Comparison between ARIMA and DES Methods of Forecasting Population for Housing Demand in Johor

Ahmad Rizal Alias, Noor Yasmin Zainun and Ismail Abdul Rahman

Universiti Tun Hussein Onn Malaysia, Faculty of Civil and Environmental Engineering.

Abstract. Forecasting accuracy is a primary criterion in selecting appropriate method of prediction. Even though there are various methods of forecasting however not all of these methods are able to predict with good accuracy. This paper presents an evaluation of two methods of population forecasting for housing demand. These methods are Autoregressive Integrated Moving Average (ARIMA) and Double Exponential Smoothing (DES). Both of the methods are principally adopting univariate time series analysis which uses past and present data for forecasting. Secondary data obtained from Department of Statistics, Malaysia was used to forecast population for housing demand in Johor. Forecasting processes had generated 14 models to each of the methods and these models where evaluated using Mean Absolute Percentage Error (MAPE). It was found that 14 of Double Exponential Smoothing models and also 14 of ARIMA models had resulted to 1.674% and 5.524% of average MAPE values respectively. Hence, the Double Exponential Smoothing method outperformed the ARIMA method by reducing 4.00 % in forecasting model population for Johor state. These findings help researchers and government agency in selecting appropriate forecasting model for housing demand.

1 Introduction

Housing sector is very significant and acts as a catalyst to a country development including Malaysia. This is because housing sector is one part of construction industry which plays vital role in economic and social growth [1]. The balance of housing supply and demand will ensure economic stability which contributes to better Gross Domestic Product (GDP) for a country. According to [2], there is significant relationship between population and housing demand where the number of housings in an area will approximately reflect the number of household's formation. This means that when there is a population growth, it will definitely affect the increasing number of households and leads to a growth in housing demand. [3] pointed out that the important thing need to be considered before forecasting housing demand is the appraisal of population growth. Hence, the population projection is important in producing a good forecasting housing demand model.

Rapid advancement of computer has contributed to the development of predication methods which able to forecast many things including population trend. However, not all these methods can forecast accurately due to their limitations. In this study, two methods which are identified suitable for predicting the population are Autoregressive Integrated Moving Average (ARIMA) and Double Exponential Smoothing. Both methods ignore independent variable in making a prediction and principally adopting univariate time series analysis for

forecasting. ARIMA method has been widely used in various fields such as socio-economic, agricultural and construction. This method is most popular in forecasting univariate time series data [4, 5]. While in contrast, Exponential Smoothing method is more convenient and practical to use because the behaviour of model is easy to understand and has a low level of complexity compare with ARIMA method [6]. Hence, Exponential Smoothing had been used widely with almost 13% by industry [7]. Therefore, this paper presents a comparison between two models which are ARIMA and DES models for forecasting yearly population of Johor state for the period of 10 years.

2 Methodologies

In the first method, ARIMA model is used for the non-stationary time series data. Type of non-stationary model is discussed by Box and Jenkins, [4] and mathematically the ARIMA (p,d,q) model can be expressed as in eq.(1):

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)a_t \qquad (1)$$

where;
ϕ_p = Autoregressive parameters.
θ_q = Moving average parameters.
d = Order of differencing
(1-B) = $Y_t - Y_{t-1}$
Y_t = Actual value at time period t
a_t = Unknown random errors (or residuals)

In Box-Jenkins method, it does not assume any particular pattern in the historical data of the series to be forecasted. Rather, it uses three main steps in modelling process which are; (1) identification of data; (2) parameter estimation and; (3) diagnostic checking to determine the best model to be applied for the final model. The Box-Jenkins forecast method is schematically shown in Fig. 1:
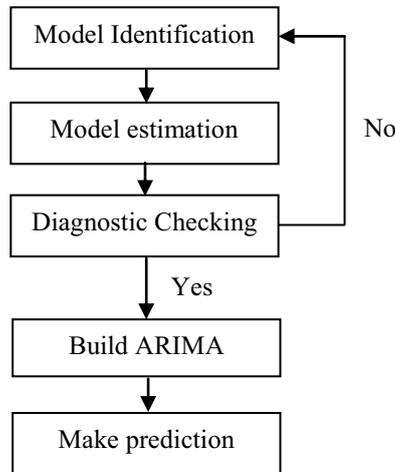

Figure 1. Box-Jenkins method for model selection [4].

