OMEGA Int. J. of Mgmt Sci., Vol. 14, No. 2, pp. 93–98, 1986
Printed in Great Britain. All rights reserved

0305-0483/86 $3.00 + 0.00
Copyright © 1986 Pergamon Press Ltd

A Pragmatic View of Accuracy Measurement in Forecasting

BENITO E FLORES

Texas A & M University, USA

(Received November 1984; in revised form July 1985)

Accuracy measurement in forecasting is always a subject of debate because of its importance. An adequate metric is necessary to properly select a forecasting method for a specific application. Competitions to determine the best method have helped the practitioner. The criteria for selection have not received as much attention. Of the two kinds of measurement statistics—relative and absolute—the former may present problems for the user if zeros or near zero values appear. This is more a practitioner problem because artificially generated time series do not usually have zeros. The relative and absolute measures are discussed and a solution for the existence of zeros in the data is given. If symmetry of the errors is a problem solutions are discussed. Managers will select the metric depending on the application and their management style. Once the metric has been selected the decision as to which forecasting method to select in a given situation becomes a less difficult problem.

INTRODUCTION

DECIDING which method to use in order to forecast the demand for a product is one of the many decisions that management has to make. If the product is considered important because of its high unit value or because of the volume utilized then the decision becomes critical. This applies to all items that constitute the important few (vs the trivial many). The forecast of these items will usually be watched more closely by management. For these, the selection process requires a more thorough analysis.

To select a forecasting method, authors such as Makridakis *et al.* [1] suggest that there are various criteria which can be used to help make the decision. They mention:

—the pattern of the data
—the type of series
—the time horizon
—the ease of application
—the cost
—the accuracy of the forecast

The first three are more related to the historical information available than to the type of demand or sales. The other three are more managerial in nature.

The ease of application includes among its components ease of understanding. If managers do not understand the forecasting method, chances are they will not use it. One of the reasons is that managers, in turn, may have to explain it to their supervisors.

The cost components that constitute the total cost function can usually be defined explicitly. The forecasting error cost, though, is far more difficult. The explicit definition of a loss cost function for a particular application is not easy to obtain, one of the reasons being that the function may be unique to the company and its cost structure. The cost is related to the accuracy measure as will be discussed afterwards.

The desire to have an accurate forecast is implicit. The definition of which accuracy measure to use is another story.

Witness the M-Competition [2]. In it five measures of accuracy were used: Mean Average Percentage Error (MAPE), Mean Square Error (MSE), Average Ranking (AR), Medians of Absolute Percentage Errors (Md), and Percentage Better (PB). As is to be expected, there is no total congruence in the results of their use. Which accuracy measure should be used by management to select a forecasting method?

Some of the results concerning the forecasting methods used in the M-Competition are becoming very useful for practitioners. In the accuracy measurement area, there are no corresponding conclusions. For instance, in the commentaries [3] to the competition, things did not get simpler. More accuracy measures are suggested. Percentage of Total Change, Coefficient of Variation, Geometric Mean Square, among others, are suggested by different authors. So, what accuracy measures should be used and what are some of the problems associated with their use? This paper takes a look at both issues from a managerial standpoint.

To provide a framework for the analysis, the accuracy measures are divided using the format of Makridakis [1] into two groups: Standard and Relative measures. Both types are presented and discussed below.

In some of the measures, like the MAPE, the user of this measure can get a very good feel of the performance of a forecasting method but sometimes there are problems with its use. One of them, the occurrence of zero or near zero values in the time series, will be discussed in detail below. The question of symmetry for MAPE is also commented on.

ACCURACY MEASURES

The following is a collection of the more common accuracy measures:

The error in period i is defined as the difference between the actual datum and the forecast.

$$E_i = X_i - F_i \tag{1}$$

Since this is usually done for n time periods then there will be n error terms. The descriptive statistics can be:

$$ME = \sum_{i=1}^{n} E_i/n \tag{2}$$

Mean Error (Bias).

$$MAD = \sum_{i=1}^{n} |E_i|/n \tag{3}$$

Mean Absolute Deviation

$$MSE = \sum_{i=1}^{n} E_i^2/n \tag{4}$$

Mean of Squared Errors

$$RMSE = \left[\sum_{i=1}^{n} E_i^2/n\right]^{1/2} \tag{5}$$

Root Mean Squared Error

Some relative measures can also be defined:

$$PE_i = [(X_i - F_i)/X_i] * (100) \tag{6}$$

Percentage Error

$$MPE = \sum_{i=1}^{n} PE_i/n * 100 \tag{7}$$

Mean Percentage Error

$$MAPE = \sum_{i=1}^{n} |PE_i|/n * 100 \tag{8}$$

Mean Absolute Percentage Error

There are, as has been mentioned before, many other accuracy measures that are used.

