

# Gaussian and Non-Gaussian Autoregressive Time Series Models with Rainfall Data

Sukhpal Kaur, Madhuchanda Rakshit



**Abstract-** The Gaussian and non-Gaussian autoregressive models are used in this paper for analyzing time series data. The autoregressive time series models with various distributions are considered here for analyzing the annual rainfall of Punjab, India. Three different types of autoregressive models are applied here for analyzing data namely autoregressive model with Gaussian, Gamma and Laplace distribution. For the goodness of fit the chi-square test is applied and the best fitted distribution is obtained for the data. Next the stationarity of data is checked, after that models are applied on data for comparing three distributions of AR models and lastly the best fitted model is obtained. The residual checking of selected model is also discussed and forecast the best fitted model based on simulated response comparison.

**Keywords:** Chi-square test, Laplace AR model, Gamma AR model, Gaussian AR model, NRMSE

## I. INTRODUCTION

The necessity of real valued and natural occurring time series observation yields the interest in this field by using various distributions with autoregressive process. Recently various different distributions are used by authors for analyzing the various time series models. However this is a very new field for the researchers and a very few work on real data have been found up to till date. The analysis of experimental data that have been observed at different points of time leads to unique problems in statistical modelling and inference. There are many natural occurring times series observations which show a tendency to follow asymmetric and heavy tailed distributions. Time series models for real valued observation using non-Gaussian distribution has been increased from the last few decades. The era of linear time series models began with autoregressive models. Autoregressive models with normal distribution used as a statistical tool for analyzing various time series data analysis. Yule and Walker introduced Autoregressive Moving Average (ARMA) model. Autoregressive Integrated Moving Average (ARIMA) model is proposed by Box and Jenkins [3] and they give detailed knowledge about ARIMA and Seasonal ARIMA models in their book.

In recent times there has been considerable research in the development of time series models with seasonal or periodic properties in the meteorological and hydro logical area. Another concept of autoregressive model known as Periodic Autoregressive (PAR) model and Laplace Autoregressive (LAR) has also been frequently used recently in various environmental, hydrological and meteorological studies [21].

Gaussian process is very convenient in environmental sciences, but it is not applicable for skewed marginal distribution. Gastwirth and Wolff [7] had developed a stationary linear first order autoregressive with the marginal distribution process, i.e. called LAR (1). Gaver and Lewis [8] developed a linear AR (1) process satisfying with the gamma marginal distribution. The Laplace LAR (1) model, and its generalizations to higher order correlation structures, proposed by Dewald and Lewis [6], the autoregressive process by using Laplace distribution and is termed as Laplace Autoregressive (LAR) model. In 1989, Damsleth and El-Shaaravi [4] developed a time series model with Laplace noise as an alternative to the normal distribution. Gibson J. D. [9] applied an AR (1) process for image source modelling in data compression tasks. C. H. Sim [23] discussed the general theory of model-building approach which consists if model identification, estimation, diagnostic checking and forecasting for a model with given marginal distribution and applied different type of autoregressive models with various distributions in paper on the discharge of the Mekong River, Thailand. Debasis Kundu [5] studied the discrimination between normal and Laplace distributions. Billard [2] proposed autoregressive model with exponential distribution and Gourierroux, C. et al. [10] introduced autoregressive model with gamma distribution. Kuttykrishnan A.P. [17] had studied about the Laplace autoregressive time series models. He had also discussed the Laplace distribution as a symmetric and asymmetric. The properties of asymmetric Laplace Autoregressive model have also explained. After that, he defines the generalized asymmetric Laplace process for first order autoregressive. Krishnan B. And George D. [15] defined a first order moving average model with Laplace marginal distributions extension of higher order. A first order moving average process with mixed Laplace distributions as marginal is developed and also introduced this process as the mixture of asymmetric Laplace marginals. Nguyen H. D. et al. [20] introduced the Laplace mixture autoregressive model (LMAR) model that utilizes a Laplace mixture conditions model, as an alternative to the Gaussian mixture autoregressive (GMAR) model. Jayakumar and Kuttykrishnan [11] developed time series models and discussed the application of asymmetric Laplace distribution in modelling currency exchange rate, interest rate, stock

Revised Manuscript Received on October 30, 2019.

