iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XII **Month of publication:** December 2017

DOI:

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

Classification of Forecasting Methods Based On Application

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 methods always differs for each scenario and application area. For example we may not apply time series forecasting for a data where there is no time recorded variable is present. Each of application area has its own specific models and it has again its own criteria for selecting best model out of it. We have discussed some of the forecasting methods in widely used applications areas.
A forecasting model is defined as a functional relationship between dependent variable which is meant for forecasting and independent variables which are having some impact on forecasting variable. Every forecasting model has its own characteristics by its way of defining and nature of scenarios applicable for the forecasting. Some of the basic characteristics of any forecasting model can be outlined as follows.

I. INTRODUCTION

Functional form of relationship represents a mathematical equation which shows the kind of relationship between the response variable and explanatory variables in the current scenario and this form is decided based on several factors like
Type of movement of dependent variabl
Nature of independent variabl
Availability of data variable
No of total variables in the analysis
is the key characteristic of any forecasting model and the success of complete forecasting process since this functional form can be used further to forecast the variable of interest in the analysis. The variable for which the statistician wants to find the forecasted values is called dependent variable or response variable and it is again decided at the time requirement analysis only. Depending on the nature of this variable like binomial, categorical, continuous etc. the researcher can decide about the functional form of forecasting model.

A. Multiple equation forecasting method:

This is very subjective and specific method of forecasting. It is also known as the 'System of simultaneous equation models' or 'econometric model building'. This method is normally used in macro-level forecasting for the economy as a whole. The method is indeed very complicated since there were many dependent variables that interact with each other via series of equations. The principle advantage in this method is that the forecaster needs to estimate the future values of only the exogenous variables unlike the regression method where he has to predict the future values of all, endogenous and exogenous variables affecting the variable under forecast. The values of exogenous variables are easier to predict than those of the endogenous variables. However, such econometric models have limitations, similar to that of regression method.

B. Time series forecasting method

Arrangement of a study variable according to its time of occurrence is called time series. Time series are comprised of four separate components: trend component, cyclical component, seasonal component, and irregular component. These four components are viewed as providing specific values for the time series when combined. Time series modeling for the purpose of forecasting and control is important for analyses of production and business systems.

There are generally two types of model representation, corresponding to whether the time series observations are mutually independent or auto correlated. In the first category, the time series model is some function of time with a superimposed random error term, sometimes called the process noise. When time series observations are auto correlated the current observation is modeled as a linear combination of the past observations or the previous noise terms. This latter representation is commonly called the Box-Jenkins model or the autoregressive integrated-moving average (ARIMA) model (Box and Jenkins 1976). Selecting the model that best describes he underlying process is important for achieving good forecasting accuracy.

C. Composite methods of forecasting

Bayesian method forecasting is the best example for the composite method of forecasting. This family of methods combines statistical methodology with structured integration of human judgment: new evidence is used to update a statistical forecast, based on application of Bayes' theorem. These methods are good for highly seasonal data with short historyWe can find more applications to this method in pharmaceutical domain since we do not get mush historical data on the dependent variables.

D. Simulation modeling methods

Simulation methods involve using analogs to model complex systems. These analogs can take on several forms. An equation to predict an economic measure would be a mathematical analog. Mathematical analogs are of particular importance to futures research. They have been extremely successful in many forecasting applications, especially in the physical sciences. In the social sciences however, their accuracy is somewhat diminished. The possible drawback of this method is that these techniques often begin with an initial set of assumptions, and if these are incorrect, then the forecasts will reflect and amplify these errors. One of the most common mathematical analogs in societal growth is the S-curve. The model is based on the concept of the logistic or normal probability distribution. All processes experience exponential growth and reach an upper asymptotic limit. The advantage of the model is that it forces planners to take a long-term look at the future. Another common mathematical analog involves the use of multivariate statistical techniques. These techniques are used to model complex systems involving relationships between two or more variables. Multiple regression analysis is the most common technique.

E. Cross-impact matrix method

In the general applications, we can find that the relationships often exist between events and developments that are not revealed by univariate forecasting techniques. The cross-impact matrix method recognizes that the occurrence of an event can, in turn, affect the likelihoods of other events. The basic step here is to assign probabilities that reflect the likelihood of an event in the presence or absence of other events. The resultant inter-correlational structure can be used to examine the relationships of the components to each other, and within the overall system. The advantage of this technique is that it forces forecasters and policy-makers to look at the relationships between system components, rather than viewing any variable as working independently of the others.

