RESEARCH PAPER | Statistics | Volume : 5 | Issue : 8 | August 2015 | ISSN - 2249-555X

Modified Auto Regression Model for Race of Adults

KEYWORDS: Auto regression, Modified Auto regression models, RMSE.

S.C. Thasleema
Department of Statistics, S.V. University, Tirupati-517502.

B. Sarojamma
Department of Physics, S.V. University, Tirupati- 517502.

ABSTRACT *Auto regressive model is relation between present time series variable and past time series values. Generally regression gives relation between dependent variable and independent variables. In time series a univariate variable relates chronologically. In this paper two modified auto regressive models by incorporating time to first modified auto regressive model and second modified autoregression model is incorporated by adding time and also dividing factorials to past time series values. Root mean square error criteria are used to choose best among three models i.e. auto regressive model, modified autoregressive model-1 and modified auto regressive model-2. The data"percent of people 25 years and over who have completed high school or college, by race, Hispanic origin and sex, for the period, of 23years i.e., 1990-2014(non institutionalized population)". Among the fitted 15 models, 5 for auto regressive, 5 for modified auto regressive model-1 and 5 for auto regressive model-2 for orders 1, 2, 3,4 and 5. Among three auto regressive models, modified auto regressive model-2 is the best for the given data.Among 15 models, 5th order modified auto regressive model-2 is the best.*

1.0 INTRODUCTION

Many univariate time series models are there in literature. Popular univariate time series models are exponential smoothing models, auto regression model, regression model, generalized regression model, dynamic regression model, auto regressive integrated moving average models, auto regressive conditionally Heteroscedastic (ARCH), generalized auto regressive conditionally Heteroscedastic (GARCH) etc.

ARIMA: $$Y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$$

ARCH: $\sigma_t^2 = \beta_0 + \beta_1 Y_{t-1}^2 + \ldots\ldots. + \beta_p Y_{t-p}^2$

GARCH: $\sigma_t^2 = \beta_0 + \beta_1 Y_{t-1}^2 + \ldots\ldots. + \beta_p Y_{t-p}^2 + \gamma_1 \sigma_{t-1}^2 + \gamma_2 \sigma_{t-2}^2 + \ldots.. + \gamma_q \sigma_{t-q}^2$

Fong-Linchu in his article, "A fractionally integrated auto regressive moving average approach to forecasting demands', gives tourism forecasting by incorporate integrated ARMA (p, d, q) ARIMA models. In his paper he compared previous models with his incorporated model. An unobserved component model for multi-rate forecasting of telephone call demand: the design of a forecasting support system" given by Wlodek Tych et al. in their paper outlines both the methodological and algorithmic aspects of the modeling, forecasting and signal extraction procedures, including the design and implementation of forecasting support software with a specially designed graphical user interface with in the MATLAB computing environment. James W. Taylor in his article "Volatility forecasting with smooth transition exponential smoothing"gives a new smooth transition adaptive exponential smoothing a variety of non-linear GARCH models, where proposed for volatility forecasting.

Gardner Jr., stated an article "exponential smoothing; the state of the art". Ameen and Harrison in their article "Discounted weighted estimation". They extended the methods to Bayesianforecasting, kalman filtering and state space modeling and compared. Brown, R.G., Smoothing gives forecasting and prediction of discrete time series.

2.0METHODOLOGY

Many Regression models are the in literature. Some of the regression models are simple linear model, multiple linear model, polynomial model, Poisson regression, Auto regression model of order '1' i.e. Markoff"s Series, Auto regression model of order '2'i.e. Yule's Series, Auto regression of order 'p', etc.In time series, regression is of the form Auto regression and in Auto regression equation, present time series Y_t is depend upon past time series valuesY_{t-1}, Y_{t-2}….Y_{t-p},gives equation of order 'p'.

Volume : 5 | Issue : 8 | August 2015 | ISSN - 2249-555X

Autoregression(p):

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots\cdots + \beta_p Y_{t-p} \cdots(1)$$

where Y_t, Y_{t-1}…$Y_{t-p}$is time series values at time t, t-1, t-2 …t-p respectively.

$\beta_o,\beta_1,\ldots..\beta_p$ are constants to be estimated.

Auto regression equation is modified by adding time to't'. Estimation of dependent variable Y_t(present time series value at time't')

Using time't' and past time series values Y_t, Y_{t-1}… Y_{t-p}at time points t-1,t-2…t-p respectively.

