International Journal of Applied Research
Journal HP: www.intjar.com, ISSN: 2411-6610

Time varying model for detection of trend in time series data of monthly count of rainy days in monsoon

Md. Asadul Islam[1*], Muhammad Shahrukh Rahman[2] and Md. Abdul Matin[1]

[1]*Department of Statistics, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh*
[2]*Agricultural Economics Division, Bangladesh Agricultural Research Institute, Joydebpur, Gazipur-1701, Bangladesh*

ARTICLE INFO

Article history

Received 9 January 2016
Accepted 21 May 2016
Online release 23 May 2016

Keyword

Rainy days,
Time Varying Model,
Classical Decomposition Method,
Box-Jenkins method

***Corresponding Author**

Md. Asadul Islam
E-mail: asad_ju333@yahoo.com
Phone: +880-1799-937979

ABSTRACT

This study examines the pattern of the number of rainy days in monsoon season and various time series methods used for detection of trend and forecast the future. Rainfall data was collected from Dhaka rainfall recording station from the periods 1973-2008. Monsoon period was considered because of the importance of rainfall affecting the crop production. The traditional classical decomposition method and Box-Jenkins methods were carried out to detect trend and forecast of future value. Time varying method was used for detection of trend and forecast future value to get the better estimated value than the traditional methods. Besides most appropriate state-space representation of the model were used for analyzing the data. This study reveals that in the classical decomposition method the predicted value of the log of the number rainy days rise slightly. It does not follow the same pattern as the original value. On the other hand the predicted value provides the similar pattern of the original data in the Box-Jenkins methods. Again the predicted value in the state-space representation methods follows the same pattern of the original data value. The predicted value which is obtained by state space representation is very similar to the original data value. For that state-space representation of the model gives better result in time series analysis for forecasting. In the time varying method the predicted values give the better result because its values are much closer to the original data set than the traditional time series model. It overcomes the limitations of the constant slope coefficient and makes the series stationary. It is revealed from the study that the numbers of rainy days are decreasing over time in the monsoon. This work will be helpful for further study of the spatial rainfall distribution over the country by considering all rainfall recording station of the country.

Introduction

Agriculture plays a vital role in the economies of Bangladesh. The contribution of agriculture in GDP is considerable. The agricultural sector of the country always depends on climatic conditions. Rainfall is most important natural factor that has significant contribution to agriculture production variability (Green, 1964). If the amount of rainfall decrease and the monsoon does not come in the proper time then its consequences will decrease the amount of agricultural production in the country. If the agricultural production decreases then it will affect the GDP of the country. It was, therefore, felt essential to construct a more rational model for detection of trend of rainy days in monsoon.

For estimating trend and forecast the series classical method and the popular Box-Jenkins methods were applied (Box & Jenkins, 1976). State-space approach was applied for estimating trend and forecasting the series to overcome the limitations of these two methods (Brockwell & Davis, 1996). In time series analysis the trend may be estimated or eliminated through the classical decomposition model where smoothing is performed to predict future events. In spite of some limitations Box-Jenkins approach was also used to forecast the time series data. Box–Jenkins methods for time series analysis are popular and widely applied. This approach is based on autoregressive integrated moving average (ARIMA) models.

In practice, some non-stationary features in the time series are present due to trend and/or seasonal effects. As a first step, the observed time series is transformed into a stationary series using time and lag functions. In practice, the trend and/or seasonal are removed from the series by *differencing*. A non-stationary random walk it can be turned into a random (stationary) process by taking the first differences (Hamilton, 1989). Despite the relationships between ARIMA and unobserved components time series models, the Box–Jenkins and state space approaches to time series analysis are distinct. In the state-space approach the non-stationary time series in terms of trend and seasonal components are explicitly modeled. In the Box–Jenkins approach, trend and seasonal effects are treated as nuisance parameters. These effects are removed from the series before any analysis can begin. As a result, state space methods provide an explicit structural framework for the decomposition of time series in order to diagnose all the dynamics in the time series data simultaneously. The Box–Jenkins methods are concerned with the short-term dynamics only and are therefore primarily concerned with forecasting only.