F. Regression forecasting method:

Regression is a way constructing relationship between the variables based on some existing data and then further the fitted relation can be used for forecasting of the dependent variable. Deciding about the format of line (linear or second degree equation) is the most critical part of the process which can be achieved initially through scatter plots.As it is building some relationship in the form an equation, we may not require orderliness in the data. Moreover, several estimation procedures are available to fit this type of equation and user can perform powerful residual analysis to choose the best model.

II. STEPS INVOLVED IN TIME SERIES FORECASTING

A. Selection of the historical data for forecasting

Data selection for the time series forecasting is the first step involved in any forecasting. Depending on the nature of data and availability, we need to select the past data and it should be at least contains more than 50 observations as a thumb rule. Selection of more data also sometimes affects the performance of the model in the sense that the latest observations get some what less importance. The following points should be considered while selecting the historical data for time series forecasting. If the forecasting objective is for long range, say for more than two years, then it is better to take all the possible historical data since there is expected some long term influential factors in the data.

If the objective for forecasting is for shorter period, then user may restrict the input data to at least 50 observations.Preference should be given to recent past data only. In the case of non availability of recent past data, the we can take old historical data with proper assumptions and limitations of the forecasting process.

B. Data preparation for time series modeling

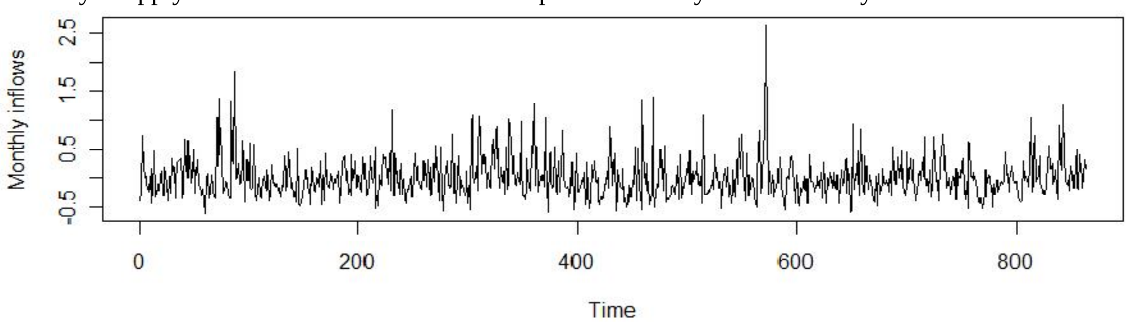
In the real time scenarios, there is very less chance to get clean and complete data for any forecasting project. The data obviously contains missing values, outliers and unavailable columns. As per time series analysis is concerned, we need to implement some of the following data preparation steps in the analysis.

1) Complete the time series for all forecasting runs: In many practical scenarios, we can observe that there are some rows missed out from each time series which we call as incomplete time series. In this case the forecasted values may go wrong and even not reliable. We need to check the completeness of the data say for example, if the historical data is monthly data, then data should be available for all the 12 months in each year.
2) Proper treatment should be given for missing values are non available values. Missing values are inevitable in any forecasting process and there are lot of procedures are described to fill the missing values like filling with zero's, filling with mean or median, removing the observations etc. One should select proper treatment depending on the type of problem for the missing value treatment.
3) There should be some graphical checks for outlier identification as they are impacting the final forecasted values in a great way. If they are not relevant, then it better always to remove them completely from the analysis. Otherwise, we need to think of replacing the outliers with proper substitutes in the analysis.
4) Standardization or normalization is one of the important steps in data preparation activity since it removes the effect of huge values and also it improves the model performance in any software application.

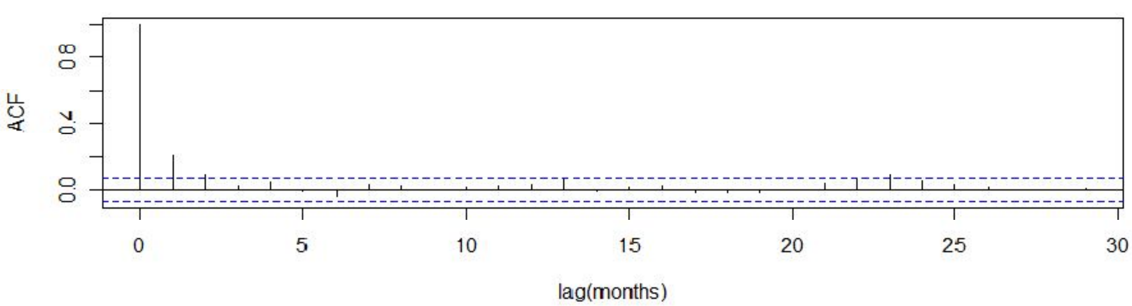
C. Primary check for stationarity in the data

A time series is said to be stationary if there is no systematic change in mean (no trend), if there is no systematic change in variance and if strictly periodic variations have been removed. In other words, in the entire time series data should approximately show uniform mean and variance. The main assumption of time series model is that the data should be stationary and hence we can check the stationary in the following ways:

1) Check for the stationarity of the data.: A normal time series plot can be used to check the stationarity in the data as it is very flexible and easy to apply. We can also use ACF or PACF plots to identify the stationarity in the time series data.