DISCUSSION

Absolute measures

The measures that were described above can be discussed under different formats. Here, some characteristics will be described and then a managerial perspective will be commented on.

Absolute errors can be difficult to use by themselves. If a forecast error is 100 units, the seriousness of the error cannot be comprehended unless the level of the series is also given.

For instance, the mean is a measure that should not be used by itself because it does not provide a measure of the error variance. Its value should (but does not always) hover around zero. Ideally it should be equal to zero. The mean, as is well known, will not tell us anything about the dispersion of the errors.

From a managerial standpoint though, means are quantities that are handled frequently. Managers are used to handling them. Means are defined in the same units as the data. They do provide information concerning the fact that bias may be present. This information is useful because forecasts can be adjusted for bias. For instance, if what is being forecasted is sales, this statistic can be very valuable if sales quota and its corresponding salary (and/or bonus) is being formulated. Conservative salesmen would like the forecasts to be biased downward.

The mean absolute deviation (MAD) is a metric defined also in the same units as the historical data. Because it uses absolute values, it does not have the problem of the mean of having errors cancelling themselves out if of opposite signs. It does not, however, distinguish between variance and bias.

The MAD is not too difficult to explain to management because it measures the value of a typical error regardless of sign. If the cost of overforecasting or underforecasting can be defined, then this measure assumes that the costs are directly proportional to the number of units the forecast is in error. The loss function, as in the case of the mean, is symmetrical.

Another thing that can make the use of the MAD more attractive is its simplicity of calculation. The values of MAD can be approximated by using a smoothing relationship:

$$MAD_i = g * E_i + (1 - g) * MAD_{i-1} \tag{9}$$

where g is a constant between 0 and 1 usually taking low values such as 0.1 as mentioned by Trigg and Leach [4].

This relationship has the advantage of requiring less memory space in a computer and is faster for computational purposes.

The squared error is a very desirable way to measure deviation. In this measure, errors are weighted according to size. Large errors will be given greater weight.

The MSE, as reported by Carbone and Armstrong [5], is one of the measures more commonly used by practitioners. The measure, because it squares the error terms, will not have the problems of the mean. Because it is scale variant, care must be exercised in its use.

If the square root of the MSE is calculated, the metric will have the same units as the data. Outliers have a strong effect on both measures.

The MSE has the inherent difficulty of not being easy to explain to management. Nevertheless, if the errors are assumed to be normally distributed, the RMSE and the MAD are related as reported by Brown [6] by the function

$$RMSE = 0.8 * MAD \tag{10}$$

and therefore switching back and forth between RMSE and MAD can be performed rather handily. Brown also discussed the relationship between RMSE and MAD for other distributions.

The use of RMSE is justified if the cost function increases as the square of the error. This type of function has been exemplified by several authors and will be described later on in this paper.

Relative measures

The percentage error can give some idea of the quality of a forecast. Nevertheless, for a series 5% error may be bad and for others 12% may be very good.

The MAPE is similar to MAD except that it is dimensionless. Problems do arise in its use if the values are zero or close to zero. This will be discussed later. Its error function is non-symmetrical as it is biased towards estimates below actual values. If a forecast is equal to zero, the percentage error can never be off more than 100%. Errors on the high side do not have a limit.

The MAPE can give us a very good indication of how well the forecasting method is performing. Because it is dimensionless, it facilitates communication to management. For instance, suppose a value of 10% for MAPE is calculated for a time series. This value can give us a very good idea of the quality of the forecasting method. From it, a company can decide whether it can live with this kind of precision. This metric implies that the cost function for the error can be defined in terms of percentages rather than units.

The comparison between forecasting methods is quite simple with this measure. It can also be used to compare results across time series.

Other statistics have been developed that perform more complex comparisons. Theil's U statistic [7] compares the forecast to the results that would have been obtained if the Naive method had been used for forecasting purposes rather than the one actually used. This in effect creates an immediate comparison of forecasting methods but has, unfortunately, the difficulty of being very hard to explain to management. Perhaps for this reason McLaughlin [8] refers to his measure as the batting average. This measure is similar to Theil's and could be calculated directly or from the U statistic (Makridakis [1]).

Many accuracy measures can be similarly discussed. Could it be that practitioners like Muir of American Software [9] have a point in saying that managers should use whatever measure makes sense to them?