A successful application of ARIMA models requires the (differenced) time series to be stationary. However, as Durbin & Koopman (2001) pointed out: In the economic and social fields, real series are never stationary however much differencing is done. The investigator has to face the question, how close to stationary is close enough? This is a hard question to answer.' In state space methods, stationary of the time series is not required. Furthermore, missing data, time-varying regression coefficients and multivariate extensions are easily handled in the state space framework. This handling is relatively difficult in a pure ARIMA modeling context.

But studies on to assess the pattern of the number of rainy days in a month in monsoon season are scant. With these ends in view, the present study has been undertaken to detect the trend of the rainfall of the number of rainy days in monsoon and compare the state space representation of the model with the traditional decomposition method and Box-Jenkins method.

Methodology

Sources of data

This study is conducted of the rainfall of the Dhaka station taking the duration 1973-2008 on account. The monthly and daily data of rainfall for Dhaka station over 1973-2008 are collected from the meteorological department of Bangladesh. The secondary data were collected from meteorological department which is on monthly rainfall in monsoon over the period 1973-2008.

Description of variables

Meteorological department stores data on a sheet or computer's hard disc where the data are arranged in rows, which contain 12 months and columns, which contains 31 days. For the comprehension of the study, the data were sorted to get a single column. Then the daily data from January 1st 1973 to 31st December 2008 were arranged on a single column. Similarly the monthly data from January 1973 to December 2008 were also arranged on a single column for each variable, the monthly counts of rainy days in monsoon.

Statistical software used

There are different kinds of statistical software which can be used to perform the analysis. The available statistical software which can be for the analysis is as follows: Eviews-3.1, MINITAB, R, Gauss, SPSS, S-plus and Microsoft Excel etc. In this Eviews-3.1 was used because it has both windows and syntax option and finally it is very easy to handle. Microsoft Excel, SAS and SPSS were also used.

The traditional time series modeling approach such as classical decomposition methods and Box-Jenkins methods were used for analyzing the data. But both the two approach have some problems. For this reasons state-apace representation of the time series was applied to estimate trend and forecasting time series because it is very easy to estimate and forecasting in relative to the other two traditional time series modeling. Time varying method was also used for detection of trend and forecast future value to get the better estimated value than the traditional methods (Cooley & Prescott, 1973).

Results and Discussion

Estimation of trend in the absence of seasonality by classical decomposition method

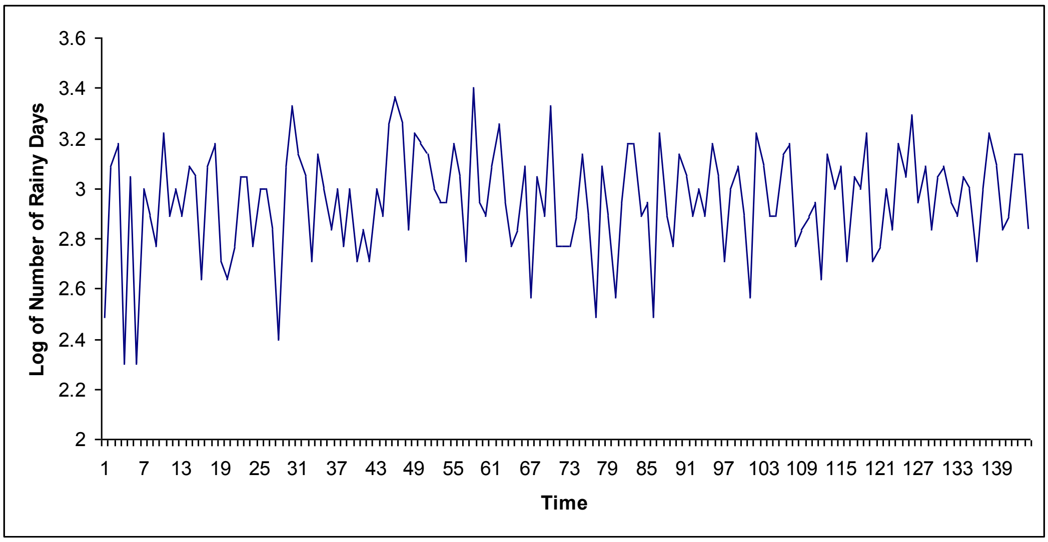

Fig. 1. Line graph of the log of the number of rainy days.