Important step to make an appropriate ARIMA model is to determine optimal model parameters (p,d,q) using autocorrelation function (ACF) and partial autocorrelation function (PACF). At the initial stage, the parameters values of ACF and PACF are determined using three types of non-seasonal theoretical Box-Jenkins models as summarized in Table 1.

Table 1. Non-seasonal theoretical Box-Jenkins Models [8].

Model	ACF	PACF
Moving average (MA) of order q	Cuts off after lag q	Dies down
Autoregressive (AR) of order p	Dies down	Cuts off after lag p
Mixed autoregressive-Moving average (p,q)	Dies down	Dies down

In a second method, Double Exponential Smoothing which is a part of Exponential Smoothing method use moving average of time series data. Principally, this method use weight past technique within the data for smoothing purposes to generate smoothed data set. The smoothed data is then use for forecasting application. The following eq. (2) and eq. (3) are used in double exponential smoothing method with separate for level (Lt) and trend (Tt):

$$L_t = \alpha X_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \qquad (2)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \qquad (3)$$

where;

α = Smoothing constants for level

β = Smoothing constants for trend

X_t = Current actual value

L_t = Level

T_t = Trend adjustment

In determine the best double exponential smoothing model, there are four steps to be followed which are; (1) obtain initial estimates for Lt and Tt; (2) update the estimates Lt and Tt by using some predetermined values of smoothing constants; (3) find the best combination of α and β that minimizes error (MAPE) and (4) make prediction using the followed eq. (4) :

$$F_t(k) = L_t + k.T_t \qquad (4)$$

where;

Ft = Forecasting equation

k (1,2,3….n) = Value for next level of forecasting

Following stage is to apply both models with population time series data for Johor state from the Department of Statistic, Malaysia. This population data comprises of age ranged from 15 to 80+ years with an interval of five years from the period of 1970 up to 2015. However for the model 13 & 14 which used age of 75-79 and 80+, there are 10 sets of missing data from period 1970 until 1979. This data is divided into two sets which is for Training (Tn) and Testing (Ts) purposes.

Data from 1970-2010 contributed to 41 sets are used for constructing the models. While, 5 sets of data from 2011 to 2015 are used for testing process. These data sets are also used for validation of the models. In view of these, various performance measures are proposed in literature to estimate forecast accuracy and to compare different models namely Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square (RMS), Mean Percentage Error (MPE) and sum of squared error (SSE) [9]. According to [10], the two most commonly used statistical methods for validations are RMS and MAPE. Therefore, for this study, MAPE is used for validation. MAPE is defined according to the following eq. (5):

$$MAPE = |PE| / n \qquad (5)$$

where;

|PE | = actual value –forecasted value

n = number of forecast

Lewis classified models as "best" if the MAPE less than 10%, "good" with MAPE range between 10% - 20%, "acceptable" for MAPE 20% - 50% and "false" if MAPE more than 50% [11].

3 Materials and study area

Data of population for Johor state obtained from the Department of Statistics, Malaysia was used as a data input for the models. The data was separated according to age category with the range of five years. Based on the selected methods which are Autoregressive Integrated Moving Average (ARIMA) and Double Exponential Smoothing, 14 forecasting models were developed to each of the methods. Using the developed models of forecasting designed, the input data are then analysed using Software Package for Social Science (SPSS) for

Mean Absolute Percentage Error (MAPE) which was used to measure the accuracy of the models

