} = \sqrt{\frac{\text{Sum of differences squared (Row 5, Col. 7)}}{\text{No. of observations} - 1}} = \sqrt{\frac{357{,}522{,}720}{4-1}}$$

$= 10{,}917$

[a]Refer to column 5, rows 1-4, in Table I.

[b]Refer to Table I.

The first step in the procedure involving two models is to determine how much weight to assign to each. Two equations are used to accomplish this; one to compute the weight for the forecast from model A and the other to compute the weight for the forecast from model B. (Note that the models are labelled "A" and "B" for convenience only; any notation to distinguish between them can be used.) These labels do not change but the weights assigned to each change through time because the MSE and SD are re-calculated at the end of each period.

The Equations

The weight assigned to model A is calculated using the following equation:

$$W_{A.t} = \frac{\text{MSE}_{B.t} \div \text{SD}_{B.t}}{(\text{MSE}_{A.t} \div \text{SD}_{A.t}) \div (\text{MSE}_{B.t} \div \text{SD}_{B.t})}. \tag{1}$$

The equation says simply that the weight assigned to model A to obtain a forecast of period (t+1) is a function of the relative errors as of the current period (t). ("Current period" designates the most recent period of the data set used to calculate the weights, which may or may not be the current period in real time.)

The weight assigned to model B is simply one minus the weight assigned to model A. This assures that the weights sum to one. In equation form, the weight for model B is:

$$W_{B.t} = 1 - W_{A.t}. \tag{2}$$

Weight Changes

As the process moves through time the weights for the two models change continually based on their respective mean-squared errors and standard deviations. So it is possible that the weight for one model can shift from a small percentage to a large one and vice versa, depending on the historical accuracy (error), of one model relative to the other. The sum of the percentages for both models cannot exceed 100 per cent (or 1).

The combined forecast is derived by taking both of the individual forecasts and multiplying them by the weights:

$$F_{C.t+1} = (F_{A.t+1} \times W_{A.t}) + (F_{B.t+1} \times W_{B.t}) \tag{3}$$

where:

- $F_{C.t+1}$ = Combined forecast of the next period.
- $F_{A.t+1}$ = Forecast value of the next period obtained from model A.
- $F_{B.t+1}$ = Forecast value of the next period obtained from model B.
- $W_{A.t}$ = Weight of model A computed as of the current period (the most recent period in the series).
- $W_{B.t}$ = Weight of model B computed as of the current period (the most recent period in the series).

Since ex-post forecasts are prepared of each period in the historical data series, values of the MSE and SD must be computed for each period. The computation of the MSE is based on the length of "memory" desired by the forecaster; that is, the number of periods of past data that are included in the computations.

At one extreme, all of the past periods in the data series may be used on a cumulative basis. Under this scheme, the squared error of the current period is added to the cumulative total of the squared errors of all previous periods; the MSE is determined by dividing that total by the number of observations. The weights under this scheme become progressively less responsive to changing data patterns.

At the other extreme, a more dynamic and responsive weighting scheme is developed by using two, three, or four of the most recent errors as the basis for determining the MSE.

Moving Average

Assume that a forecaster has decided to use the four most recent errors to determine weights. Under this scheme, a forecast of period 5 utilizes errors from period t, t–1, t–2, and t–3; that is periods 4, 3, 2, and 1. Note that the composition of the four periods changes as the process moves through time; the earliest period, that is, the period furthest removed from the most recent period, is dropped as the next is added. This is a four-point moving average scheme.

The forecast of period 6, for example, is based on errors produced by each model in periods 2, 3, 4, and 5 since we are producing the forecast of period 6 at the end of period 5.

In the following example, a four-point moving average scheme is demonstrated. Since there is no hard-and-fast rule on this point, one may wish to try a variety of memory lengths, and select the one that produces weights resulting in a combined forecast with the smallest error. An important consideration is that the cost of experimentation with different schemes could offset any value derived from increased accuracy.

Stop Command

One way to save computing time is to establish limits in the computer program, such as a command to stop when the difference in the size of the error between one trial and the next is less than some fixed percentage. To illustrate the application of the weighting scheme, consider the data series contained in Table III. The 14 quarters of actual data shown in column 2 are arbitrarily divided into two sets: the first contains four periods of actual data, the second, ten periods. The second set represents a holdout sample against which to test the accuracy of the combined forecast. This holdout sample is not required in an actual combination application; it is suggested as a way to give the forecaster confidence in using the methodology.