Time series plot of Monthly income



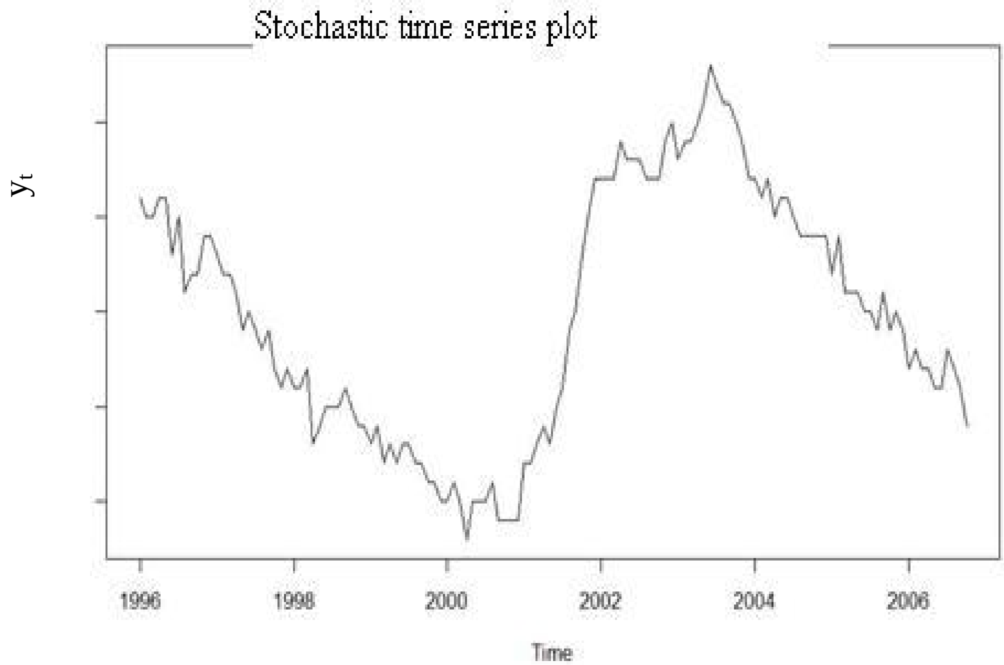
ACF plot for monthly income

In the above plot we can observe that almost the entire time series shows uniform mean and variance in the data.

2) *Removing the non stationarity in the data:* To remove the non stationarity, we need to follow any one of the following methods.

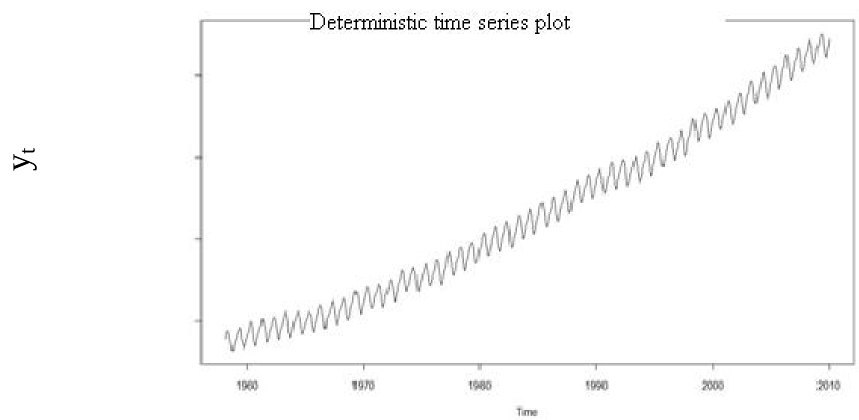
a) one of the available options for time if the non stationarity is observed due to un uniform variance, then log transformation or square root transformation will be the best idea to gain the stationarity in the data.

b) If there is nonstationarity due to stochastic trend, then differencing the time series analysis is the best method to avoid non stationarity. The time series plot of stochastic trend in this case should looks as follows.



