INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Some Aspects of Forecasting Methods

C.Narayana[1], G.Y. Mythili[2], J. Prabhakara Naik[3], K.Vasu[4], G. Mokesh Rayalu[5]

[1]Assistant professor, Department of Mathematics, Sriharsha Institute of P.G. Studies, Nellore

[2]Assistant Professor, Department of Mathematics, School of Advanced sciences, VIT, Vellore

[3]Lecturer in Statistics, SSBN Degree & PG College (Autonomous), Anantapur

[4]Assistant Professor, Vidya Educational Institutions, Yadamari, Chittoor

[5]Assistant Professor, Department of Mathematics, School of Advanced sciences, VIT, Vellore

Abstract: Forecasting is an important field of statistics that spans in to many fields including business and industry, government, economics, environmental sciences, medicine, social science, politics, and finance. Forecasting problems are generally classified as short-term, medium-term, and long-term models. The short-term forecasting problems involve predicting events only a few time periods (days, weeks, months) into the future. Medium-term forecasts extend from one to two years into the future, and long-term forecasting problems can extend beyond that by many years. Short- and medium-term forecasts are required for activities that range from operations management to budgeting and selecting new research and development projects. Long-term forecasts impact issues such as strategic planning. Short- and medium-term forecasting is typically based on identifying, modeling, and extrapolating the patterns found in historical data. Because these historical data usually exhibit inertia and do not change dramatically very quickly, statistical methods are very useful for short- and medium-term forecasting.

I. INTRODUCTION

In the literature, we can find large number of forecasting methods which are classified according to various criteria like time, nature of process etc. In the following section, we described some of the important classifications of forecasting methods.In all the forecasting projects, we can find different forecasting horizons depending on the fitted model and history of data available for prediction. The classification of forecasting method with respect to its forecast duration is as follows.

A. Short-term forecasting method:

The forecasting method in which the prediction period is lies between one hour and one week is considered as short term forecasting method. This type of projects are generally found in clinical trials where the forecast of increase of bacteria is done for next few minutes as they are growing fastly with in a short time. The short term forecasts are always utilizes the growth models where exponential growths are expected for study variables. The rate of error occurrence is also small and tends to rise if you increase the forecasting period to more than a week. In the finance region, we can find daily forecasts happening for stock market indices which fall under this method and ARCH and GARCH models are proven to be the best models for this type of scenarios.

B. Medium term forecasting:

Medium term forecasting method usually applied for the situation where prediction period lies between a week and less than year. Many of the forecasting methods may be classified as mid-term forecasting only because weekly and monthly predictions are common requirement of the projects.

The most of the time series methods as well as regression methods are getting used to find the predictions with mid- range time period. Sales and Demand forecasting are better examples of this category where they are estimated for year or less. The prediction error in this type of forecasting is considerably high and user has to depends on some model selection criteria for selection of best model.

C. Long term forecasting method:

In this method, the prediction period is more than a year and almost all demographic related surveys may fall in to this category. This method involves some strategic decision taking based on forecasts. As general trend equation may suffices the needs of this forecasting and error also minimum in this method. Acquisitions and merges, product changes, process changes are better examples of long term forecasting projects for which span of time is more than year.

D. Explanatory Models For Forecasting

The forecasting models in which forecast is expressed as a function of certain number of factors that influence its outcome is called explanatory models for forecasting. Thus, in this model, the forecasts need not be time dependent. In addition to this, an explanatory model that relates output to input facilitates better understanding of the situation and allows experimentation with different combination of input that effect the output or forecasts. In this way, explanatory models can be geared towards intervention, influencing the future through decisions made today. More accurate forecasts also result as the influence of explanatory variables on the output can be estimated. Here, the forecaster must decide on how many variables to include and what is the functional form of the forecasting model before the project actually begins in the environment.

E. Simple regression model

The regression method establishes the relationship of a variable, called response, with one or more predictor variables. If only one predictor is studied for its relationship with the response variable, then it is a simple regression analysis. Consider the following simple linear regression model for investigating a relationship of a response variable Y with a predictor X as

$$Y = a + bX + e$$

The random error e is assumed to be normally distributed with mean 0 and variance σ^2. Of this model, a, b and σ^2 are unknown constants, which are to be estimated. If it turns out that b = 0, then X is not a significant predictor of Y.