Zero values

The relative measures, because they are defined as a ratio, have an inherent difficulty in handling low and zero values in the denominator. This is not the case with the absolute measures.

One of the things that is scarcely commented on in the literature is the fairly common event of having zero sales for a particular time period. Perhaps done by design, or chance, this is not an occurrence in artificially generated time series. Sometimes this may be done on purpose to avoid problems. Nonetheless, in practice they do occur and practitioners, for instance Smith [10] and Brown [11], give examples of the occurrence of the event.

Armstrong [12] touches on the subject of the existence of zeros in the data by stating that problems may arise if values close to zero are encountered for actual results. So does Steece [13] when he mentions that problems may arise when the actual data is near zero. The fact is that if the values of the demand or sales (X) are zero, the use of the relative accuracy formulas will yield indeterminate values and the helpfulness of the measure ceases to exist.

The existence of zeros may depend on the definition of the time interval where the demand or sales occur. There may be some time periods, i.e. days, weeks or months when there are no sales for a specific product. Another problem may be the near zero values. This event can occur when the demand for a product may be declining.

Equation (6) should be defined only for values of demand that are greater than zero. Negative values of X would be highly unusual (perhaps, in the case of sales forecasting, actual values could become negative if a sufficient number of clients decided to return the product). If the data are near zero, then the percentage errors become very large and distort the results so badly that comparison between forecasting methods becomes very difficult. This will also make comparison across time series nearly impossible.

The end result of this is that PE, MPE and MAPE (as is the Theil statistic), when values of zero occur, cannot be used for accuracy measurement since the value of the error would become infinite (indeterminate). Therefore, evaluation of a forecasting method or comparison between two methods would be impossible to make. In the case of near zero values the comparison would be very difficult.

Practitioners like Smith have resorted to clever means of going around the problem of having zeros in the data. He just defines the error to be 100%. To formalize this procedure a simple example is described below.

For instance, assume $F = 10$ and $X = 0$.

$PE = (0\text{–}10)/0 = \text{indeterminate } (\infty)$

Truly the error is the total amount, i.e. 100%. Now, if the expression for equation (6) was written as

$$PE = (X - F)/X * (100) \quad \text{for all } X > 0 \quad (11)$$

$$PE = (F - X)/F * (100) \quad \text{for } X = 0 \quad (12)$$

then the PE would be

$$PE = (10 - 0)/10 * (100) = 100\%$$

which is the error in percentage points.

This can help the user circumvent the problem of having zeros in the data and the use of relative measures.

Symmetry

Another disadvantage of MAPE is that the error function is not symmetrical, that is, as was mentioned before, the method has a bias for estimates that are below the actual data.

Armstrong [12] suggests a solution by using a modified MAPE

$$\overline{MAPE} = \sum_{i=1}^{n} |(X_i - F_i)|/[0.5 * (X_i + F_i)]/n * 100 \quad (13)$$

This change makes the statistic a symmetrical measure. If an actual value is 50, a forecast of 25 is as good or bad as a forecast of 100. The limits are now from zero for a perfect forecast to 200 for an infinitely bad forecast. On the negative side, this measure is more difficult to explain to management. What do the average of the forecast and the actual data point mean? What about the ratio of the error to this value?

Perhaps due to the fact that the author is strongly influenced by the metric system, rather than using Armstrong's adjusted MAPE ($\overline{\text{MAPE}}$), a similar measure could be suggested:

$$MAPE = \sum_{i=1}^{n} |(F_i - X_i)|/(F_i + X_i)/n * 100 \quad (14)$$

which would have a limit of 100 corresponding to Armstrong's 200. An added attractiveness of both formulas is that now both the forecast and the actual data point must be zero for the ratio to become indeterminate. Both formulas, from the standpoint of ease of understanding for management, remain a problem.

Some managerial examples

Can the decision of which statistic to use be tied with the managerial style? As mentioned by Gardner [14], is the ability of the forecasting method to avoid large errors more important than their central tendency? Managers whose style is conservative and who adhere to a minimization of regret criterion (Savage [15]) should use MSE or RMSE.

Is the business manager like the baseball manager (see comment about McLaughlin's measure) who will play basing his decisions on the law of averages? If this is the case, metrics such as MAD or MAPE would be more towards his style.

