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:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

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

Estimation and Model Selection for Time Series Forecasting

G.Y. Mythili[1], C.Narayana[2], J. Prabhakara Naik[3], K.Vasu[4], G. Mokesh Rayalu[5]
[1]*Assistant Professor, Department of Mathematics, School of Advanced sciences, VIT, Vellore*
[2]*Assistant professor, Department of Mathematics, Sriharsha Institute of P.G. Studies, Nellore*
[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: ***Arrangement of any study variable by its time of occurrences is called time series and hence time is one of the key variables in time series analysis. The analysis of experimental data that have been observed at different points in time leads to new and unique problems in statistical modeling and inference. The obvious correlation introduced by the sampling of adjacent points in time can severely restrict the applicability of the many conventional statistical methods traditionally dependent on the assumption that these adjacent observations are independent and identically distributed. The systematic approach by which one goes about answering the mathematical and statistical questions posed by these time correlations is commonly referred to as time series analysis. Time-series analysis is used when observations are made repeatedly over 50 or more time periods. Sometimes the observations are from a single case, but more often they are aggregate scores from many cases. For example, the weekly output of a manufacturing plant, the monthly number of traffic tickets issued in a municipality, or the yearly GNP for a developing country, all of these tracked over considerable time. One goal of the analysis is to identify patterns in the sequence of numbers over time, which are correlated with themselves, but offset in time. Another goal in many research applications is to test the impact of one or more interventions (IVs). Time-series analysis is also used to forecast future patterns of events or to compare series of different kinds of events. Time series analysis provides tools for selecting the best suited model that can be used to forecast of future events. Modeling the time series is a statistical problem that involves some statistical tools like estimation technique and testing of hypothesis. Forecasts are used in computational procedures to estimate the parameters of a model being used to allocate limited resources or to describe random processes such as those mentioned above. Time series models assume that observations vary according to some probability distribution about an underlying function of time.***

I. INTRODUCTION

Time series forecasting is very importance even in the financial as well as other domains which needs some kind of prediction over time. One of the reasons for its importance is preventing undesirable events by forecasting the event, identifying the circumstances preceding the event, and taking corrective action so the event can be avoided. It means that we can avoid a low sales even if it is predicted correctly by implementing several sales development programs. Forecasting also helps to reduce the impact of some unavoidable events by predicting them well in advance.

We can save the lives of thousands in case if predictions of weather forecasts done properly. Many public and private organizations are concentrating more on improving their predictive power so that they can reduce the bad effects of un avoidable events. Finally, many people, primarily in the financial markets, would like to profit from time series forecasting. Whether this is viable (or) not is most likely a never-to-be-resolved question, nevertheless many products are available for financial forecasting. Because of this importance, a new branch of time series called financial time series has been comein to the picture and several models like ARCH and GARCH are developed as a result of this.

II. ESTIMATION OF PARAMETERS OF THE MODEL

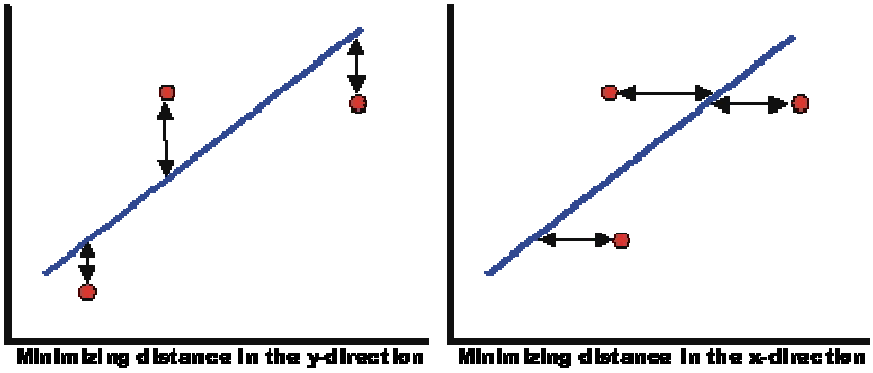
Once a Model has been selected, the parameters of the model need to be estimated. Several parameter estimation methods are available. This section will present an overview of these methods, starting with the relatively simple method of probability plotting and continuing with the more sophisticated least squares and maximum likelihood methods.