\* Correspondence Author

**Sukhpal Kaur\***, Research Scholar of Applied Science Department, Guru Kashi University.

**Dr. Madhuchanda Rakshit**, Assistant Professor at Guru Kashi University.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

price changes etc. Jose et al. [16] and Lishamol and Jose [14] developed a unified theory for a Gaussian and non-Gaussian autoregressive processes through normal- Laplace and generalized normal- Laplace distributions. Kuttykrishnan and Jayakumar [18] introduced bivariate semi  $\alpha$ - Laplace distribution, its characterizations and associated autoregressive models. In recent years Johannesson et al. [12] applied Laplace distribution models in their paper on data related to road topography and roughness. Miftahurromah B. [19] et al. used Bayesian mixture Laplace Autoregressive approach for modelling the Islamic stock risk investment. In 1990 Sim, C. H. [22] studied Gamma and Exponential autoregressive process. Johannesson et al. [13] also studied autoregressive model with gamma variance for road topography modelling. Berger, K. et al. [1] has done studied on goodness of fit measures for model evaluation. In this study for simulated response comparison used the concept of goodness is used. In this paper three different types of autoregressive time series distributions such as Gaussian, Gamma and Laplace distributions are applied for analyzing the annual rainfall of Punjab, India which is sourced from Indian Metrological Department (IMD) on the basis of subdivisonal area of Punjab from the period 1871 to 2016.

## II. METHODOLOGY

Autoregressive time series models are the most widely used models among the stochastic models. The main reason behind this is that the autoregressive form has an inbuilt factor of time dependence. The three models used for this paper are discussed as follows:

### A. Gaussian AR model

The general structure of an autoregressive process of order p for AR (p) model is written as:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \varepsilon_t$$

where  $\varepsilon_t$  is white noise, i.e.  $\{\varepsilon_t\} \sim N(0, \sigma^2)$  and  $\varepsilon_t$  is non-correlated with  $X_s$  for each. The conditional density and the joint probability density function for Gaussian AR process is symmetric and it is also time reversible model.

### B. Gamma AR model

A linear autoregressive models for gamma distributed processes, is called Gamma autoregressive (GAR) model. The model is based on the assumption that the underlying series have a gamma marginal distribution and the equation is obtained as

$$X_i = \phi X_{i-1} + \varepsilon_i$$

where  $\varepsilon_i$  is a sequence of independent and identically distributed gamma  $(\alpha, \beta)$  random variables with  $\alpha, \beta > 0$ . The marginal density of  $\{X_i\}$  and its conditional density are respectively,

$$f_X(x) = [\alpha(1-\phi)]^\beta x^{\beta-1} \exp[-\alpha(1-\phi)x] / \Gamma(\beta)$$

and

$$f_{X_{i+j}|X_i}(x|y) = \delta \left( \frac{x}{\phi^j y} \right)^{(\beta-1)/2} \exp[-\delta(x + \phi^j y)] I_{\beta-1} [2\delta(\phi^j xy)^{1/2}]$$

where  $\delta = \alpha(1-\phi)/(1-\phi^j)$ ,  $0 < \phi < 1$ , and  $I_k(z)$  is the modified Bessel function of the first kind and order is  $k$ . The conditional density and probability density function of gamma model is not symmetric in  $x$  and  $y$  and is time reversible.

### C. Laplace AR model

The Laplace AR process is similar to the exponential AR process on the basis of their construction. If  $\{\varepsilon_t\}$  be a sequence of independent and identically distributed random variables then

$$\varepsilon_t = \begin{cases} 0 & \text{with probability } \rho^2 \\ l_t & \text{with probability } 1-\rho^2 \end{cases}$$

where the relation  $X_t = \rho X_{t-1} + \varepsilon_t$ ,  $|\rho| < 1$  and  $\{l_t\}$  is a sequence of iid standard Laplace variates. The marginal probability density function and conditional density of  $X_{t+i}$  given  $X_t = y$  are respectively,

$$f_X(x) = \frac{1}{2} \exp(-|x|), \quad -\infty < x < \infty$$

$$\text{and } f_{X_{t+i}|X_t}(x|y) = \rho^{2i} \lambda(x - \rho^i y) + \frac{1}{2} (1 - \rho^{2i}) \exp(-|x - \rho^i y|), \quad -\infty < x, y < \infty$$

The conditional density and the joint probability density function for Laplace AR process is symmetric and also the model is time reversible.