In some functional areas, like in production planning, large errors can create serious problems. Overforecasting can lead a company to add temporary capacity via subcontracting but worst of all, permanent capacity can be added if equipment is purchased. Because of this, practitioners use the MSE more frequently than any other measure. This coincides with the results reported by Carbone and Armstrong [6]. In the same paper, the authors report that practitioners use other criteria. For instance, the results show that users more frequently desire ease of interpretation as a criterion to use. Also, ahead of MSE is cost/time. The next two statistics used by practitioners are MAD and MAPE. Some other measures are mentioned but their use by practitioners is much lower compared to the ones previously mentioned.

In inventory control, accuracy measures that can be tied to the variability of the forecast errors will be used more frequently. This goes back to the need to calculate safety stocks in order to cover the variability of demand during the lead time period. In this case MSE or RMSE may be more desirable.

Care must be exercised in the use of the loss function. For instance, the function which Makridakis [16] mentions is asymmetric, but the question of whether over- or underestimation is more or less desirable depends on whether the inventory cost or the stockout cost is more expensive. This cannot be answered *a priori*.

Sales forecasting in financial management terms, according to Brigham [17], is the expected value of a probability distribution of possible levels of sales. An accuracy measure that minimizes the variability of the forecast errors would be preferable. This can be MSE or any other related metric.

In sum, every managerial application of forecasting will require a different metric. The particular use should determine the metric to be applied.

CONCLUSIONS

There are no easy answers to the question of accuracy of forecasts. The relative measures can be of significant help in measuring and contrasting forecasting methodologies. One of the caveats in the use of relative measures is when the actual values are zero or near zero. If this is the case, the proposed way of calculating the forecast error can solve the problem. If symmetry or the presence of zeros becomes a problem, any of the adjusted MAPE formulations can be of help.

The managerial application will help to determine which metric to use in the selection of the forecasting method. This is very clearly the case in inventory control. In other situations, the managerial style can define the type of metric to use. Managers by nature tend to be conservative. Maybe, because of this, measures that tend to avoid large errors (which managers will regret most) will be preferred. This is what has been found out in surveys. In some cases, measures like MSE can be connected to other ones like MAD and make managers more comfortable with their use.

The manager will select the metric, either because of the application, the managerial style or both. Once the metric is selected, the decision as to which forecasting method to use will become a less difficult problem.

REFERENCES

1. Armstrong SJ (1978) *Long Range Forecasting*. Wiley, New York, USA.
2. Armstrong JS and Lusk EJ (1983) Commentary on the Makridakis Time Series Competition. *J. Forecasting* **2** (3), 259–311.
3. Brigham EF (1982) *Financial Management: Theory and Practice*. Dryden Press, Chicago, IL, USA.
4. Brown RG (1962) *Smoothing, Forecasting and Prediction of Discrete Time Series*. Prentice-Hall, Englewood Cliffs, NJ, USA.
5. Brown RG (1967) *Decision Rules for Inventory: Management*. Holt, Reinhart and Winston, New York, USA.
6. Carbone R and Armstrong JS (1982) Evaluation of extrapolative forecasting methods: results of a survey

of academicians and practitioners. *J. Forecasting* **1** (2), 215–217.
7. Gardner ES (1983) The trade-offs in choosing a time series method, *J. Forecasting*, **2** (3), 265–267.
8. McLaughlin RL (1972) Measuring the accuracy of forecasts. *Busin. Econ.* **7** (3), 27–35.
9. Makridakis S et al. (1982) Accuracy of extrapolation (time series) methods: results of a forecasting competition. *J. Forecasting* **1** (2), 111–153.
10. Makridakis S (1983) Empirical evidence vs personal experience. *J. Forecasting* **2** (3), 295–306.
11. Makridakis S et al. (1983) *Forecasting: Methods and Applications*. Wiley, New York, USA.
12. Muir JW (1983) Problems in sales forecasting needing pragmatic solutions. Paper presented at the Third International Symposium on Forecasting, Philadelphia, PA, USA.
13. Savage LJ (1951) The theory of statistical decision. *J. Am. statist. Ass.* **46** (253), 55–67.
14. Smith BT (1978) *Focus Forecasting Computer Techniques for Inventory Control.* CBI Publishers, Boston, MA, USA.
15. Steece B (1982) In *Handbook of Forecasting: A Manager's Guide* (Edited by Makridakis S and Wheelwright S), Chap. 27. Wiley, New York, USA.
16. Theil H (1966) *Applied Economic Forecasting*. North Holland, Amsterdam.
17. Trigg DW and Leach AG (1967) Exponential smoothing with an adaptive response rate. *Opl Res. Q.* **18** (1), 53–64.

ADDRESS FOR CORRESPONDENCE: *Benito E Flores, Business Analysis Department, College of Business Administration, Texas A & M University, College Station, TX 77843, USA.*