To estimate the value of a and b, we can apply the least square estimation which arrives at the following expressions

$$b = \frac{SP}{SS_x} = \frac{\Sigma(X - \overline{X})(Y - \overline{Y})}{\Sigma(X - \overline{X})} \text{ and } a = \overline{Y} - b\overline{X}$$

This method of finding the best fitting regression line is called the least squares method because. It minimizes the sum of the squared errors in prediction. In other words, if we let Y'= the predicted value of Y, then $\Sigma(Y - Y')^2$ = a minimum .The quantity (Y - Y') is the amount of error in prediction. It is the difference between the actual Y score and the predicted Y score. We square this term before summing, because the sum of the errors in prediction equals zero. Some key points about linear regression is

1) The relationship between X and Y must be linear. If the relationship is nonlinear, prediction will not be very accurate
2) We need to have representative sample to construct regression equation and use the equation to predict Y scores when only X values are given.
3) A regression equation in general is fitted for a range of values of input variables. If the forecasting is done for the variable values which are out of range of linear property, then the predictions may not be accurate.
4) Prediction is most accurate if the data have the property of homoscedasticity i.e., if the variability of the Y scores is constant at all points along the regression line.

F. Multiple liner regression

In multiple regression, there is one variable to be predicted but there are several explanatory variables. The general form of multiple regression is

$$Y = b_0 + b_1 * X_1 + b_2 * X_2 + ... + b_k * X_k + e$$

Just like simple linear regression, multiple linear regression also had some assumptions about X_i's and e's.

The explanatory variables X_1, X_2, …X_k are random and independent to each other.

The error term e's are un correlated with each other

The error terms are normally distributed with mean zero and variance $\hat{\sigma}_e^2$

To the estimated values of the coefficients, one can apply the method of least squares and get the unbiased estimates of the coefficients. Similarly, we can apply the parametric t-test to find the significance of the each individual variables

1) *Forecasting based on logistic regression:* Logistic regression is used to analyze relationships between a dichotomous dependent variable and metric or dichotomous independent variables. In the logistic regression, the response variable is dichotomous categorical outcomes (*e.g.,* dead vs. alive, cancer vs. none). Logistic and linear regression are both based on many of the same assumptions and theory.

As the outcome is dichotomous, logistic regression has great importance in pharmaceutical field where the outcome is dichotomous. Since the outcome is dichotomous, predicting unit change has little or no meaning. As an alternative to modeling the value of the outcome, logistic regression focuses instead upon the relative probability (odds) of obtaining a given result category. As it turns out the natural logarithm of the odds is linear across most of its range, allowing us to continue using many of the methods developed for linear models.

The general model of logistic regression can be expressed as follows:

$$\text{Ln} [p / (1-p)] = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 \ldots b_k x_k + e$$

Where represents the probability of an event (e.g., death), b_0 is the y-intercept, and x_1 to x_k represent the independent variables included in the model. As with the linear model, each independent variable's association with the outcome (log odds) is indicated by the coefficients b_1 to b_k. Again, an error term is included to account for differences between the observed outcome values and those predicted by the model

2) *Estimation and model performance:* Logistic regression uses maximum-likelihood estimation to compute the coefficients for the logistic regression equation. This method finds attempts to find coefficients that match the breakdown of cases on the dependent variable. The overall measure of how will the model fits is given by the likelihood value, which is similar to the residual or error sum of squares value for multiple regression. A model that fits the data well will have a small likelihood value. A perfect model would have a likelihood value of zero. The overall test of relationship among the independent variables and groups defined by the dependent is based on the reduction in the likelihood values for a model which does not contain any independent variables and the model that contains the independent variables.

3) *Forecasting based on Decision tree :*We have observed that in recent years have seen an increasing cross-fertilization between the fields of decision analysis and forecasting. Decision-analytic models often require forecasts as inputs, and aspects of the Bayesian decision-theoretic framework underlying decision analysis have proved useful to forecasting .Decision trees are used to select the best course of action in situations where you face uncertainty. Many business decisions fall into this category. For example, a manufacturer must decide how much inventory to build before knowing precisely what demand will be. A litigant must choose between accepting an out-of-court settlement or risking a trial. A speculator must decide to buy an asset before knowing if it can be sold for a profit. In all of these cases, the decision-maker faces an unknown that seems to make it impossible to choose the right option with any certainty. Although the decision-maker does not know what the outcome of the unknown will be, he or she generally has some knowledge about what the possible outcomes are and how likely each is to occur. This information can be used to select the option that is most likely to yield favorable results. Decision trees make this type of analysis easy to apply.

