Transformed Variable ARMA Model

P. Ramakrishna Reddy, B. Sarojamma*

Department of Statistics, S.V. University, Tirupati – 517502, Andhra Pradesh, INDIA.

*Corresponding Address:

saroja14397@rediffmail.com

Research Article

Abstract: There are many forecasting models like moving average MA (1), MA (2), weighted moving average, multiple moving average, simple exponential smoothing, double exponential smoothing, triple exponential smoothing, adaptive smoothing, auto regression AR (1), AR (2), ARMA, ARIMA, ARCH, GARCH, etc. In this paper we discussed about transformed variable ARMA models. ARMA model is a combination of auto regression with moving averages. Generally auto regression and moving averages are calculated with time series values. By taking log values and geometric mean ARMA model values smoothing the data. Arithmetic mean effects the extreme values where as geometric mean cannot effects extreme values as much as A.M. By taking logarithm and geometric mean transformations to the time series observations to perform ARMA models, we get logarithm ARMA model and geometric mean ARMA model. Logarithm ARMA model and geometric mean ARMA model are tested for goodness of fit by using Kolomogrov-Smirnov test. Mean Square Error (MSE) criterion is used for choosing best model among ARMA (1, 1), Logarithmic ARMA (1, 1) and geometric mean ARMA (1, 1) models.

Keywords: Autoregressive Moving Average, Geometric Mean Autoregressive Moving Average, Logarithmic Autoregressive Moving Average, Kolmogrov-Smirnov test, Mean Square Error.

1. Introduction

A sequence of numerical data points in successive order, usually occurring in uniform intervals. A time series is simply a sequence of numbers collected at regular intervals over a period of time. Time series data may be in the form of years, or months, days, hours, minutes. Forecasting plays an important role in atmospheric sciences, population growth, in urbanization for estimation of growth of houses, etc. Time series forecasting is used to predict future values based on previously observed values. Forecasting involves making the best possible judgment about some future event. In other words, Forecasting is a numerical estimate of an event for some future data that can be achieved with a specified level of support and are reproducible.

1.1 Moving average

The simplest way to smooth a time series is to calculate a simple moving average. The smoothed statistic S_t is then just the mean of the last k observations.

$$S_t = \frac{1}{k}\sum_{n=0}^{k-1} X_{t-n} = \frac{X_t + X_{t-1} + \ldots\ldots + X_{t-k+1}}{k}$$

$$= S_{t-1} + \frac{X_t - X_{t-k}}{k}$$

where the choice of an integer k > 1 is arbitrary. A small value of k will have less the smoothing effect.

1.2 Exponential Smoothing

Exponential smoothing is the special type of moving average method in forecasting. In simple moving average, the mean of the past k observations that mean the weight of k time series observations having equal value 1/k. Where as in exponential smoothing, if observations get older weights are also exponentially deceasing. The most recent observations will usually provide the highest weight value and observation is get older its weight is decreasing.

1.3 Single Exponential Smoothing

Single exponential smoothing is also called a simple exponential smoothing. The parameter in simple exponential smoothing is 'α'. If we estimate forecast value of some 't+1' point then the equation becomes

$$F_{t+1} = F_t + \alpha(Y_t - F_t)$$

where F_{t+1} = forecast for time point 't+1'

F_t = forecast for time point 't'

Y_t = our time series observation at time 't'

α = constant

α lies between 0 and 1.

$\alpha + \beta = 1$

1.4 ARMA model

The basic elements of autoregressive (AR) and moving average (MA) models can be combined to produce a great variety of models. Combination of p^{th} order autoregressive model and q^{th} order moving average model called an ARMA (p, q) model and is expressed as

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

where Y_t is time series value at time 't'

$Y_{t-1}, Y_{t-2} \ldots Y_{t-p}$ is time series values at time t-1, t-2 …t-p.

$e_t, e_{t-1}, e_{t-2} \ldots e_{t-q}$ are error terms at time t, t-1, t-2 … t-q respectively

$\phi_1, \phi_2 \ldots \phi_p$ are autoregressive coefficients

$\theta_1, \theta_2 \ldots \theta_q$ are error constants.