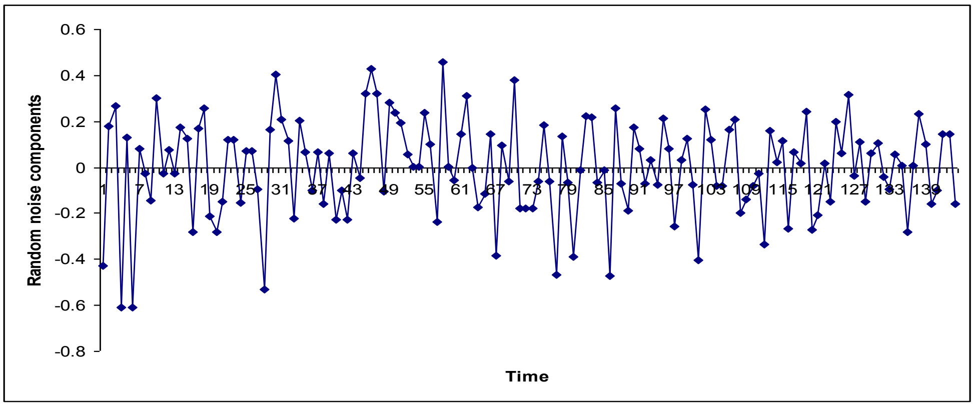

Fig. 2. Line graph of the random component of the fitted model.

Figure 1 reveals that the log of the number of rainy days shows almost no trend. It also reflects that the log of the number of rainy days have almost no changes over time.

In the absence of seasonal component the model becomes

$X_t = m_t + Y_t, t = 1, ...n$, Where m_t is the trend component and Y_t is the random noise components. We can estimate trend component by polynomial fitting (Duncan & Horn, 1972) i.e.

$m_t = a_0 + a_1 t$, Where a_0 intercept and a_1 is the slope.

For my research the model is $\ln(NRD) = a_0 + a_1 t$ (Harvey, 1989), Where NRD stands for the number of rainy days

Thus the fitted model is $\ln(NRD) = 2.911 + 0.00056t$

Above plot of the random noise components of the log of the number of rainy days reveals that there is no systematic pattern (Fig. 2). So from the decomposition part of the log of the number of rainy days the number of rainy days in the future years in monsoon season is predicted.

Prediction by using classical method

The predicted value of the number of rainy days in monsoon in the next year monsoon season is as follows:

Table1. Predicted value by using classical method (Gersch & Kitagawa, 1983).

Name of the month	Lrd	Predicted value
June-2009	2.992392158	19.93331
July-2009	2.992955598	19.94454
August-2009	2.993519038	19.95578
September-2009	2.994082478	19.96703

Table 1 show the number of rainy days was increased very slowly from one month to another in the monsoon season.