A. Probability Plotting

The least mathematically intensive method for parameter estimation is the method of probability plotting. As the term implies, probability plotting involves a physical plot of the data on specially constructed probability plotting paper. This method is easily implemented by hand, given that one can obtain the appropriate probability plotting paper. The method of probability plotting takes the cdf of the distribution and attempts to liberalize it by employing a specially constructed paper. Let us consider the linear function involving two parameters say, a and b. Now the estimates of these parameters are obtained from probability plot with slope and intercept parameters. Besides the most obvious drawback to probability plotting, which is the amount of effort required, manual probability plotting is not always consistent in the results. Two people plotting a straight line through a set of points will not always draw this line the same way, and thus will come up with slightly different results. This method was used primarily before the widespread use of computers that could easily perform the calculations for more complicated parameter estimation methods, such as the least squares and maximum likelihood methods.

B. Least squares parameter estimation

The method of linear least squares is Most frequently used for all regression analysis. The method of least squares requires that a straight line be fitted to a set of data points, such that the sum of the squares of the distance of the points to the fitted line is minimized. This minimization can be performed in either the vertical or horizontal direction. If the regression is on *X*, then the line is fitted so that the horizontal deviations from the points to the line are minimized. If the regression is on *Y*, then this means that the distance of the vertical deviations from the points to the line is minimized. This is illustrated in the following figure.



Least squares perameter estimation line

Now the following steps are performed to get the least square estimates.

Formulate the error term involved in the estimation as $e_t=Y-\hat{Y}$

Now sum squares of error is given by $Q = \sum e_t^2$

Now the sum of squares is differentiated with respect to parameters and equated to zero.

These equations are called normal equations and they are solved for getting the least square estimates of the parameters

1) *MLE parameter estimation:* From a statistical point of view, the method of maximum likelihood estimation is, with some exceptions, considered to be the most robust of the parameter estimation techniques discussed here. The basic idea behind MLE is to obtain the most likely values of the parameters, for a given distribution, that will best describe the data.

If *x* is a continuous random variable with *PDF:*

$$f(x; \theta_1, \theta_2, \ldots. \theta_k)$$

where θ_1, θ_2,...,θ_k are *k* unknown parameters which need to be estimated, with *R* independent observations, x_1, x_2,...x_R, which correspond in the case of life data analysis to failure times. The likelihood function is given by:

$$L(\theta_1,\theta_2,.....\theta_k \ \ |k_1,k_2,.....k_R) = L = \prod_{i=1}^{R} f(x_1,\theta_1,\theta_2,.....\theta_k) \; i = 1, 2, ...,R$$

The logarithmic likelihood function is given by:

$$\Lambda = \ln \ \ L = \sum_{i=1}^{R} \ln \ f(x_i, \ \theta_1,\theta_2,..... \ \theta_k)$$

The maximum likelihood estimators (or parameter values) of θ_1, θ_2,..., θ_k are obtained by maximizing L or Λ.
By maximizing Λ, which is much easier to work with Λ than L, the maximum likelihood estimators (MLE) of θ_1, θ_2,..., θ_k are the simultaneous solutions of k equations such that:

$$\frac{\partial \Lambda}{\partial \theta_j} = 0 \text{ , j=1,2,…..k}$$

The MLE method has many large sample properties that make it attractive for use. It is asymptotically consistent, which means that as the sample size gets larger, the estimates converge to the right values. It is asymptotically efficient, which means that for large samples, it produces the most precise estimates. It is asymptotically unbiased, which means that for large samples one expects to get the right value on average. The distribution of the estimates themselves is normal, if the sample is large enough, and this is the basis for the usual Fisher Matrix Confidence Bounds. These are all excellent large sample properties.
the size of the sample necessary to achieve these properties can be quite large: thirty to fifty to more than a hundred exact failure times, depending on the application. With fewer points, the methods can be badly biased. It is known, This bias can cause major discrepancies in analysis. There are also pathological situations when the asymptotic properties of the MLE do not apply.

III. MODEL SELECTION FOR TIME SERIES FORECASTING

The selection of a best forecasting method is an important task and it is generally based on the several practical scenarios that a researcher can face in the field. Currently there are large numbers of forecasting models are available and user has to decide between these models very carefully after proper judgment for his model selection criteria.

A. Factors that effecting the model selection

The following are the important factors that will influence the model selection process. All the following factors may not arise at single data but they are appear at different scenarios of forecasting.

B. Prior knowledge about the data variable

Prior knowledge about the research data is very important factor in deciding model selection. It is purely based on the forecaster experience in the field and knowledge about the variables in the data. If the forecasting problem is related to identify the potential customers for a sales promotion, then usually there are infinite variables in the system and only forecasters can test the efficiency of only few parameters for model building based on his experience and prior knowledge.