In the present study, we fitted ARMA (1, 1) model and we transform the data by logarithms and geometric moving averages, then fitted ARMA (1, 1) models of these data. These models are compared by using error measures like mean absolute error, and mean square error.

2. Methodology

In general ARMA is mixture of auto regression with moving average models. Autoregressive moving average model is also a forecasting model with time‘t’ as independent variable with one dependent variable time series variable ‘u_t’. By taking log and geometric means, we estimate equations for ARMA.

ARMA (1, 1) model is a combination of auto regression of order ‘1’ combined with moving average of order ‘1’. Auto regression of order ‘1’ is passing equation present time series value ‘Y_t’ with one past time series value Y_{t-1} is combination of present and past error terms. ARMA (1, 1) is of the form

$$Y_t = c + \phi_1 Y_{t-1} + e_t - \theta_t e_{t-1}$$

where Y_t is time series value at time ‘t’

Y_{t-1} is time series value at time t-1

e_t, e_{t-1} are error terms at time t, t-1 respectively

ϕ_1 is autoregressive coefficient

θ_t is error constant.

2.1 Log transformed variable ARMA model

We take log to time series variable data and is considered as main time series data, for performing ARMA (1, 1) model.

Data may be as follows

time	time series value	transformed time series value
t	Y_t	$\log Y_t = u_t$
t_1	Y_1	u_1
t_2	Y_2	u_2
$\vdots$	$\vdots$	$\vdots$
t_n	Y_n	u_n

By taking time‘t’ as independent variable and time series values as ‘u_t’, we perform autoregressive moving average (1, 1) then we get the model of the form

$$u_t = \phi_1 u_{t-1} + c + e_t - \theta_t e_{t-1}$$

where u_t is log Y_t, log transformed time series value at time‘t’

u_{t-1} is log transformed time series value at time ‘t-1’

$u_{t-1} = \log Y_{t-1}$

c is a constant

e_t, e_{t-1} are error terms at time ‘t’ and ‘t-1’

we get predicted values of log transformed model in the form of log, for taken out log we take antilog to predicted values and take that as estimated values.

Antilog u_t = estimated value of Y_t

2.2 Length of moving averages

Length of moving average is calculated by using the following steps

Step 1: For original data draw line graph by taking time on X-axis and time series value on Y- axis.

Step 2: Point out the peaks in graph.

Step 3: List out periods of different cycles exhibited by the data.

Step 4: Calculate arithmetic mean of periods of different cycles exhibited by the data. A.M. gives length of moving averages.

2.3 Geometric mean transformed variable ARMA model

Generally Arithmetic means are using for calculation of moving averages. Mean or arithmetic mean plays an important role in time series analysis and forecasting methods. Mean is affected by extreme observations. Whereas G.M is not much affected as A.M. Geometric mean is n square root of product of x_1, x_2 … x_n observations.

$$G.M. = \sqrt[n]{x_1.x_2 \cdots x_n} = \left(x_1.x_2 \cdots x_n\right)^{1/n} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

where x_1, x_2 … x_n are n time series observations. Geometric mean is not affected as arithmetic mean due to extreme observations. ARMA model can be written in the form

$$ARMA(1,1): Y_t = c + \phi_1 Y_t + e_t - \theta_1 e_{t-1}$$

The above ARMA (1, 1) is obtained by combining auto regression of order ‘1’ with moving average of order ‘1’.

Geometric mean transformed ARMA model is obtained by first transforming time series observations into geometric averages. If you compute 3 term geometric averages, we cannot obtain time series values for the beginning 2 terms. Forecast also possible for future years by using these geometric means. Data converted for geometric mean variable is