4 Results and discussion

Comparison between two methods, ARIMA and DES aims to identify which methods are suitable to be use as a population model in Johor. Forecasting methods is chosen based on the low value of MAPE in the models where it shows that the DES has a high accuracy rate. Fig. 2-15 shows the comparison between ARIMA and Double Exponential Smoothing (DES) against actual data in term of accuracy of Training (Tn) and Testing (Ts) process.
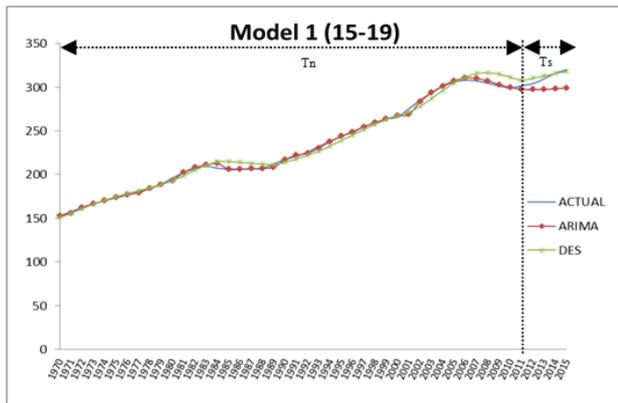
Figure 2. Tn and Ts data set for model 1 (15-19).
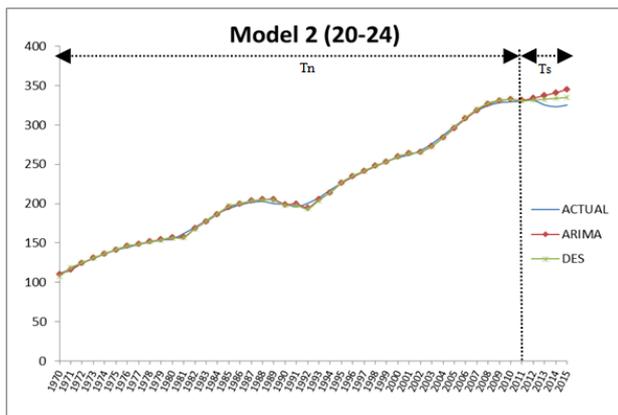
Figure 3. Tn and Ts data set for model 2 (20-24).
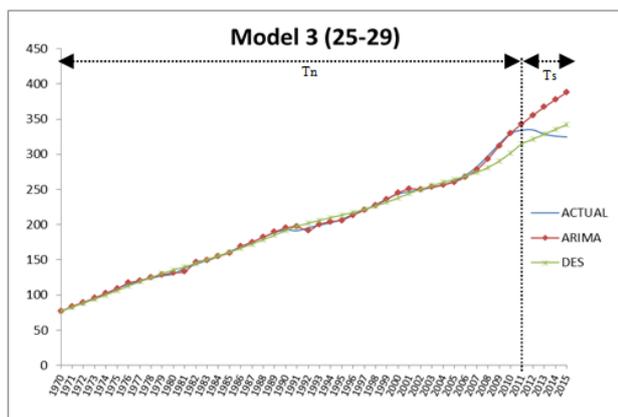
Figure 4. Tn and Ts data set for model 3 (25-29).
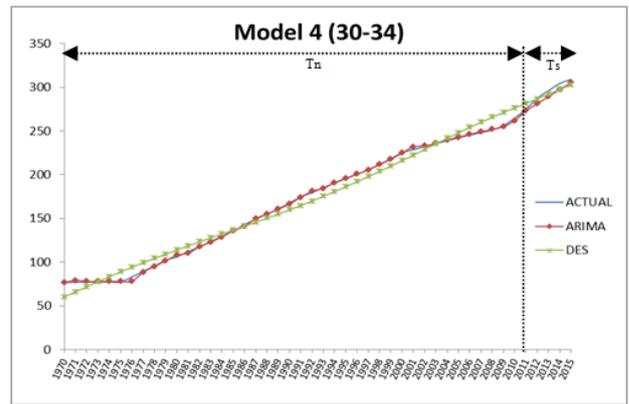
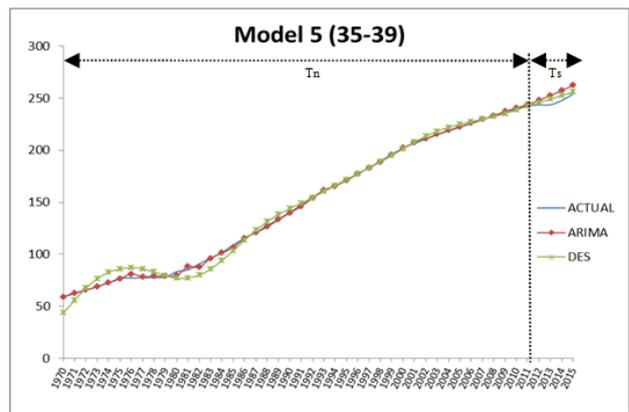**Figure 5.** Tn and Ts data set for model 4 (30-34).