Trend pattern in time series

If the non stationarity is due to deterministic trend, then we need to use regression modeling for dealing with this problem. A deterministic trend in the data looks as follows.

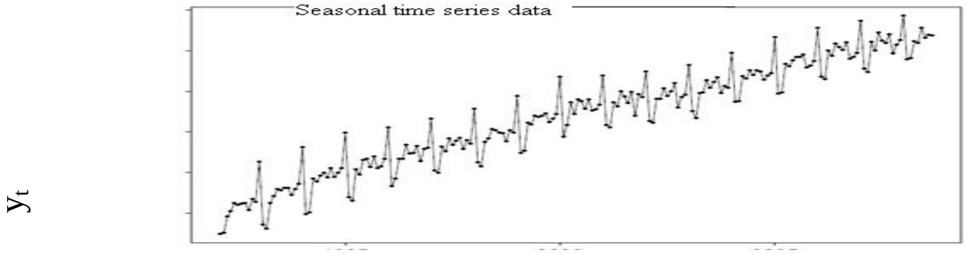


Deterministic trend equation

If the non stationarity is due to change in the mean, then we can go for first order or second order differencing of time series data.

3) *Validation of seasonality factor :* The uniform changes in the data at periodical intervals likes seasons are called as seasonal fluctuations and we need to consider the seasonality in the data while forecasting.The identification of seasonality can be done with regular time series plot and a typical seasonal data looks as follows.

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

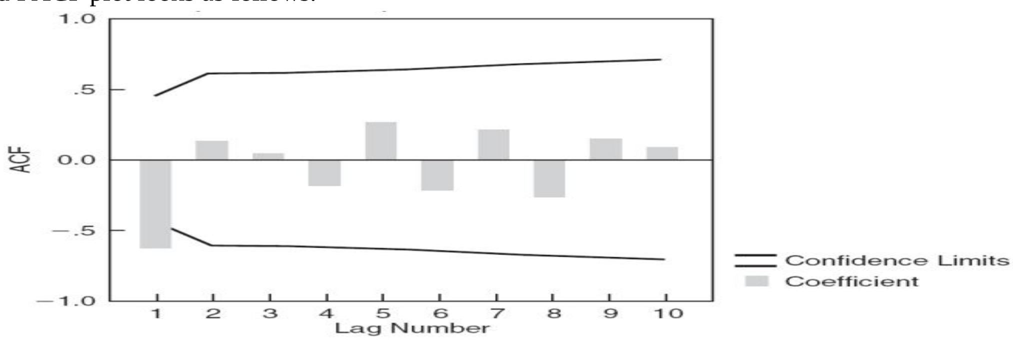


Identification of seasonality

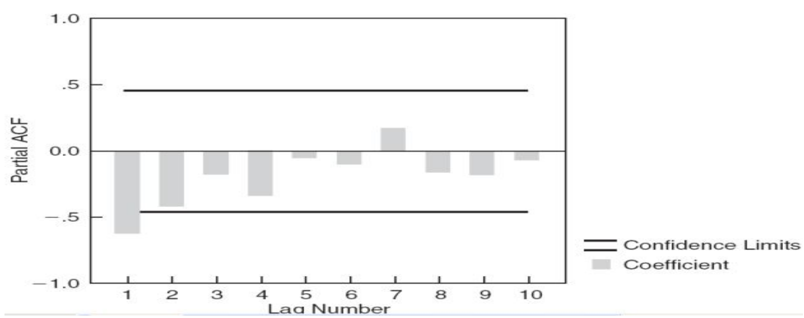
To deal with seasonality, we can add the adjustment factor in the model as well as seasonal parameter in the model. Many software's will allow the user to forecast the values with seasonality in the data.

4) *Check for auto regressive factor:* Order of auto regressive factor is an important input for time series models like ARIMA, ARCH etc. It determines the strength of correlation between subsequent values of the time series data. To know the auto regressive factor, we can construct ACF (Auto Correlation Function) plot or PACF (Partial ACF) plot and the spikes in the plots suggest the order of auto regression.

A general ACF and PACF plot looks as follows.



ACF Plot



PACF Plot

III. MODEL SELECTION CRITERIA IN TIME SERIES

Model selection, the extraction of relevant covariates or lags in explaining time series observations, is an important issue in applied econometrics. A widely followed strategy is to start with some general, unrestricted model which is subjected to subsequent reduction. There are several criteria for best model selection in the literature and some of the main criteria are described in the following sections.