Time (t)	Time series value (u_t)	Geometric mean for time series value
t_1	u_1	
t_2	u_2	G_m
t_3	u_3	G_{m+1}
$\vdots$	$\vdots$	$\vdots$
t_n	u_n	G_n

where m is number of observations in geometric average

G_m is m^{th} geometric average

$$G_m = \sqrt[m]{x_1 x_2 \cdots x_m}$$

Now we take time series values as geometric means for time ‘t’. If we perform ARMA (1,1) for this data, we get equation of the form

$$G_{T+1} = c + \phi_1 G_T + \varepsilon_T - \theta_1 \varepsilon_{T-1}$$

where G_{T+1} is geometric mean time series observation at time T+1

G_T is geometric mean time series observation at time 'T'

ε_T , ε_{T-1} are error terms at times T, T-1 respectively

ϕ_1 is parameter for auto regression of geometric variables

θ_1 is a parameter for error terms

c is a constant

2.4 Kolmogrov - Smirnov test

Various steps are involved in performing K-S test.

1. The data consist of a random sample X_1, X_2 …. X_n of size n associated with some unknown distribution function, denoted by F(x).
2. The sample is a random sample.
3. Let S(x) be the empirical distribution function based on the random sample X_1, X_2 …. X_n. Let $F^*(x)$ be a completely specified hypothesized distribution function.
4. Let the test statistic T be greatest (denoted by "sup" for supremum) vertical distance between S(x) and $F^*(x)$. In symbols, we say

   For testing $T = \sup_x |F^*(x) - S(x)|$

5. $H_0 : F(x) = F^*(x)$ for all x from $-\infty \; to \; \infty$

   $H_1 : F(x) \neq F^*(x)$ for at least one value of x
6. If T exceeds the 1-α quantile as given by Table, then we reject H_0 at the level of significance α. The approximate p-value can be found by interpolation in Table.

Kolmogrov-Smirnov test is performed for testing good fit of ARMA.

2.5 Mean square error (MSE)

Mean of squared error terms gives mean square error. If we have two or more models, which model is the best can be determined by using MSE criteria. Formula for MSE is as follows

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$$

where n is number of observations

$\sum_{t=1}^{n} e_t^2$ is sum of squares of error terms 'e_t'

e_t is error term at time 't'

Among many models, a model which possesses least MSE is the best model.

$$MSE\ of\ ARMA(1,1) = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \widehat{Y}_t\right)^2$$

$$MSE\ of\ \log ARMA(1,1) = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \widehat{u}_t\right)^2$$

$$MSE\ of\ GM\ ARMA(1,1) = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \widehat{G}_t\right)^2$$

If MSE of GM ARMA (1, 1) is least when compared with MSE of ARMA (1, 1) and MSE of log ARMA (1, 1) model, then GM ARMA (1, 1) is the best model among the three models.

3. Empirical investigations

In this ARMA model, we are using transformed variables x_i's, for which we use two transformations, they are log and geometric mean of the original values. We compare ARMA (1, 1), log transformed variable ARMA (1, 1) and geometric mean observations for original variable ARMA (1, 1) model.

The above three ARMA models are tested for its good fit using Kolmogrov-Smirnov one sample test. By using mean square error criteria we can determine which model is best among these three models.

3.1 ARMA (1, 1):

The combinations of auto regression of order'1' with moving average of order '1'. ARMA (1, 1) and the fitted equation for the given data is

Y_t (general): ARMA (1, 1): $Y_t = 78.890 - 0.999Y_{t-1} + 0.989e_{t-1} + e_t$

Table 1

Year	Yield	Estimated	Error	(Error)^2
1993	29.4	27.4998	1.9002	3.6108
1994	27.1	27.3954	-0.2954	0.0873
1995	27.9	27.5386	0.3614	0.1306
1996	26.1	27.3185	-1.2185	1.4847
1997	25.9	27.5454	-1.6454	2.7073
1998	25.7	27.2809	-1.5809	2.4992
1999	29.1	27.4893	1.6107	2.5944
2000	28.8	27.1218	1.6782	2.8164
2001	28.3	27.4368	0.8632	0.7451
2002	27.6	27.0978	0.5022	0.2522
2003	24.7	27.3965	-2.6965	7.2711
2004	27.1	27.1413	-0.0413	0.0017
2005	27.5	27.2664	0.2336	0.0546
2006	27	27.0823	-0.0823	0.0068
2007	26.3	27.2233	-0.9233	0.8525
2008	25	27.0532	-2.0532	4.2156
2009	29.7	27.2001	2.4999	6.2495
2010	27.8	26.8857	0.9143	0.8359
2011	28	27.1887	0.8113	0.6582
2012	26.1	26.8376	-0.7376	0.5441
			Total	37.6180
			MSE	1.8809

The above table-1 explains yield as time series values, estimated values and also mean square error values.