The following paper is summarized as follows -

- At the very first step the distribution is fitted on raw data and most appropriate distribution is chosen on the basis of chi-square test.
- Next, AR models are fitted for the various distributions and chi-square test is applied for finding out the appropriate model.
- Next, simulated response comparison is used for verification of the chosen appropriate model on the basis of normalized root mean square error.
- Further for checking the adequacy of the selected model The residual checking is done on the basis of ACF residual plot and Q-Q plot.

- Lastly, decision is taken for forecasting the rainfall data by selecting the appropriate model.
- The whole analysis of this paper is done by Matlab (2018b).

The layout of the paper is provided in the following flow chart.

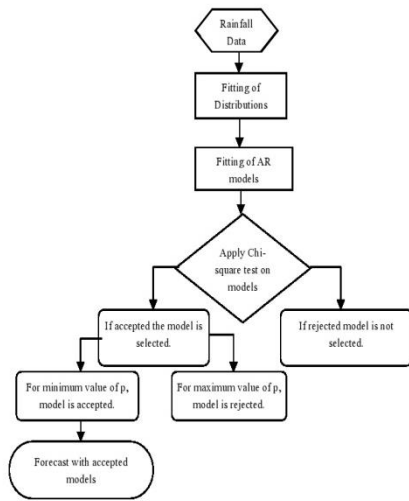


Fig. 1: Layout of Paper

### III. Result and discussion

In this section at first the graphical analysis of three distributions are discussed and chi square test for goodness of fit is applied. After that the stationarity of data is checked before fitting the AR models. Next step is to fit the three AR models and to find their simulated plots. Then the chi square test is used for the best model selection and checking the residuals of selected model and finally forecast the fitted model.

#### A. Graphical Analysis of Distributions

In this section, the three different distributions namely Gaussian, Gamma and Laplace distribution with respect to rainfall data of Punjab, India are analyzed. Fig. 2 depicts the fitting of Gaussian and Gamma distribution with the actual rainfall data. Fig. 3 shows the Laplace distribution with actual data.

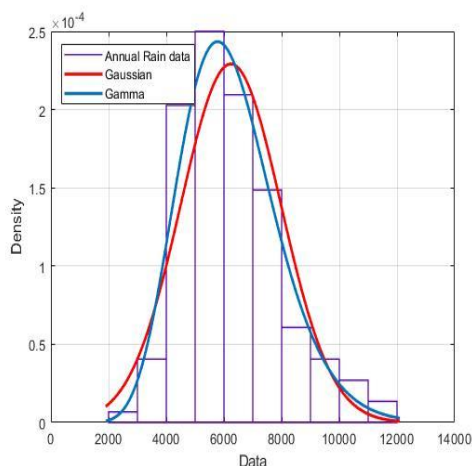


Fig. 2. Graph of Gaussian and Gamma Distributions

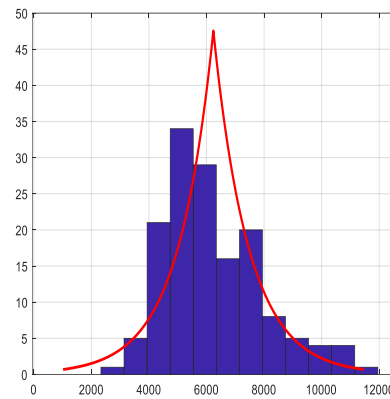


Fig. 3. Graph of Laplace distribution

For the goodness of fit, the chi-square test is applied on three distributions for actual and log transformed data. Following table I shows the h and p values of chi-square test, according to this test if the value of ‘h’ is 1 then test rejects the null hypothesis otherwise it will be accepted.

Table- I: Chi- Square Test Values on Three Distributions

Data		Gaussian Distributio n	Gamma Distributio n	Laplace Distributio n
Chi-square Test	Actual Data	h = 0, p = 0.8918	h = 1, p=0.01	h = 1, p=0.0004
	Log Data	h = 0, p = 0.3775	h = 0, p=0.0081	h = 1, p= 5.9593e-08

From the above Table I of Chi square distribution it can be concluded that among the applied three distributions Gaussian distribution is more appropriate for the data set on the basis of least p values.