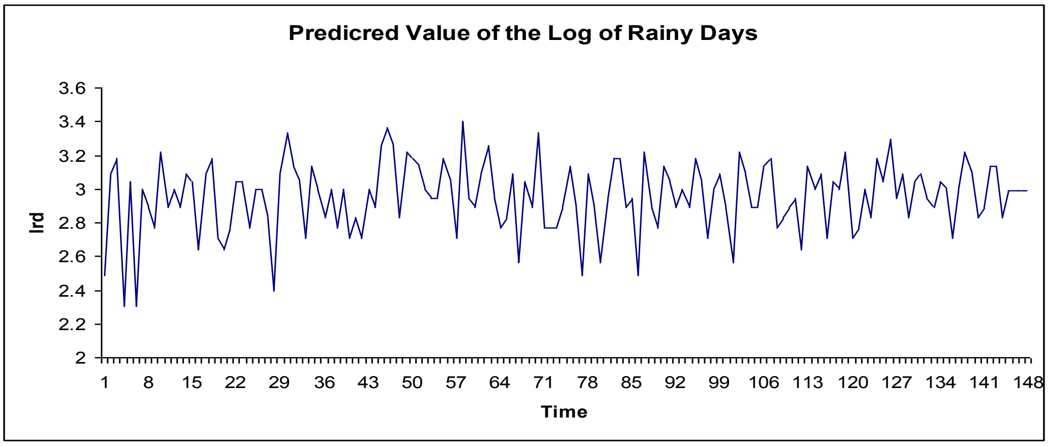

Fig. 3. Line graph of the predicted value by using classical method. where lrd stands for the log of the number of rainy days.

Estimation of ARIMA Model

A time series plot of the log of the number of rainy days (Fig. 4), it is clear that there exists almost no trend and the data shows no specific pattern. Hence the original series is nearly stationary but not so smoothly the data are stationary. A plot of the sample autocorrelation function, ACF and the sample partial autocorrelation function, PACF is shown in another figure 5 below. The graph of ACF of the series displays no specific pattern in the size of ACF values which is typical pattern of stationary series. But to ensure the series is stationary it was checked by using different kinds of diagnostic checking. From the above plot (Fig. 4) it is shown that there is almost no trend which is close to stationary but not so smoothly stationary and there is no seasonality in the data. Fig. 5 of the sample ACF and PACF reveals that the data follows typically pattern of stationary time series.

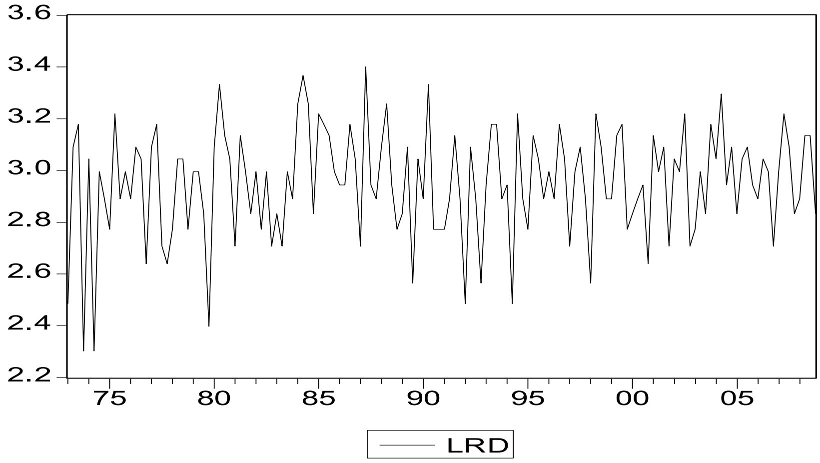

Fig. 4. Log of the number of rainy days, where lrd stands for the log of the number of rainy days.

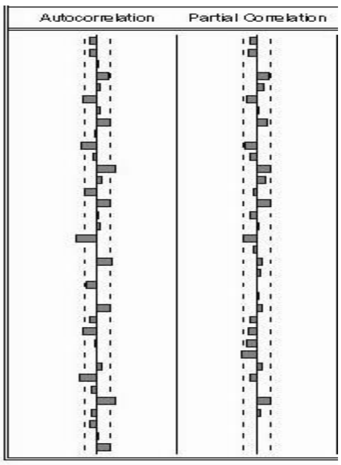

Fig. 5. Sample ACF and PACF of the log of the number of rainy days.

Unit Root Test

At the formal level, stationary can be checked by finding out if the time series contains a unit root. The Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF), tests can be used for this purpose (Dickey & Filler, 1979).