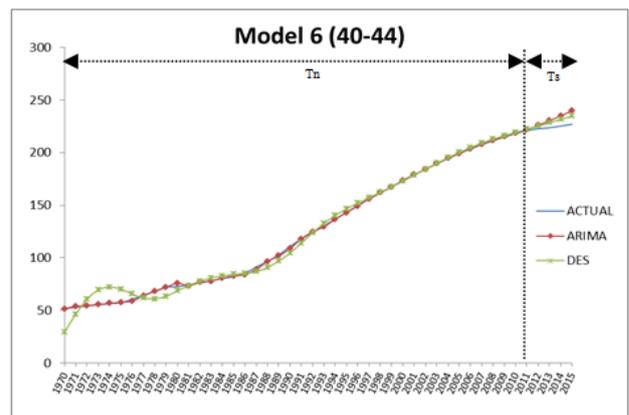**Figure 6.** Tn and Ts data set for model 5 (35-39).
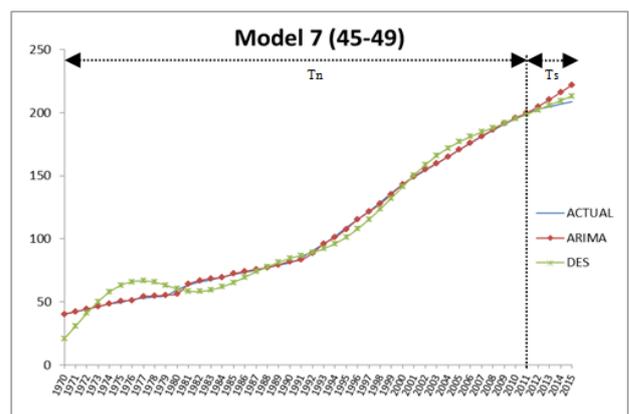

Figure 7. Tn and Ts data set for model 6 (40-44).
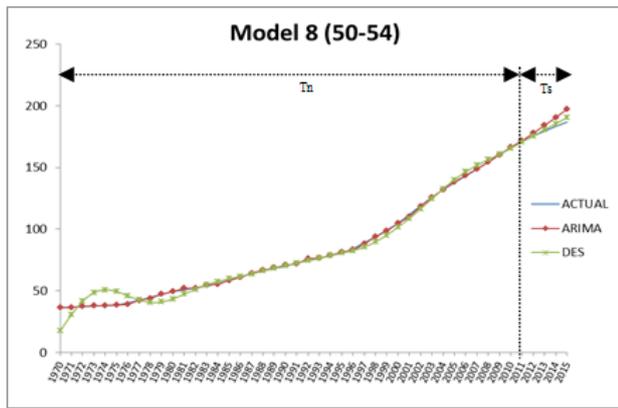

Figure 8. Tn and Ts data set for model 7 (45-49).
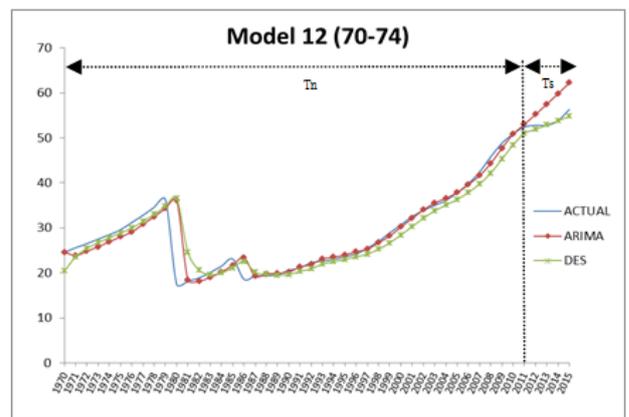
Figure 9. Tn and Ts data set for model 8 (50-54).
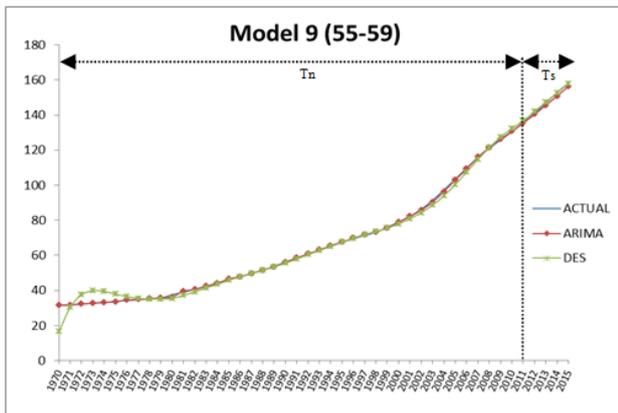

Figure 10. Tn and Ts data set for model 9 (55-59).