A. Akaike Information criteria:

Akaike (1974) proposed a very simple model comparison tool that is based on likelihood function. This criterion used a penalty term. This criterion uses a penalty term to penalize the log maximum likelihood for lack of parsimony. If In L is the log maximum likelihood, Akaike's criterion is computed as

$$A = \text{In } L - (\text{number of parameters})$$

In the regression case, with v independent variables, there are v + 1 total estimated regression coefficients and hence this criterion reduces to

$$A = \text{Ln } L - v - 1$$

B. Schwarz information criteria

Schwarz (1978) criticized Akaike's criterion as being asymptotically nonoptimal and provided a simple alternative, based on a Bayesian argument. His mathematical results lead to a revised form of the penalty function, but again one which is simple computationally. This criterion can be described as

$$B = \text{Ln } L - \left[\frac{Ln\ L}{2}\ (\text{no of parameters})\right]$$

C. Mallows Cp criterion

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

$$C_p = SS_{res}/MS_{res} - N + 2p, \text{ where}$$

SS_{res} is the residual sum of squares for the model with p-1 variables, MS_{res} is the residual mean square when using all available variables, N is the number of observations, and p is the number of variables used for the model plus one. Thegeneral 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.

IV. CONCLUSIONS

The basic idea of applied forecasting is to try to forecast the next value in a time series of observations given only a knowledge of previous values. During the past few years, a number of applied forecasting techniques have been introduced which seemed initially to have a great deal of promise. Forecasting techniques that are based on regression analysis are substantially different in their underlying concepts and theory from the techniques of time series analysis, smoothing and decomposition. Regression techniques are generally referred to as causal or explanatory approaches to forecasting .In this paper we discuss about various types of classification methods in forecasting.

REFERENCES

[1] A. R. Venkatachalam and Jeffrey E. Sohl. (1999), “An Intelligent Model Selection and Forecasting System”, Journal of Forecasting. 18, pp. 167 – 180.
[2] Aaron Smith. (2005), “Forecasting in the Presence of Level Shifts”, Journal of Forecasting. 24, pp. 557 – 574.
[3] Abraham, B. and Ledolter, A. (1983), “Statistical Methods for Forecasting”, New York: Wiley.
[4] Box, G. E. P., and Jenkins, G. M. (1976), “Time Series Analysis, Forecasting and Control”. San Francisco: Holden-Day, Inc.
[5] Box, G.E.P., Jenkins, G. M., &Reinsel, G. C. (1994), “Time Series Analysis: Forecasting and Control (3rd ed.)”, Englewood Cliffs, New Jersey: Prentice Hall.
[6] Eugene F. Fama and Kenneth R. French. (2000), “Forecasting Profitability and Earnings”, Journal of Business, Vol. 73, No. 2, pp. 161 – 175.
[7] Evan F. Koenig, Sheila Dolmas and Jeremy Piger. (2003), “The Use and Abuse of Real – Time data in Economic Forecasting”, The Review of Economics and Statistics, 85 (3), pp. 618 – 628.
[8] Gardner, E. S. Jr., and E. McKenzie. (1985), “Forecasting Trends in Time Series”, Management Science, 31, pp. 1237-1246Hamilton, J. D. (1994), “Time Series Analysis”, Princeton University Press, New Jersy.
[9] Jonathan D. Linton. (2002), “Forecasting the Market Diffusion of Disruptive and Discontinuous Innovation”, IEEE Transactions on Engineering Management”, Vol. 49, No. 4, pp. 365 – 374
[10] Jose Manuel Pavia – Miralles. (2005), “Forecasts From Nonrandom Samples: The Election Night Case”, Journal of American Statistical Association, Vol. 100, No. 472, pp. 1113 – 1122
[11] Mc Laughlin, R. L. (1962), “Time Series Forecasting”, Marketing- Research Technique, Ser.6. New York: American Marketing Association
[12] McQuarrie, A. D. R., & Tsai, C. (1998), “Regression and Time Series Model Selection”, Singapore: World Scientific.
[13] Nelson, C. R. (1973), “Applied Time Series Analysis for Managerial Forecasting”, San Francisco: Hoden-Day, Inc.
[14] Ruey S. Tsay. (2000), ‘Time Series and Forecasting: Brief History and Future Research”, Journal of American Statistical Association, Vol. 95, No. 450, pp. 638 – 642.
[15] Thomopoulos, N. T. (1980), “Applied Forecasting Methods”, Englewood Cliffs, N. J. Prentice – Hall
[16] Tiao, G. C., and Guttman, I. (1980), “Forecasting Contemporal Aggregates of Multiple Time Series”, Journal of Econometrics, 12, pp. 219 - 230.