The actual procedure of implementing the DF tests involves the null and alternative hypothesis as follows

$H_0 : \partial = 0$ means there is a unit root that is the time series is non-stationary.

$H_1 : \partial < 0$ means there is no unit root that is the time series is stationary.

Decision rule

If the calculated value is greater than the critical value, the null hypothesis of the unit root is rejected i.e. the series is stationary, otherwise the null hypothesis of the unit root is accepted that is the series is non-stationary.

Table 2. ADF unit root test of the log of the number of rainy days.

ADF Test Statistic	-13.20083	1% Critical Value*	-4.0245
		5% Critical Value	-3.4417
		10% Critical Value	-3.1452

Table 2 depicts that the original time series data of the log of the number of rainy days is not smoothly stationary. In order to further smooth the fluctuation existing in the data the original data is differenced to get the stationary series. Since the modulus of the calculated value is greater than the critical value, the hypothesis of there is no unit root in the series is rejected. It presents that the original series is stationary. But if the first difference is taken, the series pattern shows more appropriately stationary. The Fig. 6, is the plot of the first difference of the original series. The first difference shows that the data shows completely no systematic pattern which shows completely random pattern. It reveals that the data of the first difference of the log of the number of rainy days is stationary.

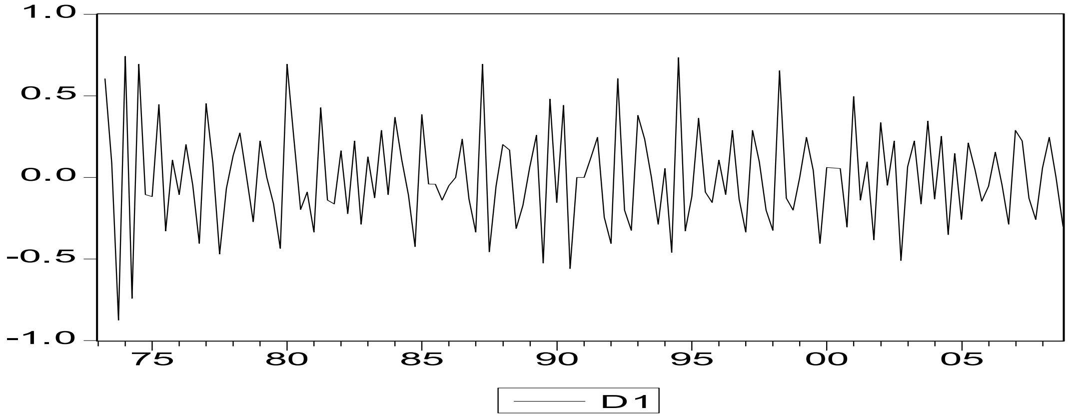

Fig. 6. Line graph after taking the first difference of the original series.

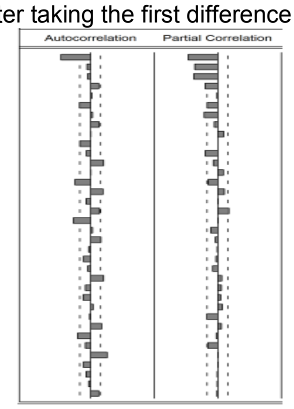

Fig. 7. Correlogram of the sample ACF and PACF of taking the first difference.

Table 3. ADF unit root test of the log of the number of rainy days

ADF Test Statistic	-20.30503	1% Critical Value*	-3.4773
		5% Critical Value	-2.8818
		10% Critical Value	-2.5774

Table 3 describes that the first difference of the original time series data at lag 1 is stationary i.e. the time series of the log of the number of rainy days is stationary and we modeled a stationary ARIMA model. Thus we will stop further test. Now we will identify the tentative model for the transformed series by inspection of ACF and PACF. It is obvious from the sample ACF of the differenced series the most dominating spike at lag 1 is outstanding. On the other hand, the spike at lag 1, 2 and 3 are statistically significant for PACF.