3.2 Log transformed ARMA (1, 1)

We transformed original variable by using 'log', we get log $x_t = u_i$. Using u_t, we estimated ARMA (1, 1) equation and from that equation, we are estimated the values. Log transformed ARMA (1, 1) is as follows

$Y_t = 2.258 - 0.999Y_{t-1} + 0.99e_{t-1} + e_t$

3.3 Geometric mean ARMA (1, 1) model We estimated geometric mean of line length for original data. That geometric mean value is taken as original data, we fitted geometric mean ARMA (1, 1) model and we predicted the observations.

Geometric mean ARMA (1, 1) model is expressed GMARMA (1, 1):
Y_t (with GMA): $Y_t = 7.426 + 0.514\ Y_{t-1} + 0.934\ e_{t-1} + e_t$

Table 2

Year	Yield	Predicted log values	Estimated Log ARMA	$(Error)^2$	Predicted G.M Yield	Estimated GMARMA	$(Error)^2$
1993	29.4	1.4388	27.4663	3.7392			
1994	27.1	1.4373	27.3716	0.0738			
1995	27.9	1.4392	27.4916	0.1668	27.1650	27.1320	0.0011
1996	26.1	1.4361	27.2961	1.4306	26.5613	27.1668	0.3667
1997	25.9	1.4392	27.4916	2.5332	26.5250	26.5096	0.0002
1998	25.7	1.4355	27.2584	2.4285	26.9951	26.8501	0.0210
1999	29.1	1.4383	27.4347	2.7733	27.6338	27.1967	0.1910
2000	28.8	1.4331	27.1082	2.8623	28.1918	27.7714	0.1768
2001	28.3	1.4375	27.3842	0.8387	27.8674	28.0616	0.0377
2002	27.6	1.4327	27.0832	0.2671	27.0953	27.3828	0.0827
2003	24.7	1.4369	27.3464	7.0034	26.7938	26.9066	0.0127
2004	27.1	1.4333	27.1206	0.0004	26.6247	26.9070	0.0797
2005	27.5	1.4349	27.2207	0.078	26.7607	26.6730	0.0077
2006	27	1.4323	27.0583	0.0034	26.7010	27.0777	0.1419
2007	26.3	1.4342	27.1769	0.769	26.6886	26.6329	0.0031
2008	25	1.4318	27.0271	4.1093	27.0451	27.0221	0.0005
2009	29.7	1.4338	27.1519	6.4929	27.3574	27.1807	0.0312
2010	27.8	1.4293	26.872	0.8612	27.7212	27.4866	0.0550
2011	28	1.4335	27.1331	0.7515			
2012	26.1	1.4285	26.8225	0.5221			
			Total	37.7045		Total	1.2092
			MSE	1.8852		MSE	0.0756

Above table-2 contains eight columns, first column tells about time‘t’ in years, second column tells about time series values (yields), third column tells about logarithm of predicted time series values, forth column gives estimated values using log transformed ARMA (1, 1) model, fifth column says about error squares. By using average of fifth column, we find mean square error value of log transformed ARMA (1, 1). Sixth column tells about Geometric mean of predicted time series values, seventh column gives estimated values using GMARMA (1, 1) model, and eighth column says about error squares. By using average of eighth column, we find mean square error value of GMARMA (1, 1).

Comparison

Kolomogrov-Smirnov one sample test is used for testing goodness of fit, Mean square error (MSE) is used for telling which model is better model compared with other models.

Model	MSE	S K value	P - value
ARMA(1, 1)	1.8809	0.442	0.990
LARMA(1, 1)	1.8852	0.421	0.994
GMARMA(1, 1)	0.0756	0.672	0.757

A plot is drawn taking years on x – axis and time series values, LARMA predicted values and GMARMA predicted values on y – axis, and is shown in figure-1.