C. Data availably and missing values

Availability of the data is one of the significant factor in deciding about the forecasting model. If we have at least 50 observation in the history, then probably any model can be tried out. If less number of historical data points are there, then one should be very careful about the model selection. For very less number of historical data, he may prefer to consider some models like Unobserved Components Model (UCM) etc.

D. Availability of tools and technology

In many situations, user may not able to have enough technology to run the decided model. In this case, we need to take the technology factor in to account while selection the model.
Some of models like UCM requires large memory to run their iterative procedures and not all software will support this. Depending on the available technology and tools, we can decide the best model in the time series forecasting.

E. Knowledge concerning the quantity being forecast:

This is also one of the influencing factor about the model selection in time series. If the quanitiy being forecasted is large, then a standard model like ARIMA will be useful since it gives better estimates for long range forecasting. For the short range forecasting we can go for simple graphical or moving average methods as they are accurate and easy to apply for such small forecasts.

F. Best model selection on different scenarios

Practically, it is not possible to put hard rules for model selection but we can highlights some of the models for particular scenarios depending on the various factors like expertise in modeling and data knowledge. With forecasting procedures, we are usually trying to recognize a change in the process of a time series while remaining insensitive to variations caused by purely random effects. The goal of planning is to respond to fundamental changes, not to spurious effects.

a) With a method based purely on historical data, we can not eliminate the noise factor from the data.. The problem is to set parameters that find an acceptable trade-off between the fundamental process and the noise. If the process is changing very slowly, both the moving average and the regression approach should be used with a long stream of data. For the exponential smoothing method, the value of α should be small to de-emphasize the most recent observations. Stochastic variations will be almost entirely filtered out.

b) If the process is changing rapidly with a linear trend, the moving average and the exponential smoothing methods are at a disadvantage because they are not designed to recognize trends. Because of the rapid changes, the time range of the moving average method must be set small and the α parameter of the exponential smoothing method must be set to a larger value so that the forecasts will respond to the change. Nevertheless, these two methods will always fall behind a linear trend. The forecasts will never converge to a trend line even if there is no random variation. Of course, with the adjustment of parameters to allow a response to a process change, the forecasts become more sensitive to random effects.

c) If there is perfectly linear trend in the data, then the exponential smoothing method with a trend adjustment and the regression method are both useful and appropriate since they will eventually converge to a trend line. Thus in the absence of a change in trend, the time range of the regression data can be large, and the α and β values of the exponential smoothing method can be small, thus reducing the random effects.

d) If the process is changing rapidly with rapid changes in the linear trend, each of the methods described in the above will have trouble, because it is difficult to separate

Changes in the process from random changes. The time ranges must be set small for the moving average and regression methods, resulting in sensitivity to random effects. Similarly, the α and β parameters for exponential smoothing must be set to larger values with a corresponding increase in sensitivity to randomness. Both the moving average and regression methods have the disadvantage that they are most accurate with respect to forecasts in the middle of the time range.