#### B. Fitting of AR models

Three types of AR processes viz. Gaussian AR process, Gamma AR process and Laplace AR processes are applied in this section. Effect of seasonality is also observed by studying the graphical presentation of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

#### Autocorrelation Plot

The ACF and PACF plots are shown in Fig. 4. These plots are useful for finding the existence of seasonality and stationarity of data. From the autocorrelation plots it is analyzed that there is no seasonality and stationarity effects found on data as all spikes are under the level of significance. Therefore there exist no outliers in time series and from the autocorrelation plots the order of AR- process can also be determined for analysis. So now data set is ready to apply the autoregressive models.

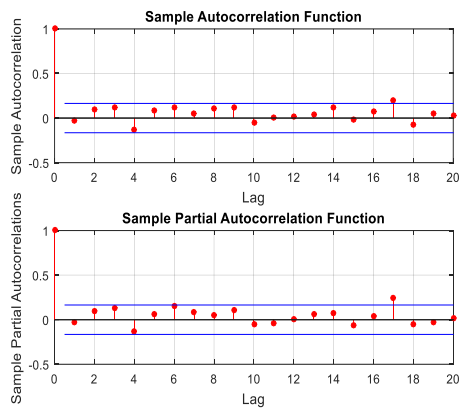


Fig. 4. Graph of Autocorrelation Functions

▪ Fitting of AR models with various distribution

In this section the simulated plots of Gaussian AR- process, Gamma AR- process and Laplace AR- process are provided in the following fig. 5 which shows the trend of our fitted AR models for these three distributions models with parameters  $a = -0.734$ ,  $\sigma_y = 7.73$  and  $\nu = 14.2$ .

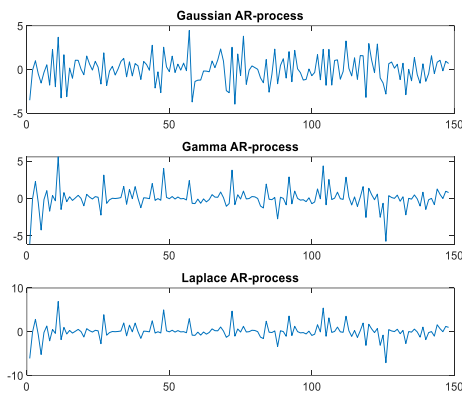


Fig. 5. Simulated AR- process with three distribution

Now the chi square test is again applied on various AR processes for obtaining the better conclusion among the most fitted AR models for forecasting data. Following table shows the chi-square test values of the three models.

Table -II: Chi –Square test values for goodness of fitted Models

Models	Gaussian AR Process	Gamma AR Process	Laplace AR Process
Chi-square Test	h=0 p=0.352	h=0 p=0.629	h=0 p=0.578

From the above table value of calculated chi square test, it is observed that all the ‘h’ values are zero, therefore all the distribution models are accepted for the data as according to the chi square test and among all p values the Gaussian AR model has minimum value of p i.e. 0.352. So it can be concluded that Gaussian AR model is more appropriate AR model among three AR models.

As the most suitable distribution and AR model is selected, now the simulated response comparison is provided here for

comparing the simulated Gaussian AR model with actual data. Normalized root mean square error (NRMSE) measure is used here for goodness of fit between simulated response and actual data.

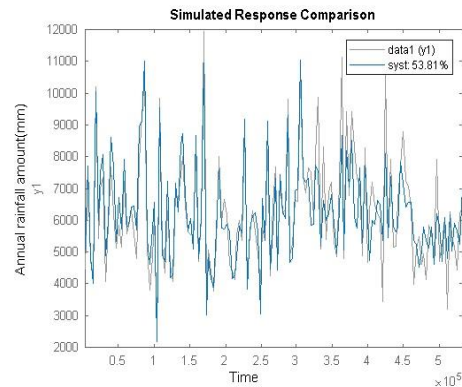


Fig. 6. Graph of Simulated Response Comparison

Fig.6 illustrates the differences between model response (i.e. Gaussian AR model) and the actual data. The percentage of NRMSE fitness value is 53.81% which represent the closeness of the selected forecast model to the actual data.

▪ Residual Plot of fitted model

Residual checking is done for finding whether the information gathered for the most suitable model is adequate or not. The residual should be uncorrelated with zero mean, constant variance and should follow the normal distribution for a good forecasted model. From the ACF residual plot in fig.7, it is observed that the residuals are close to zero and there is no significant correlation in the residuals series.