Table III also shows quarterly ex-post forecasts from two models. A (column 4) and B (column 5). (These models are not identified by type because the types of models used to prepare the forecasts have no effect on the weighting technique.)

Table III. *Actual and Forecast Data*

Date	Time period	Actual	Forecast values Model A	Model B
(1)	(2)	(3)	(4)	(5)
1988-I	1	1,928	2,124	2,084
1988-II	2	1,996	2,063	2,118
1988-III	3	2,078	1,995	2,182
1988-IV	4	2,030	2,037	2,218
1990-I	5	2,064	2,125	2,272
1990-II	6	2,386	2,106	2,464
1990-III	7	2,432	2,116	2,517
1990-IV	8	2,542	2,386	2,722
1991-I	9	2,663	2,513	2,804
1991-II	10	2,767	2,597	2,919
1991-III	11	2,896	2,728	3,017
1991-IV	12	3,014	2,849	3,032
1992-I	13	3,146	2,964	3,107
1992-II	14	3,015	3,135	3,147

Calculating MSE, SD

Since we chose to develop weights on the basis of a four-point moving average as discussed above, the MSE and SD of each of the two models, A and B, are computed from the actual and forecast data of periods 1 through 4 (columns 2, 3, and 6, Table I) for the purpose of preparing an ex-post forecast of period 5.

The MSE of each of the two models is shown at the foot of Table I. The SD of each is shown at the foot of Table II. (Methods used to compute the MSE and SD are shown in Tables I and II respectively.) The MSE and SD values calculated in Tables I and II are entered into equation 1 to calculate the weight of model A:

$$W_{A.5} = \frac{(MSE_{B.4}) \div (SD_{B.4})}{(MSE_{A.4} \div SD_{A.4}) + (MSE_{B.4} \div SD_{B.4})}$$

$$W_{A.5} = \frac{21{,}345 \div 10{,}917}{(12{,}461 \div 17{,}534) + (21{,}345 \div 10{,}917)}$$

$$= 0.733, \text{ or } 73.3 \text{ per cent (weight for model A).}$$

(*Note*: The weight assigned to period 5 ($t + 1$) is the weight calculated as of period 4(t).)

The weight of model B is computed by substituting values in equation (2):

$$W_{B.5} = 1 - W_{A.4}$$
$$= 1 - 0.733 = 0.267.$$

Note that in two-model combination the weight for model B could be determined first by replacing the numerator in equation 1 with the MSE of A divided by the SD of A and keeping the denominator the same. The weight of A would then be one minus the weight determined for B. The combined ex-post forecast value for period 5 using equation 3, is:

$$F_{c.5} = (F_{A.5} \times W_{A.4}) + (F_{B.5} \times W_{B.4})$$
$$= (733 \times 2125) + (267 \times 2272)$$
$$= 2164 \text{ (combined forecast for period 5).}$$

Note that $F_{A.5}$ is the forecast value of period 5 obtained from model A and $F_{B.5}$ is the forecast value of period 5 obtained from model B. Ex-post combined forcasts are prepared for each of the remaining quarters in the holdout sample following the procedures outlined above. For example, the four periods 2 to 5 are used to calculate weights for period 6, the four periods 3 through 6 for period 7, etc. Combined ex-post forecasts for the ten-period holdout sample were calculated in this way. The results appear in Table IV.

Once the forecaster is satisfied by the use of the ex-post process that the procedure is operating correctly, and that errors are within established boundaries, the next step is to repeat the steps to obtain an ex-ante (before-the-fact) combined forecast for the next period – in this case period 15. Forecasts of several periods beyond the next are prepared as follows:

(1) Prepared forecasts using model A and model B for as many periods ahead as is prudent.

(2) Assign the same weights to each future period as those used to prepare the combined forecast of period 15. (New weights cannot be computed as of periods beyond period 14, because there are no actual data available beyond period 14 on which to compute them.)

Forecasts of periods beyond 16 are updated at the end of each period by using the actual data of period 16 when it becomes available, and using it as the most recent data (period *t*) to calculate new weights as of period 16. Use those weights to update forecasts of all periods beyond period 16. Perform this procedure at the end of each period for as long as the combined forecast is performing within reasonable doubts.

Does the Scheme Work?