Figure 11. Tn and Ts data set for model 10 (60-64).

Figure 12. Tn and Ts data set for model 11 (65-69).

Figure 13. Tn and Ts data set for model 12 (70-74).
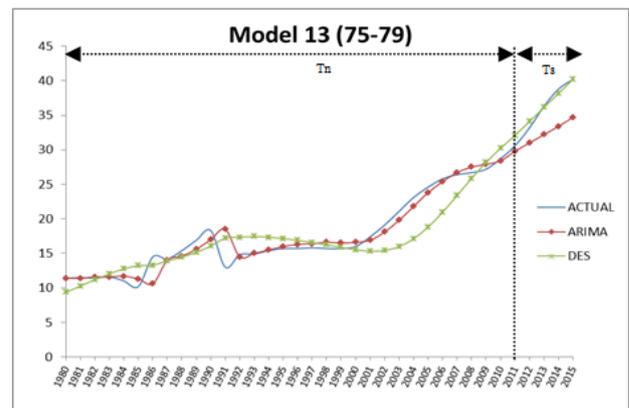

Figure 14. Tn and Ts data set for model 13 (75-79).

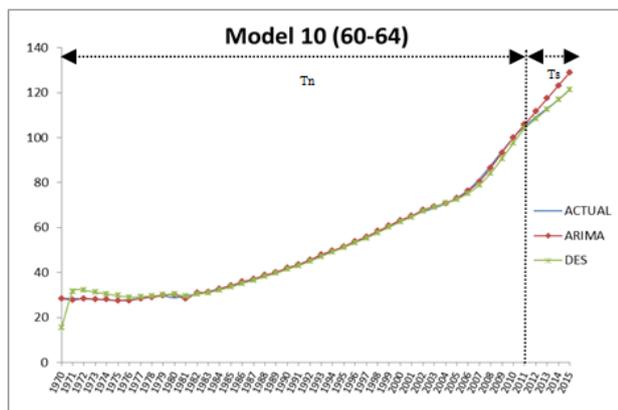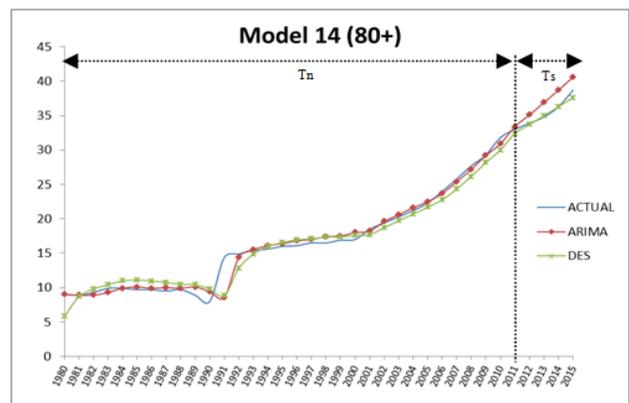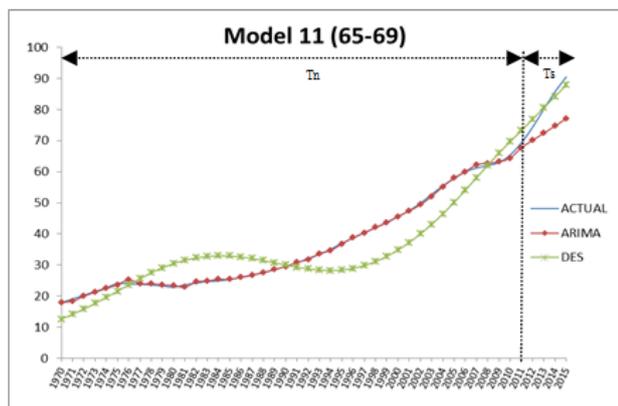**Figure 15.** Tn and Ts data set for model 14 (80+).

Based on the Fig. 2-Fig. 15, obviously it shows that the DES model is more accurate for validation process rather than ARIMA model using the data set 2011-2015. In term of testing, ARIMA model is seems good for that particular part but it is not enough for the purpose of model development without accuracy in future prediction (testing). According to Lewis [11], best model development if the MAPE less than 10%, and MAPE range between 10% - 20%, is consider acceptable. In view of this, both of model is less than 20% error, but DES is superior with every single model is less than 5% MAPE comparing with ARIMA models. Table 2 shows the comparison between ARIMA and DES model in term of model parameters and error (MAPE).