The Q- Q plot is a diagnostic tool for checking the normality of residuals. If the quintiles of the actual data and selected model are in the same distribution then the Q-Q plot would be a roughly straight line. Fig.8 is the Q-Q plot of Gaussian AR model, It is clear from the figure that the plot is roughly normally distributed.

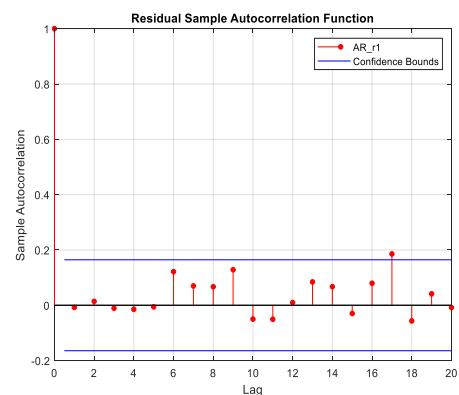


Fig. 7. ACF residual plot of Gaussian AR model



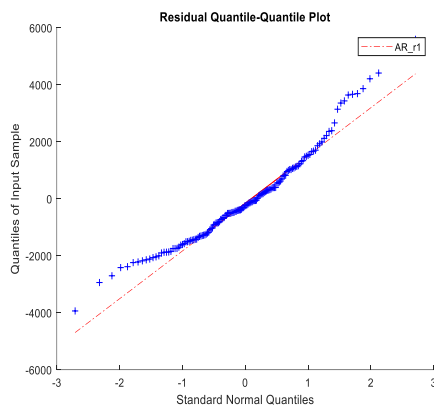


Fig. 8. Q – Q residual plot of Gaussian AR model

Forecasting of Gaussian AR model

Lastly, the forecasted plot of rainfall data by using Gaussian AR model is shown in fig.9, which shows the plot of actual rainfall data of past (i.e. from 1871 to 2016) and forecasted annual rainfall of Punjab up to 2036. According to the trend of plot the minimum and maximum forecasted annual rainfall value of Punjab is recorded as apprx. 2440 mm and 8320 mm.

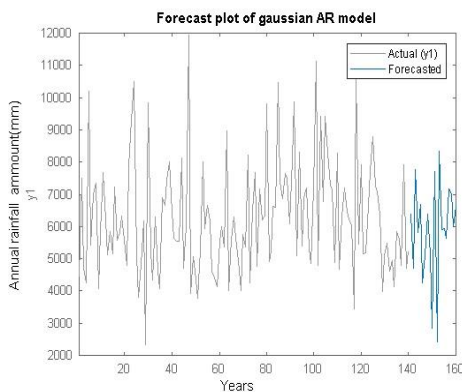


Fig. 9. Forecast plot of Gaussian AR model

IV. CONCLUSION

Estimation of rainfall data is very important nowadays for utilization of water resources in all over world as well as for India and Punjab. It is useful for designing drainage systems irrigation. So concern is increasing for its valid estimation and proper methods for predicting annual rainfall. Various statistical methods are available in literature but the appropriate method for a particular data is not fixed. In this study, three different distributions are fitted for annual rainfall data of Punjab based on Autoregressive models. Among them Gaussian AR process is established as more accurate than others. And the forecast is done on the basis of Gaussian AR model up to next 20 years and maximum and minimum amount of annual rainfall is also recorded and mention in subpart of forecasting section B of result and discussion. Various statistical tools are used for selecting appropriate distribution and AR model such as chi square test, simulated response comparison, ACF residual and QQ plot for residual checking etc. From the application point of view the rainfall data with Gaussian and non Gaussian distribution is a very new concept. Lots of future research

work can be carried out by future researchers keeping in view the application of non Gaussian AR models.