Does combining forecasts using the weighting scheme presented in this article perform better than either of the two individual models used in the combination? On average, the combination is better.

Evidence for this conclusion is found in Tables IV and V. The foot of Table IV shows the level of accuracy for each of the two models and the combined forecast measured by the overall MSE of the 10-quarter holdout sample. Model A has an MSE of 36,069, and model B 16,587, and the combined forecast 8,614.

The MSE of the combined forecast of 76 per cent less than model A and 48 per cent less than model B. This represents a significant reduction in errors and a considerable improvement in overall forecasting accuracy.

Table V. *Evaluation of Combined Forecasts*

Periods	Squared error Model B Wins	Squared error Model B Losses	Squared error Combined Wins	Squared error Combined Losses
(1)	(2)	(3)	(4)	(5)
1	–	43,264	10,000	–
2	6,084	–	–	14,400
3	7,225	–	–	36,481
4	–	32,400	169	–
5	–	19,881	64	–
6	–	23,104	4	–
7	–	14,641	3,600	–
8	324	–	–	900
9	1,521	–	–	3,364
10	–	17,424	17,161	–
Sum	15,154	150,714	30,998	55,145
Conservations	4	6	6	4
MSE	3,788	25,119	5,166	13,786

Table IV. *Forecast Accuracy Evaluation*

		10-period holdout sample					
		Model A		Model B		Combined	
Period	Actual	Forecast	Square error	Forecast	Square error	Forecast	Square error
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
5	2,064	2,125	3,721	2,272	43,264	2,164	10,000
6	2,386	2,106	78,400	2,464	6,084	2,266	14,400
7	2,432	2,116	99,856	2,517	7,225	2,241	36,481
8	2,542	2,386	24,336	2,722	32,400	2,529	169
9	2,663	2,513	22,500	2,804	19,881	2,655	64
10	2,767	2,597	28,900	2,919	23,104	2,765	4
11	2,896	2,728	28,224	3,017	14,641	2,836	3,600
12	2,014	2,849	2,722	3,032	324	2,984	900
13	3,146	2,964	33,124	3,107	1,521	3,088	3,364
14	3,015	3,135	14,400	3,147	17,424	3,146	17,161
Mean square error			36,069		16,587		8,614

Model Performance

An examination of the individual quarters, shows that the combined forecast outperforms model A in 8 out of 10 quarters, and is superior to model B in 6 of 10 quarters. Closer examination of the four “wins” by model B does not weaken the argument in favour of the combined forecast.

Looking at Table V, in which the squared errors of the individual and combined forecasts are presented, we see that on the four occasions that model B outperforms the combined forecasts (column 2 in Table V), the latter had an MSE of 13,786 which is approximately 3.5 times the MSE of the winning model B (3,788).

Alternatively, on the six occasions that the combined forecasts were more accurate than model B, the latter had an MSE of 25,119 which is approximately five times that of the combined forecast (5,166).

Hence, even when the combined forecast is less accurate than one of the other models from which it is derived, the chances are that its error will not be as great as that produced by any of the models that may outperform it on some occasions.

In summary, it should be noted that combining forecasts offers a solution to the choice among competing models. It warrants experimentation as a parallel system to determine the best weighting scheme and the number of periods to include in the “memory”.

References and Further Reading

Ang, J., Chua, J., and Fatemi, A., “How to Use AKAIKE Criterion to Automatically Specify and Construct Box-Jenkins Models”, *Journal of Business Forecasting*, Spring 1982, pp. 38-42.

Brandon, C., Fritz, R. and Xander, J., “Combining Time Series and Econometric Forecasts of Tourism Activity”, *Annuals of Tourism Research*, Vol. II, 1984, pp. 219-29.

Granger, C.W. and Newbold, P., *Forecasting Econometric Time Series*, Academic Press, New York, NY, 1977.

Jain, C.L., “How to Make One Good Forecast by Merging Two Fair Forecasts”, *Journal of Business Forecasting*, Spring 1984, pp. 26-8.

Makridakis, S. and Winkler, R., “Averages of Forecasts: Some Empirical Results”, *Management Science*, September 1983, pp. 987-96.

Conway Lackman is an Associate Professor in the Palumbo School of Business at Duquesne University, Pittsburgh, Pennysylvania, USA. Charles Brandon is Associate Professor at Rollins College, Winter Park, Florida, USA.