Regression model with time and time series values

$$Y_t = b_0 + b_1 t + b_2 Y_{t-1} + \cdots\cdots + b_p Y_{t-p} \cdots\cdots(2)$$

Another regression model is introduced by taking time't' and past time series values Y_t, Y_{t-1}… $Y_{t-p}$and their respective coefficients are informed of variable and respective factors.

$$Y_t = b_0 + b_1 t + \frac{b_2}{2!} Y_{t-1} + \frac{b_3}{3!} Y_{t-2} + \cdots\cdots + b_{p+1} \frac{Y_{t-p}}{(p+1)!} \cdots\cdots(3)$$

Here Y_t is time series value at time't' (dependent variable)

Y_t, Y_{t-1}… $Y_{t-p}$are independent time series variables at time t-1, t-2…t-p (independent)

The method used for estimation of coefficients for three auto regression equations using ordinary least squares method of estimation.

Root Mean Square Error (RMSE) Criteria:RMSE is obtained estimation by using positive square root of mean square errors.This criterion is used especially when we have two or more models compared at a time and choose which model is the best among several models.

Error is defined as difference of original value (u_t) and estimated value (F_t). Mean of squared error gives mean square error (MSE). Positive square root of mean squares gives root mean square error (RMSE).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(u_i - F_i)^2}{n}}$$

3.0EMPIRICAL INVESTIGATIONS

Auto regression models gives relation between present time series value (Y_t) with past time series values Y_t, Y_{t-1}… Y_{t-p}in regression form as follows

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots\cdots + \beta_p Y_{t-p}$$

Present time series value Yt is depend upon past time series of time 4 or 5, after that it is almost depend upon Y_6, Y_7……..Y_{t-p}.

We estimated equation of auto regression of order 1,2,3,4 and 5 using time 't' as independent variable from 1990-2014 (25 years) and time series variable is percent of people 25 years and over, who have completed high school or college, by race, Hispanic origin and sex, selected period 1990-2014(non institutionalized population).

The fitted equations of auto regression for different orders are

$$AR(1) : Y_t = \beta_0 + \beta_1 Y_{t-1}$$

=85.255+0.835 Y_{t-1}

$$AR(2) : Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}$$

=85.312+0.961 Y_{t-1} - 0.147 Y_{t-2}

$$AR(3) : Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3}$$

=85.337+0.911 Y_{t-1} + 0.078 Y_{t-2} - 0.216 Y_{t-3}

$$AR(4) : Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4}$$

=85.332 + 0.852 Y_{t-1} + 0.079 Y_{t-2} + 0.063 Y_{t-3} - 0.286 Y_{t-4}

Volume : 5 | Issue : 8 | August 2015 | ISSN - 2249-555X

$$AR(5): Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \beta_5 Y_{t-5}$$

= 85.431 + 0.806 Y_{t-1} + 0.069 Y_{t-2} + 0.069 Y_{t-3} - 0.123 Y_{t-4} - 0.174 Y_{t-5}

Modified auto regression model by adding time '1' to auto regression equation. The modified auto regression equation is

$$AR'(1): Y_t = \beta_0 + \beta_1 Y_{t-1}$$

=85.255 + 0.835 Y_{t-1}

$$AR'(2): Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}$$

=85.312 + 0.961 Y_{t-1} - 0.147 Y_{t-2}

$$AR'(3): Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3}$$

=85.337 + 0.911 Y_{t-1} + 0.078 Y_{t-2} - 0.216 Y_{t-3}

$$AR'(4): Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4}$$

= 85.332 + 0.852 Y_{t-1} + 0.079 Y_{t-2} + 0.063 Y_{t-3} - 0.286 Y_{t-4}

$$AR'(5): Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \beta_5 Y_{t-5}$$

= 85.431 + 0.806 Y_{t-1} + 0.069 Y_{t-2} + 0.069 Y_{t-3} - 0.123 Y_{t-4} - 0.174 Y_{t-5}

Another modified auto regression model is developed by adding time't' and coefficients are also modified as coefficient divided by factorials. The fitted second modified auto regression equation is

$$AR''(p): Y_t = \beta_0 + \beta_1 t + \beta_2 \frac{Y_{t-1}}{2!} + \beta_3 \frac{Y_{t-2}}{3!} + \cdots\cdots + \beta_{p+1} \frac{Y_{t-p}}{(p+1)!}$$

Modified auto regression equations for different orders are as follows

$$AR''(1): Y_t = \beta_0 + \beta_1 t + \frac{\beta_2}{2!} Y_{t-1}$$

= 85.255+0.187*t + 0.835(Y_{t-1})/2!

$$AR''(2): Y_t = \beta_0 + \beta_1 t + \frac{\beta_2}{2!} Y_{t-1} + \frac{\beta_3}{3!} Y_{t-2}$$

=85.312+0.18*t+ 0.961 (Y_{t-1})/2! - 0.147 (Y_{t-2})/3!

$$AR''(3): Y_t = \beta_0 + \beta_1 t + \frac{\beta_2}{2!} Y_{t-1} + \frac{\beta_3}{3!} Y_{t-2} + \frac{\beta_4}{4!} Y_{t-3}$$

=85.337+0.174*t+0.911(Y_{t-1})/2!+0.078(Y_{t-2})/3!-0.216(Y_{t-3})/4!