Now we consider the difference types of tentative model as much as possible from which we select the best model using the model selection criteria. Since the characteristics of a good ARIMA model is parsimonious ignoring the higher order of p and q, the tentative model on the basis of model selection criterion are as follows:

Table 4. Comparison of different ARIMA model.

Model	AIC	R^2	Adjusted R^2
ARIMA(1,1,1)	-0.277	0.537	0.53
ARIMA(0,1,1)	-0.282	0.542	0.539
ARIMA(1,1,0)	0.21	0.23	0.23
ARIMA(2,1,2)	0.32	0.16	0.156

From the above Table 4, we see that for the model ARIMA (0, 1, 1) has smaller AIC and R^2 and adjusted R^2 are high than the other model. So the model ARIMA (0, 1, and 1) is the best tentative model and we can also use this model for our foresting purpose.

Prediction by Using ARIMA Model

The ultimate application of ARIMA model is to forecast future value of a time series. The forecasting of the tentative ARIMA model is used when the finite number of past observations is available. The forecasted value of the ARIMA (0, 1, 1) model are as follows:

Table 5. Estimated value by using ARIMA model.

Name of the predicated month	lrd	Predicted value of the number of rainy days
June-2009	2.8832	18
July-2009	2.9740	20
August-2009	3.0292	21
September-2009	2.7510	16

Table 5 illustrates that the number of rainy days in monsoon follows the same pattern as the original value. It is clear that the number of rainy days decreasing as the time passes in relative to the original value.

From the Fig. 8, the one step predictors of the log of the number of rainy days for ARIMA model in the above figure together with the actual data. For this model, the predictors follow the movement of the original data quite well but the predicted value is small in relative to the original data.

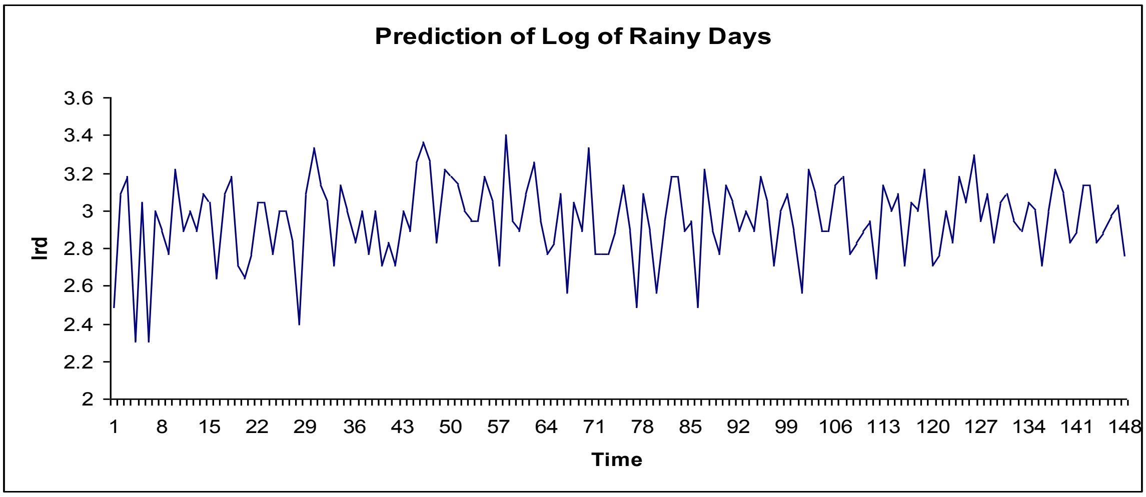

Fig. 8. One step prediction for the log of the number of rainy days.

State-space representation of the model

The trend model with deterministic level and stochastic slope coefficient are as follows

$$y_t = \mu + v_t + \varepsilon_t, \quad where \ \varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$$

$$v_{t+1} = v_t + \xi_t, \quad where \ \xi_t \sim nid(0, \sigma_\xi^2)$$, where

y_t represent the log of the number of rainy days. The estimated component of the state-space representation (Durbin & Koopman, 1992) of the model is as follows:

At convergence the value of the log-likelihood function equal -30.48063. The maximum likelihood estimates of the variance of the irregular components are $\hat{\sigma}_\varepsilon^2 = 0.018469$, and the maximum likelihood estimate of the state disturbance variance is $\hat{\sigma}_\xi^2 = 4.83E^{-7}$. The estimated value of the level $\hat{\mu} = 2.61$ and the final estimated value of the slope component are $\hat{v} = .00241$. The state variance of the slope component is almost zero, meaning that the value of the slope hardly chance over time

The below Fig. 9, contains the separate development of the slope over time. At initial stage the change of the slope over time is considerable but after few periods it is clear that the slope is effectively constant. This is consistent with the close-to-zero disturbance variance for this component.

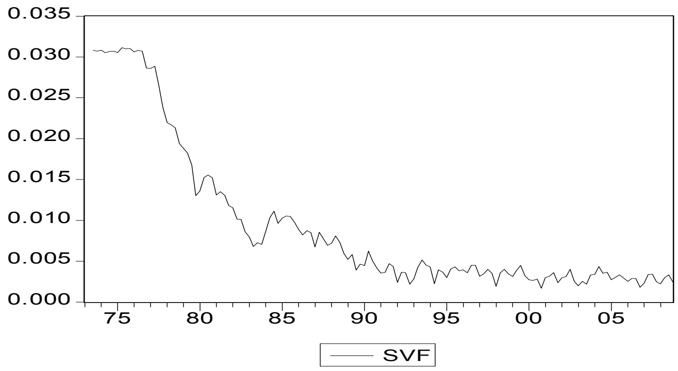

Fig. 9. Slope of stochastic linear Trend model.

The predicted value of the number of rainy days in the state space representation

If we plot the predicted value of the log of the number of rainy days then we get the graph as follows

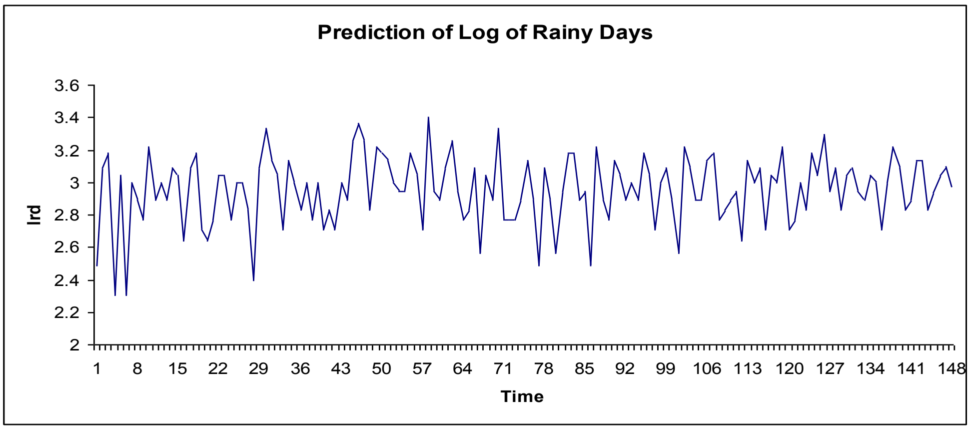

Fig 10. Predicted value of the log the number of rainy days.

Table 6. Predicted value by using state apace representation.