Table 2. Comparison between ARIMA and DES Models.

Model No	Age group	MODEL			MAPE	
		ARIMA (p,d,q)	DOUBLE EXPONENTIAL SMOOTHING (DES)			
			Alpha (α)	Beta (β)	ARIMA	DES
1	15-19	(1,1,0)	0.5123	0.1416	3.930	1.17
2	20-24	(1,1,0)	0.9612	0.9569	3.240	1.72
3	25-29	(1,1,0)	0.2123	0.1121	11.16	3.64
4	30-34	(1,1,0)	0.0157	0.9571	1.720	1.64
5	35-39	(1,1,0)	0.2351	0.9836	2.162	1.20
6	40-44	(1,1,0)	0.2863	0.9462	2.951	1.97
7	45-49	(1,1,0)	0.2565	0.5121	2.979	0.90
8	50-54	(1,1,0)	0.2863	0.8562	2.738	0.79
9	55-59	(1,1,0)	0.5234	0.4651	0.800	0.42
10	60-64	(1,1,0)	0.9325	0.1813	3.706	0.45
11	65-69	(1,1,0)	0.0611	0.5086	9.034	2.86
12	70-74	(1,2,0)	0.6420	0.0391	11.294	1.46
13	75-79	(1,2,0)	0.1051	0.7596	9.410	1.93
14	80+	(1,2,0)	0.6092	0.1331	12.210	1.16
AVERAGE					5.524	1.522

Referring to table 2 in ARIMA method, the 14 models generated an average MAPE value of 5.524 with the lowest MAPE of 0.8 for model no. 9 (Age group 55-59) and the highest MAPE of 12.210 for the model no. 14 (age group 80+). While, in DES method, values of MAPE recorded for all models are below to 3 except for model no 3 (age group 25-29) having MAPE 3.640. In addition, every single model in DES generated lowest MAPE compare with ARIMA model. Therefore, the DES is considered a better method for modelling population for Johor state for the next 10 years. Hence, DES method performs better than ARIMA with these data.

5 Conclusion

This study has shown comparison between ARIMA and DES methods which are popularly used for linear data trend analysis. The comparison applied Johor state yearly population data sets. It was found that the most appropriate forecasting method for population of Johor state is DES method with lower MAPE for all 14 models. DES method for all models based on age group have the lower MAPE can be assumed better accurate models for forecasting population in Johor for the next 10 years.

Acknowledgement

References

1. I.A. Rahman, A.H. Memon, A. Azis and N. Abdullah. Modeling causes of cost overrun in large construction projects with partial least square-SEM approach: Contractor's perspective. Engineering and Technology **5**, 1963–1972, (2013).
2. C.H. Mulder. Population and housing: A two sided relationship. Demographic Research **15**, 401-412. (2006).
3. W.S Thompson. Population growth and housing demand. The ANNALS of the American Academy of Political and Social Science **190**, 131-137, (1937).
4. G.E.P Box & G. Jenkins. *Time series analysis, forecasting and control,* Holden-Day, San Francisco, California. (1970).
5. G.P. Zhang. Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing **50**, 159–175, (2003).
6. K.K. Lai, L. Yu, S.Y. Wang and W. Huang. Hybridizing exponential smoothing and neural network for financial time series predication. Computational Science–ICCS. Springer Berlin Heidelberg **3994**, 493-500, (2006).
7. G.L. Lilien, P. Kotler. Marketing decision making: A model-building approach. Marketing research **21**, 339-341, (1983).
8. B.L Bowerman, R.T. Connell, and A.B. Koehler. *Forecasting, Time Series, and Regression*, South-Western College Pub; 4th edition, (2005).
9. R.Adhikari, & R.K. Agrawal. *An Introductory Study on Time Series Modeling and Forecasting*. 21-45, (2013).
10. N.Y. Zainun. Computerized Model to Forecast Low-Cost Housing Demand in Urban Area in Malaysia Using Artificial Neural Networks (ANN). Loughborough University, (2011).
11. C.D. Lewis. *Industrial and Business Forecasting Methods*. Butterworths. **2**,194-196, (1982).