$$AR''(4): Y_t = \beta_0 + \beta_1 t + \frac{\beta_2}{2!} Y_{t-1} + \frac{\beta_3}{3!} Y_{t-2} + \frac{\beta_4}{4!} Y_{t-3} + \frac{\beta_5}{5!} Y_{t-4}$$

= 85.332+0.169*t+0.852(Y_{t-1})/2! +0.079(Y_{t-2})/3!+0.063(Y_{t-3})/4!-0.286(Y_{t-4})/5!

$$AR''(5): Y_t = \beta_0 + \beta_1 t + \frac{\beta_2}{2!} Y_{t-1} + \frac{\beta_3}{3!} Y_{t-2} + \frac{\beta_4}{4!} Y_{t-3} + \frac{\beta_5}{5!} Y_{t-4} + \frac{\beta_6}{6!} Y_{t-5}$$

= 85.431+0.16*t+0.806(Y_{t-1})/2!+0.069(Y_{t-2})/3!+0.069(Y_{t-3})/4!- 0.123 (Y_{t-4})/5! - 0.174(Y_{t-5})/6!

The above equations are for orders 1,2,3, 4 and 5 of modified auto regression model-1 and modified auto regression model-2.

Root mean square error criteria (RMSE): A root mean square error criterion is used for choosing the best auto regression models of modified auto regression model-1 and modified auto regression model-2.

RMSE values for auto regression models, different orders of modified auto regression model-1 and different orders of modified auto regression model-2 are as follows.

Volume : 5 | Issue : 8 | August 2015 | ISSN - 2249-555X

Order	Auto regression model	Modified Auto regression model-1	Modified Auto regression model-2
1	69.2117	69.3949	35.9794
2	68.0612	68.2374	41.8432
3	65.4237	65.5938	39.4962
4	61.1364	61.3010	37.6482
5	**56.8787**	**57.0342**	**35.6256**

4.0 CONCLUSIONS

Auto regression model deals with present time series value Y_t is related with past time series values Y_t, Y_{t-1}… Y_{t-p}. Relation between u_t and Y_t, Y_{t-1}… $Y_{t-p}$is auto regression of order p [$AR_{(p)}$]. In this paper, we are introduced two univariate models by modifying auto regressive model. Generally auto regressive moving average model is of the form.

AR(p):

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots\cdots + \beta_p Y_{t-p}$$

where Y_t, Y_{t-1}… $Y_{t-p}$are time series values at time point t, t-1…t-p respectively.

β_o, β_1,…..β_p are constants to be estimated.

Modified auto regressive model-1 is estimated by adding one more independent variable time‘t’ to independent Y_t, Y_{t-1}… Y_{t-p}, andthe modified AR model-1 is as follows.

$$Y'_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots\cdots + \beta_p Y_{t-p-1} + \beta_{p+1} Y_{t-p}$$

The modified auto regressive model-2 is estimated by taking time‘t’ and past time series values divided by factorial values as follows

$$AR''(p) : Y_t = \beta_0 + \beta_1 t + \beta_2 \frac{Y_{t-1}}{2!} + \beta_3 \frac{Y_{t-2}}{3!} \cdots\cdots + \beta_{p+1} \frac{Y_{t-p}}{(p+1)!}$$

RMSE values for fitted auto regression, modified auto regression model-1, modified auto regression model-2 for orders 1,2,3,4and 5 are as follows

Order	Auto regression model (RMSE) values	Modified Auto regression model-1(RMSE) values	Modified Auto regression model-2(RMSE) values
1	69.2117	69.3949	35.9794
2	68.0612	68.2374	41.8432
3	65.4237	65.5938	39.4962
4	61.1364	61.3010	37.6482
5	**56.8787**	**57.0342**	**35.6256**

Among fitted auto regression models, modified auto regression model-2 is the best model compared with auto regression model, and modified auto regression model-1. Modified auto regression model-2 of order 5 is the best model for 25 years of datafor percent of people 25 years and over, who have completed high school or college, by race, Hispanic origin and sex, selected period 1990-2014 (non institutionalized population), using root mean square error.

REFERENCE REFERENCES | [1] Fong-Lin Chu,(2006). “Fractionally integrated auto regressive moving average approach toforecasting tourism demand”, Tourism Management, 29, pp.79-88. | [2] James W. Taylor, (2004). “Volatility forecasting with smooth transition exponential smoothing”, International journal of Forecasting, 20, pp. 273-286. | [3] Gardner Jr., E.S. 1985. “Exponential smoothing: the state of the art”, Journal of Forecasting 4, pp. 1-28. | [4] Ameen, J.R.M& Harrison, P.J.(1984).“Discounted weighted estimation”, Journal ofForecasting”, 3, pp. 285-296. | [5] Ministry of Human Resource Development, Govt. of India. Data.gov.in