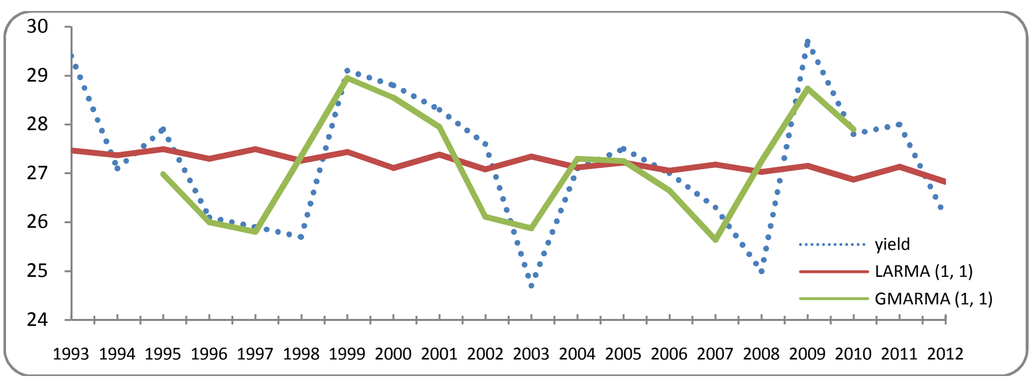

Figure 1

4. Summary and Conclusions

The fitted ARMA models are

The ARMA(1,1)model: $Y_t = 78.890 - 0.999Y_{t-1} + 0.989e_{t-1} + e_t$

Log transformed ARMA (1, 1) model: $Y_t = 2.258 - 0.999Y_{t-1} + 0.99e_{t-1} + e_t$

Geometric moving average of ARMA (1, 1) model: $Y_t = 7.426 + 0.514\ Y_{t-1} + 0.934\ e_{t-1} + e_t$

These three ARMA fitted models are good fit models, and tested using K-S test, and MSE is used for choosing best model among fitted three models.

MSE of ARMA (1, 1) model: 1.880898373

MSE of Log transformed ARMA (1, 1) model: 1.885224

MSE of Geometric mean of ARMA (1, 1) model: 0.075577

MSE of Geometric mean of ARMA (1, 1) model is less than general ARMA (1, 1) and log transformed ARMA (1, 1) models. So, we conclude that the geometric moving average model is the best model than general ARMA (1, 1) and log transformed ARMA (1, 1) models.

Acknowledgement

The first author would like to thank the UGC-BSR for providing the financial assistance of this research work.

References

1. Charles M. Beach and James G. Mackinnon, "A maximum Likelihood Procedure for Regression with Auto correlated errors", Econometrica, Vol. 46, No.1 (1978).
2. Savin. N.E. and White Kenneth J, "Testing for auto correlation with missing observations", Econometrica, Vol. 46, No.1 (1978).
3. Asker H. Choudhury, Simon power & Robert D. St. Louis (1947), "linear estimation of the regression model with ARMA disturbances: a Simulation study", Communications in statistics – simulation and computation, 26:1,315-332.
4. Spyros Makridakis, Steven C. Wheelwright and Rob J. Hyndman, "Forecasting methods and applications", Third Edition, Wiley India Pvt. Ltd., New Delhi.
5. Box, G.E.P. and G.M. Jenkins (1970) time series analysis: Forecasting and control, San Francisco: Holden-Day.
6. Box, G.E.P. and D.A. Pierce (1970) Distribution of the residual autocorrelations in autoregressive-integrated moving-average time series models, Journal of the American Statistical Association, 65, 1509-1526.
7. Mc Kenzie, E. (1984) General exponential smoothing and the equivalent ARIMA process, Journal of Forecasting, 3, 333-334.
8. Mc Kenzie, E. (1986) Error analysis for Winters' additive seasonal forecasting system, International Journal of Forecasting, 2, 373-382.
9. www.nationalagriculturalstatisticsservice.org.in.