G. *Steps involved in decision tree*

1) List out the choices you will make by defining the objective and constraints on the situation. Typically, you will want to maximize some objective such as profits or minimize something undesirable, such as waste.
2) List the possible actions between which you need to choose and list the possible events that may happen after you have acted.
3) Draw up a Payoff matrix to determine the value of each decision, should any of the possible subsequent events occur. This will require a common unit of measurement, such as time, return on investment, etc.
4) Draw the basic Decision Tree, showing the actions, events and outcomes from the previous steps, showing actions emanating from squares and events from circles.
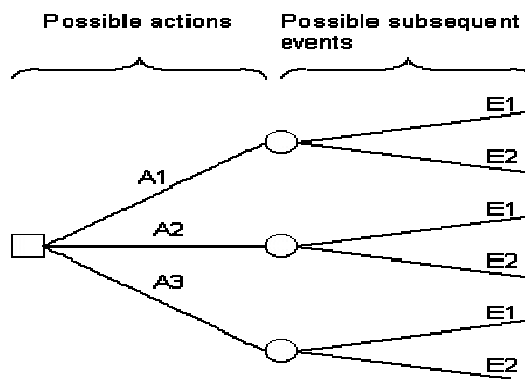
Model of decision tree

Identify the prioritization strategy, which may place different emphasis on things such as risk and potential return.
Calculate the value of each alternative action, according to the strategy being used. This may, for some strategies, require that you identify the probability of the alternative events or consequences.

II. CRITERIA FOR MODEL SELECTION IN FORECASTING

Model selection is the task of selecting an appropriate model from the list of existing models which is having proven superior quality and performance. Criteria here refers to any method or tool to find the best models out of group of models. As it is the key part of any forecasting project, many approaches have been proposed over the years for dealing with this key issue and in the following sections, we discussing some of them briefly.

A. R square and Adju

Ted R square: The coefficient of determination, known also as the R^2, is a common measure in regression analysis. In general, R^2 can be defined as the ratio of proportion of variation in dependent variable(Y) explained by independent variable (X).The usual representation of R^2 is given by

$$R^2 = \frac{SST - SSE}{SSE} = \frac{\Sigma(y - \bar{y})^2 - \Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2}$$

Each additional X variable added to the model increases R^2. Since, R^2 can be made larger, simply by adding more predictor variables to the model, a modification of R^2 has been proposed. This adjusted R^2 does not automatically increase when new predictor variables are added to the model. In fact, the adjusted R^2 may actually decrease, because the decrease in SSE may be more than offset by the corresponding decrease in the error df. To avoid the drawback, there is slight correction proposed in R^2 which we call as adjusted R^2 and the formulation of it is as follows.

$$Adj\,R^2 = 1 - \frac{(1 - R^2)(N - 1)}{(N - K)}$$

Where N=total number of cases and K=No of independent variable +1
The adjusted R^2 values gives better picture of model efficiency in the econometric forecasting projects.

B. Mallows Cp criterion

C.L. Mallows developed a method to find adequate models by plotting a special statistic against the number of variables+1.

$$C_p = \frac{SS_{res}}{MS_{res}} - N + 2p,$$

where SS_{res} = residual sum of squares for the model with p-1 variables, MS_{res} = residual mean square when using all available variables, N = number of observations, and p = number of variables used for the model plus one. The general procedure to find an adequate model by means of the C_p statistic is to calculate C_p for all possible combinations of variables and the C_p values against p. The model with the lowest C_p value approximately equal to p is the most "adequate" model.
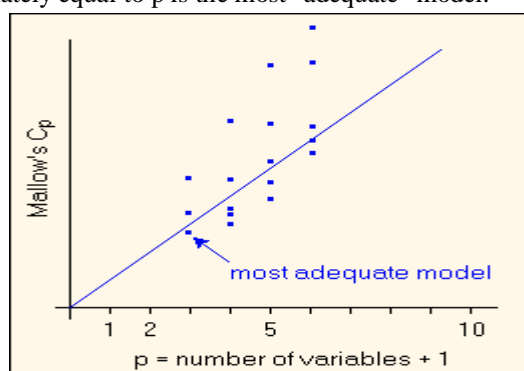
Cp value determination

As we can see that at the lowest value of Cp, the best model is selected for the given scenario. However, the construction of the Cp criterion is based on the assumption that s^2, the MSE from fitting the full model, is an unbiased estimate of σ^2. If the full model happens to contain a large number of parameters (β's) that are possibly not significantly different from zero, then this estimate of σ^2 will be large. This is because the variables that are not contributing to significantly decreasing the SSE are still counted toward the degrees of freedom when computing the MSE in the full model. If this is the case, Cp will not be a suitable criterion to use for determining a good model.

B. Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) is a way of selecting a model from a set of models. The chosen model is the one that minimizes the Kullback-Leibler distance between the model and the truth. The formulation of AIC is

$$AIC(M) = -2 \log L(M) + 2 \cdot p(M)$$

Where L(M) is the likelihood function of the parameters in model M evaluated at the MLE (Maximum Likelihood Estimators).AIC approach is valid for any families of statistical models such as: ANOVA (can use it for multiple comparisons), time series models, categorical data models, multivariate model etc.

Model selection procedure using AIC is

1) Decide, on scientific grounds, upon a small set of models to be compared (not a fishing expedition), that is, narrow down the number of variables to be considered for inclusion in a multiple regression
2) Fit all possible models in the set (ML estimates of parameters in all the models under consideration)
3) For each model, calculate its AIC (Gerc the number of parameters in a multiple regression is k + 2)
4) Pick the model with the smallest AIC. That is the model in the suite with the best overall statistical properties and parameter balance Akaike'srule.Also, it should be noted that too many models (or variables) will not hold good for AIC properties and using AIC is equivalent to pickingby "cross-validation" measure only.

The Bayesian information Criterion (BIC) is a criterion for selecting a best model among formal econometric models. This measure takes into account both the statistical goodness of fit and the number of parameters that have to be estimated to achieve this particular degree of fit, by imposing a penalty for increasing the number of parameters.

$$BIC = n.\ln\left(\sigma_e^{\hat{2}}\right) + k.\ln(n)$$

Where $\sigma_e^{\hat{2}}$ is the error variance and in this case is defined as

$$\sigma_e^{\hat{2}} = \frac{1}{n}\sum_{i=1}^{n}\left(xi - \bar{x}\right)^2$$

For any model with lower value of BIC is the one to be preferred as best model. The BIC generally penalizes extra parameters, though it depends on the size of n and relative magnitude of *n* and *k*. It is important to keep in mind that the BIC can be used to compare estimated models only when the numerical values of the dependent variable are identical for all estimates being compared.

C. Goodness Of Fit Tests For Parameters Of Forecasting Models

Significance tests are good tools to find whether the parameter or coefficient influences the dependent variable or not. These tests can only be carried out if it can be assumed that the random error terms, $\in_i$, are normally and independently distributed with a mean of zero and variance of σ^2.

D. Test for Significance of Regression

The hypothesis test for significance of regression is usually performed by using the analysis of variance. The test is used to test if a linear statistical relationship exists between the response variable and at least one of the predictor variables. The statements for the hypotheses are:

$H_0 : \beta_1 = \beta_2 = \ldots.. = \beta_k = 0$

$H_1 : \beta_j \neq 0$ for at least one j

The test for H_0 is carried out using the following F statistic:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

$$F_0 = \frac{MS_R}{MS_E}$$

where MS_R is the regression mean square and MS_E is the error mean square. If the null hypothesis, H_0, is true then the statistic F_0 follows the F distribution with k degrees of freedom in the numerator and n-(k+1) degrees of freedom in the denominator. The null hypothesis, H_0, is rejected if the calculated statistic, F_0, is such that:

$$F_0 > f_{a,k,n-(k+1)}$$

E. Calculation of the Statistic F_o

To obtain the test statistic F_o, we need to calculate MS_R and MS_E in the beginning. The mean sums of squares are obtained by dividing the sum of squares by their degrees of freedom. For example, the total mean square, MS_T, is obtained as follows:

$$MS_T = \frac{SS_T}{dof(SS_T)}$$

where SS_R is the total sum of squares and dof(SS_T) is the number of degrees of freedom associated with SS_R .

In multiple linear regression, the following equation is used to calculate SS_R :

$$SS_T = y' \left[I - \left(\frac{1}{n}\right) J \right] y$$

where n is the total number of observations, y is the vector of observations, I is the identity matrix of order n and J represents an nXn square matrix of ones. The number of degrees of freedom associated with SS_R,   dof (SS_R), is (n-1).