Name of the predicted month	lrd	Predicted value of the number of rainy days
June-2009	2.93335	19
July-2009	3.044642	21
August-2009	3.099804	22
Septerber-2009	2.966828	19

The one step prediction of the log of the number of rainy days in the above Figure 10 together with the actual data. For this model, the predictors follow the same movement of the original data quite well. From the above analysis it can be concluded that the predicted value of the log of the number rainy days in the classical decomposition method are very slowly increasing which are contradict the original data pattern. In the original data the values are not always increasing. The original data in the monsoon mostly follows that the first month it has small value then the next month it increase then the next month it decrease and then in the last month in the monsoon it decrease. But the predicted value in the classical method does not follow the same pattern as the original value. On the other hand the predicted value of the log of the number rainy days in the Box-Jenkins method follows the same pattern of the original data and provides a good predicted value. But the predicted value is relatively small than the original value. Again the predicted value of the log of the number rainy days in the state-space representation methods follows the same pattern of the original data value. The predicted value which is obtained by state space representation is very similar to the original data value. For that state-space representation of the model gives better result in time series analysis for forecasting.

Conclusion

The analysis of rainfall behavior is important to the water resource management as well as agricultural department. The predictions of rainy days in monsoon contribute to the proper policy implementation in the agricultural fields. For this purpose this study was undertaken considering variable number of rainy days in monsoon per month. To detect trend and forecast future value several time series methods are available. The traditional time series methods are classical decomposition method and Box-Jenkins method. But both the methods have some limitations. In the classical method the observations of the dependent variables y_t and the independent variable time t are assumed independent but in time series analysis they are interrelated. Again in the classical method the slope coefficient is always constant value but in time series analysis it should not be a constant value. On the other hand in the Box-Jenkins method a successful application requires the time series to be stationary. Again Box-Jenkins method concerned with the short term dynamic only and is therefore primarily concerned for forecasting. To overcome the above limitations of the traditional time series methods, time varying method is used. Because in time varying method the slope coefficients vary over time and the series need not to be stationary for detection of trend and forecasting future value.

From the analysis of the data it is clear that in the classical method the predicted value increases over time very slowly which contradict with the original data pattern. The line graph of the predicted value does not follow the preceding pattern of the graph. In the Box-Jenkins method the predicted values provides the similar pattern of the original data. The line graph of the predicted values also shows a reasonable similar pattern of the original data series of the number of rainy days in monsoon. But in the time varying method the predicted values give the better result because its values are much closer to the original data set than the Box-Jenkins method and the classical decomposition method. The line graph of the predicted values also suggests a better result provides by the time varying method than the traditional time series method. Because time varying method overcomes the limitations of the constant slope coefficient and make the series stationary. It can be concluded from the analysis that the numbers of rainy days are decreasing over time in the monsoon.

References

Box, G. E. P. and Jenkins, G. M. (1976), Time Series Analysis: Forecasting and Control, Revised Edition, Holden-Day, San Francisco.

Brockwell, P. J. and Davis, R. A. (1996), Introduction to Time Series and Forecasting, Springer-Verlog, New York.

Cooley, T. F., and Prescott, E. C. (1973), “Varying Regression : A Theory and Some Application,” Annals of Economic and Social Measurement, 2, 463-473

Dickey, D. A. and Filler, W. A. (1979), Distributions of the estimators for autoregressive time series with unit root, J. Amer. Stat. Assoc., 74, 427-431.

Duncan, D. B., and Horn, S. D. (1972), “Linear Dynamic Regression Estimation From the Viewpoint of Regression Analysis,” Journal of the American Statistical Association, 67, 845-821.

Durbin, J. and S. J. Koopman (2001). Time Series Analysis by State Space Methods. Number 24 in Oxford Statistical Science Series. Oxford: Oxford University Press.

Durbin, J., & Koopman S. J. (1992) Filtering, smoothing and estimation for time series models when the observations come from exponential family distributions, London School of Economics and Political Science, London. Unpublished.

Gersch, W. and Kitagawa, G. (1983) The prediction of time series with trend and seasonalities. J. Bus. Econ. Statist., 1, 253-264.

Green, J. R. (1964) “A Model for Rainfall Occurrence”, Journal of Royal Statistical Society, Vol. 29, pp. 155-156.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econimetrica 57, 357-84.

Harvey, A. C. (1989) Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge: Cambridge University Press.