REFERENCES

- Berger K., "Goodness of fit measures: what do they tell about vegetation variable retrieval performance for earth observation data", Proceedings of SPIE- The International society for Optical Engineering, 2016
- Billard L. , "Exponential Autoregressive Models", Wiley StatsRef: Statistics Reference Online, John Wiley & Sons, Ltd, 2014
- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C., Time Series Analysis, Forecasting and Control,3rd Edition, 2009 Pearson Education, Inc.
- Damsleth E. and Shaarawi A. H., "ARMA Models with Double – exponentially Distributed Noise", Journal of Royal Statistical Society, Series B, Vol. 51, Issue-1, 1989 pp. 61-69
- Debasis Kundu, "Discriminating between the normal and laplace distributions", Advances in Ranking and Selection, Multiple Comparisons, and Reliability, January, 2005, pp.65-79,
- Dewald L. S. and Lewis P. A. W., "A New Laplace Second – Order Autoregressive Time Series Model- NLAR (2)", IEEE Transaction on Information Theory, Vol. 31, 1985, pp. 645-652
- Gastwirth J. L. and Wolff S. S., "A Characterization of the Laplace Distribution", John Hopkins University, Department of Statistics Report, 1965, 28
- Gaver D. P. and Lewis P. A. W., "First –Order Autoregressive Gamma Sequence and point processes", Advances in Applied Probability, Vol No. 3, 1980, pp. 727-745
- Gibson J. D. , "Data compression of a first order intermittently excited AR process", In Statistical Image Processing and Graphics, 1986, pp. 115-126
- Gourieroux C. and Jasiak J., "Autoregressive Gamma Processes", Journal of Forecasting, Vol. No. -25, Issue 2, March 2006, pp. 129-152,
- Jayakumar K. Kuttykrishnan A. P. , "A time series model using asymmetric Laplace distribution", Statistics and Probability Letters, Vol. No. 77, 2007, pp. 1636-1640
- Johannesson, P., Podgorski, K. & Rychlik, I., "Laplace distribution models for road topography and roughness", Working Papers in Statistics, No. 14, Department of Statistics, Lund University, 2015
- Johannesson, P., Podgorski, K., Rychlik, I. and Shariati and Nima , " AR(1) time series with autoregressive gamma variance for road topography modelling", Working Papers in Statistics, No. 5, Department of Statistics, Lund University, 2015
- Jose K. K. , Lishamol T. and Sreekumar J., "Autoregressive processes with normal Laplace marginal", Statistics and Probability Letters, Vol. No. 78, pp. 2456-2462, 2008
- Krishnan B. and George, "A Mixed Asymmetric Laplace Moving Average Process", Global Journal of Pure and Applied Mathematics, vol. 13, Issue – 9, pp. 6919-6928, Nov. 2017
- K.K. Jose and M. M. Thomas, "A Product Autoregressive Model with Log- Laplace Marginal Distribution", Statistica, Vol. 3,2012
- Kuttykrishnan A. P., "Laplace Autoregressive Time Series Models", Department of Statistics, University of Calicut, 2006
- Kuttykrishnan A. P and Jayakumar K., "Bivariate semi  $\alpha$  – Laplace distribution and processes", Statistical Papers, Vol No. 49, pp. 303-313, 2008
- Miftahurrohman B., Iriawan N. and Kartika F.(2017), "On the value at risk using Bayesian mixture Laplace autoregressive approach for modeling the Islamic stock risk investment", international Conference on mathematics: Education, Theory and Application, IOP Conf. Series: Journal of Physics: Conf. Series 855(2017)012026
- Nguyen H. D., Geoffrey J. M., Jeremy F. P. Ullmann and Andrew L. J., "Laplace Mixture Autoregressive Models", Statistics and Probability Letters, Vol. No. 110,2016, pp. 18-24
- Sales, J. D., dealer J. W., Yevjevich V. And Lane, W. L., "Applied Modeling of Hydrologic Time series", Water Resources Publication, LLC, 1997

22. Sim C. H., "First – Order Autoregressive Models for Gamma and exponential Processes", Journal of Applied probability, Vol. No.27, 1990, pp. 325-332
23. Sim C. H., "Modeling Non-normal First – Order Autoregressive Time Series", Journal of Forecasting, Vol. No.13, pp. 369-381, 1994

## AUTHORS PROFILE



**Sukhpal Kaur**, M.Sc. Mathematics from Punjabi University, Patiala. Presently a research scholar of Applied Science Department, Guru Kashi University. She has published five research papers in international journals and presented few research paper related to her thesis work in various National and International conferences. Her area of research is time series models.



**Dr. Madhuchanda Rakshit**, Ph.D in Operation research. Presently working as Assistant Professor at Guru Kashi University. She has more than 35 research papers on her own credit. Her area of research is Queneing theory, Time series analysis and Quadratic and fractional programming problems.