The regression mean square, MS_R, is obtained by dividing the regression sum of squares, SS_r , by the respective degrees of freedom,   dof (SS_r) as follows:

$$MS_R = \frac{SS_R}{dof(SS_R)}$$

The regression sum of squares, SS_R, is calculated using the following equation:

$$SS_R = y' \left[H - \left(\frac{1}{n}\right) J \right] y$$

where n is the total number of observations, y is the vector of observations, H is the hat matrix and J represents an nXn square matrix of ones. The number of degrees of freedom associated with SS_r, dof(SS_E) , is k, where k is the number of predictor variables in the model.

The error mean square, MS_E, is obtained by dividing the error sum of squares, SS_E, by the respective degrees of freedom, dof (SS_E), as follows:

$$MS_E = \frac{SS_E}{dof(SS_E)}$$

The error sum of squares, SS_E, is calculated using the following equation:

$$SS_E = y' [I - H] y$$

where y is the vector of observations, I is the identity matrix of order n and H is the hat matrix. The number of degrees of freedom associated with SS_E, dof(SS_E), is n-(k+1), where n is the total number of observations and k is the number of predictor variables in the model. Knowing SS_E and dof(SS_E), the error mean square, MS_E, can be calculated. The error mean square is an estimate of the variance, σ^2, of the random error terms, $\in_i$.

$$\hat{\sigma}^2 = MS_E$$

F. Test for Individual Regression Coefficients ('t' Test)

To test significance of the individual regression coefficient, we will use the student 't' test. Adding a significant variable to a regression model makes the model more effective, while adding an unimportant variable may make the model worse. The hypothesis statements to test the significance of a particular regression coefficient, β_j, are:

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j \neq 0$$

The test statistic for this test is based on the t distribution:

$$To = \frac{\hat{\beta}_j}{se\left(\hat{\beta}_j\right)}$$

The analyst would fail to reject the null hypothesis if the test statistic lies in the acceptance Region:

$$-t_{\alpha/2,n-2} < T_0 < -t_{\alpha/2,n-2}$$

This test measures the contribution of a variable while the remaining variables are included in the model.

For the model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$, if the test is carried out for β_1, then the test will check the significance of including the variable x_1 in the model that contains x_2 and x_3 (i.e. the model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$). Hence the test is also referred to as partial or marginal test.

III. CONCLUSIONS

The problem identifying right objective for the given forecasting problem is the first step in any forecasting process. It involves deep understanding of how the forecasts will be used, who is the end user of the forecasts, and how the forecasting function fits within the organization. It is always advisable to spend considerable amount of time in finalizing these objective before starting with the actual task. In this paper we discuss about some aspects in fore casting methods

REFERENCES

[1] Azoff. E. M. (1994) "Neural network time series forecasting of financial market." John Wiley & Sons Ltd.
[2] Eitan Michael Azoff. (1993) Reducing Error in Neural Network Time Series Forecasting, Neural Computing & Applications, pp.240-247.
[3] Hair, Anderson, Tatham, Black, (1998) "Multivariate Data Analysis," Pearson Education press.
[4] Hamilton, J.D. (1994), Time Series Analysis, Princeton University Press.
[5] Mendenhall and Beaver, (1994) "Introduction to Probability and Statistics," Ninth Edition, International Thomson Publishing.
[6] Michael McInerney, Atam P. Dhawan. (1993) Use of Genetic Algorithms with Back-Propagation in training of Feed-Forward Neural Networks, Proc. of IEEE Int'l Joint Conference on Neural Networks, Vol. 1, pp. 203-208.
[7] SWALES, G.S. and Y.YOON, (1997) "Applying artificial neural networks to investment analysis," Financial Analysts Journal, 48(5).
[8] Tang, Almeida and Fishwick, Simulation, "Time series forecasting using neural networks vs. Box-Jenkins methodology," pp. 303-310.
[9] Thimm G. and E. Fiesler. (1997) High Order and Multilayer Perceptron Initialization, IEEE Transactions on Neural Networks, Vol. 8, No.2, pp.349-359.
[10] Toffler, A. et al., (1972). The Futurists. New York: Random House.
[11] W. Press et al., (1994) Numerical Recipes in C, The Art of Scientific Computing, 2nd Edition, Cambridge University Press.
[12] Walczak, S. (2001). An empirical analysis of data requirements for financial forecasting with neural networks. Journal of Management Information Systems, 17 (4), 203-222.
[13] Wasserman, P.D (1989). Neural Computing: Theory and Practice, Van Nostrand Reinhold: New York, 1989.
[14] Weatherford, L. R., T. W. Gentry, and B. Wilamowski (2003). "Neural network forecasting for airlines: A comparative analysis." Journal of Revenue and Pricing Management 1, no. 4. 319-331.

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)