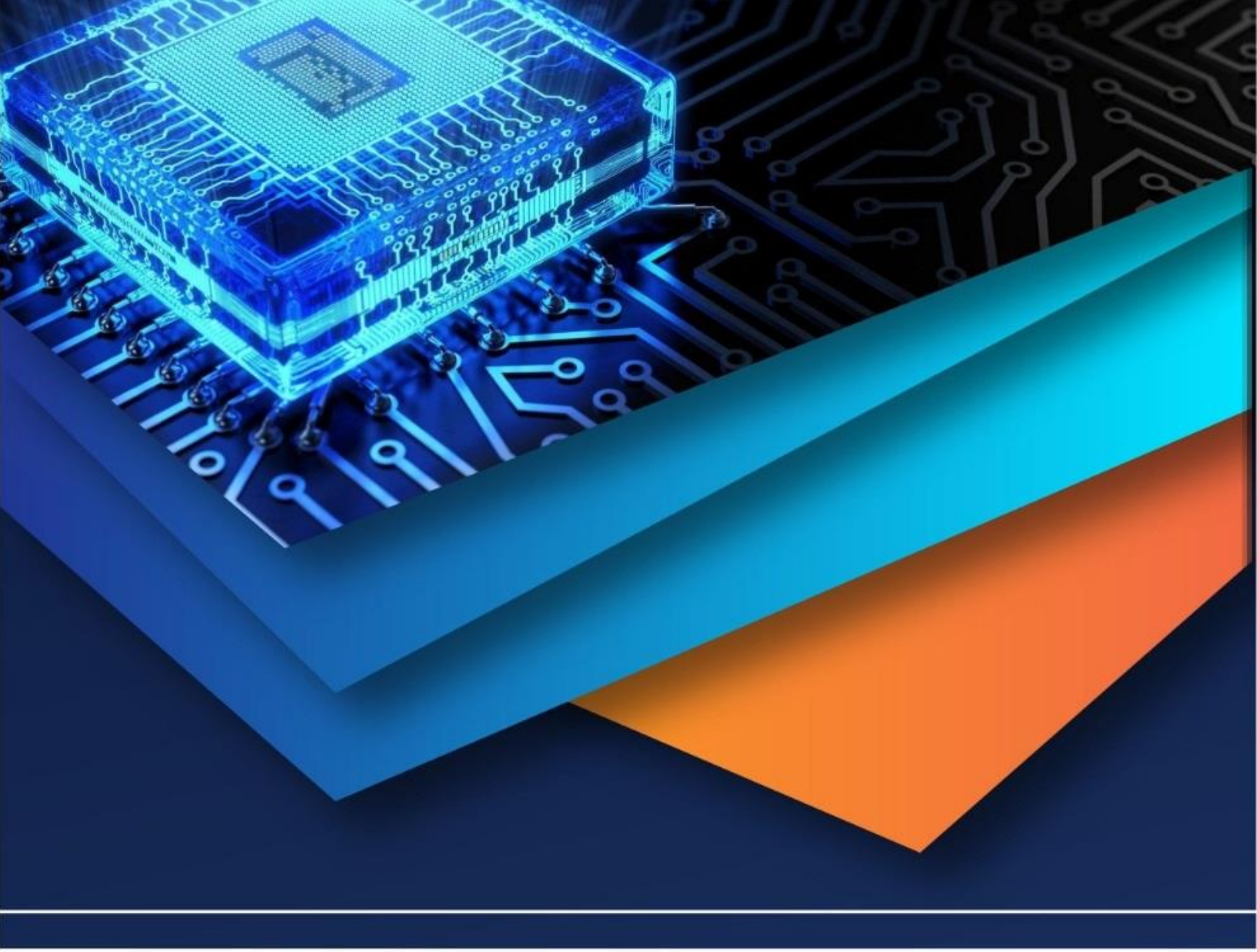
IV. CONCLUSIONS

Even though lot of research has been made in the area of time series forecasting, there are is lot of scope to further improvements as it is evident from the following limitations. The time series forecasts are not capable of capturing random or irregular fluctuations in the data. In other words, if there is some unknown random factor causes the changes in the data, then this model is not capable of including those factors in the model. Most of the time series predictions are based on historical values of the same variable and if it is influenced by external factors, then the model is not able to describe those factors. Time series modeling is dependent of stationarity and seasonality factors unless which we can not define time series models. Seasonal factors are not always capable in any time series data. In this paper we discuss about various seasonal effects are hidden in the data which are not able to identify through general graphical plots.

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. 1237-1246.
[9] Hamilton, J. D. (1994), "Time Series Analysis", Princeton University Press, New Jersy.
[10] 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.

[11] 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.
[12] Mc Laughlin, R. L. (1962), "Time Series Forecasting", Marketing- Research Technique, Ser.6. New York: American Marketing Association.
[13] McQuarrie, A. D. R., & Tsai, C. (1998), "Regression and Time Series Model Selection", Singapore: World Scientific
[14] Nelson, C. R. (1973), "Applied Time Series Analysis for Managerial Forecasting", San Francisco: Hoden-Day, Inc
[15] 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.
[16] Thomopoulos, N. T. (1980), "Applied Forecasting Methods", Englewood Cliffs, N. J. Prentice - Hall.
[17] Tiao, G. C., and Guttman, I. (1980), "Forecasting Contemporal Aggregates of Multiple Time Series", Journal of Econometrics, 12, pp. 219 - 230.
[18] Yulia Gel, Adrian E. Raftery and Tilmann Gneiting. (2004), "Calibrated Probabilistic Mesoscale Weather Field Forecasting: The Geostatistical Output Perturbation Method", Journal of American Statistical Association, Vol. 99, No. 467, pp. 575 – 583.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY