FORECASTING OF PERCENTAGE TIME THAT PARTS FOR INDUSTRIAL PROJECTS IN AUSTRALIA

Mahwish Rabia¹, Ramisha Irshad¹, Humma Nawaz¹, Maryam Khalid¹ and Ayesha Raana²

1. Department of Statistics, GC Women University Sialkot, Pakistan.
2. Department of Computer Science, GC Women University Sialkot, Pakistan.

Abstract

This paper endeavors to study the Australia percentage time that parts for industrial projects available when needed. Box-Jenkins methodology is used to forecast next 20 observations. To apply Box-Jenkins methodology data should be stationary and unit root test is used to check stationarity of data. It is found that a minor trend is found in data set and to remove trend 1st difference is applied which makes the data stationary. For model identification ACF and PACF are plotted. Furthermore, using Box-Jenkins methodology best model is selected on the basis of smaller AIC, SBC and MSE. Since, ARIMA (0,1,1) has the lowest value of AIC, SBC and MSE. So, this model is recommended as best for forecasting. In addition, to check the accuracy of forecasted values MAPE, MAE, RMSE are also computed. It can be concluded that the percentage time that parts for industrial projects in Australia will increase gradually.

Journal Homepage: www.journalijar.com

Article DOI:10.21474/IJAR01/10365
DOI URL: http://dx.doi.org/10.21474/IJAR01/10365

RESEARCH ARTICLE

MANUSCRIPT HISTORY

Received: 30 November 2019
Final Accepted: 31 December 2019
Published: January 2020

Key words:- Arima, Stationary, Aic, Sbc, Mse, Rmse, Mae, Mape, Box-Jenkins-Methodology

INTRODUCTION

Time series is an important area of forecasting based on past observation. Time series is a collection of data which are generally used for prediction and forecasting. Time series prediction refers to the process by which the future values of a system is forecasted based on the information obtained from the past and current data. Generally, pre-defined mathematical model is used to make accurate predictions. Time series prediction model are mostly used in financial area. In time series two main techniques are used for prediction. These are Auto Regressive (AR) and Moving Average (MA). With the help of these technique we can establish Auto Regressive Moving Average (ARMA), Auto Regressive Integrated Moving Average (ARIMA), Seasonal Regressiv Integrated Moving Average (SARIMA) and Box Jenkins models. Many business and economic time series exhibit seasonal and trend variation. Seasonality is a periodic and recurrent pattern caused by factors such as weather, holidays, repeating promotions as well as the behavior of economic agent (Hyllbberg,1992). A time series with trend is considered to be non-stationary and often needs to be made stationary before most modeling and forecasting processes take place. There are several different approaches to time series modeling. Traditional statistical models including Moving Average (MA), exponential smoothing and Auto-Regressive Integrated Moving Average (ARIMA) are linear in that predictions of the future values are constrained to be linear function of past observations. But one of the most important and widely used time series model is the Auto-Regressive Integrated Moving Average (ARIMA) model. The popularity of the Auto-Regressive Integrated Moving Average (ARIMA) model is due to its statistical properties as well as the well-known Box-Jenkins methodology in the model building process. ARIMA model are quite flexible in that they can represent several different types of time series, i.e. pure auto-regressive, pure moving average and combined AR and

Corresponding Author:- Mahwish Rabia
Address:- Department of Statistics, GC Women University Sialkot, Pakistan.
MA (ARMA) series, their major limitation is the pre-assumed linear form of the model and therefore, no nonlinear patterns can be captured by the ARIMA.

Literature Review:
Hudson and Ethridge (1999) used an export tax on raw cotton from 1988–1995 in order to suppress the internal price of cotton to benefit the domestic yarn industry. An analysis was conducted to estimate the impact of this policy on both the cotton and yarn sectors. These effects were simulated using the results of a structural econometric model of these sectors of Pakistan's economy. Results indicated that the export tax had a negative impact on the growth rate in the cotton sector, while having little or no impact on the yarn sector. Thus, the export tax did not achieve its objective of increasing the growth rate of value-added (yarn) production above what would have occurred naturally. Dutta and Ahmed (2004) used the framework of an endogenous growth model, this study empirically analyses the relationship between trade policies and industrial growth in Pakistan during the period 1973–1995. The cointegration and error correction modeling approaches have been applied. The empirical results suggest that there exists a unique long-run relationship among the aggregate growth function of industrial value added and its major determinants of the real capital stock, the labor force, real exports, the import tariff collection rate and the secondary school enrolment ratio. The short-term dynamic behavior of Pakistan's growth function of industrial value added has been investigated by estimating an error correction model in which the error correction term has been found to be correctly signed and statistically significant. Zhang and Qi (2005) proposed a study on Neural Network forecast for seasonal and trend time series. It was observed that time series model is affected with both seasonal and trend patterns. It was studied effectiveness of data preprocessing including de-seasonalization and de-trending, on neural network modeling and forecasting performance. It was examined the capability of artificial neural network in modeling and forecasting seasonal and trend time series. It was founded that neural network with both de-trending and de-seasonalization are able to significantly outperform seasonal ARIMA model in out-of-sample of forecasting. Finally, it was founded that neural network may yield much worse forecasting performance than ARIMA models. Zhang (2003) examined a study on time series forecasting using a hybrid ARIMA and artificial neural network (ANN) models. It was observed that linear ARIMA model and nonlinear ANN model to capture different forms of relationship in the time series data. It was explained the problem of overfitting by fitting the ARIMA model that is more strongly related to neural network can be eased. It was founded that experimental results of the combined model can be an effective way to improve forecasting accuracy achieved rather than models used separately. Kayacan et,al., (2010) observed a study on Grey system theory-based models in time series prediction. It was defined different Grey models such as GM(1,1), Grey Verhulst model, modified grey models using Fourier series is investigated. It was showed the simulation results that modified Grey models had higher performances not only on model fitting but also on forecasting. It was also observed that GM (1,1) using Fourier series in time is the best in model fitting and forecasting.Thomassey (2010) observed that sales forecasts in clothing industry: the key success factor for the supply chain management. It was observed that they forecast different models which perform more accurate and reliable sales forecasts. It was used fuzzy logic, neural network and data mining. All the techniques which applied in the paper had effective results. If a company wants to future implement a suitable forecasting system and also rethinking about their supply chain to reduce lead times and minimum order quantities.Stambuli (2013) examined how we can control oil importation in the long-run without disturbing the normal functioning of the economy as the demand for oil in Tanzania changes due to change in income of the country and international oil price. He found that the demand for oil in the short run is both income and price inelastic, while in the long run it is income elastic and price inelastic, showing that income has more effect on oil demand than price.Laura Boemeke et al., (2015) examined that coconut oil (CO) has positive effect on health and also cure from disease. It protect us from breast cancer. The effect of long term consumption of coconut oil (CO) are unknown.

Methodology:--
In this study, the Australia percentage time that parts for industrial project available when needed (Appendix 1) has been forecasted through using ARIMA methodology. Data of these years has been extracted and then we utilized the Box-Jenkins methodology to forecast the future observation.

Box-Jenkins methodology:
A mystery to know, what happened in the future. So, in that case it is most suitable model selection method for forecasting of time series variable. Box-Jenkins methodology of forecasting is different from most methods because it does not assume any particular pattern in the historical data of the series to be forecasted. Box-Jenkins Analysis refers to a systematic method of identifying, fitting, checking and using autoregressive moving average (ARIMA) time series models. The model fits well if the residuals are generally small or randomly distributed. The Box-Jenkins
method is appropriate for time series of medium to long length (at least 50 observations). The Box-Jenkins methodology is valid only when the variable fulfills some assumptions which are:
1. Postulate the general model
2. Model identification
3. Parameter estimation
4. Forecasting

**Postulate the general model:**
In postulating the model, the time series is checked for stationarity. In our study to make our time series stationary we took the differencing of lag 1 as the time series was non-stationary.

**Model identification:**
We have to determine the required model. In this step we use graphical method and plotting the series of ACF and PACF. If the graphical plot indicate that series is not stationary then we move towards difference. Difference is the best way to transform the non-stationary series.

**Parameter estimation:**
After the identification of an appropriate model the next step is to estimate the parameters. The parameters of the selected model are estimated using maximum likelihood techniques etc. as outlined in Box-Jenkins (1976) through ARIMA (p,d,q).

**Diagnostic checking:**
In the last if the model is adequate then use it for forecasting otherwise repeat the process again until the required model is obtained. Predictions resulted from this model is especially for short-term predictions and in most cases. It is more reliable than traditional modeling method of econometrics. Of course, it is necessary to separately judge about each special case.

**Flow Chart of Box-Jenkins Methodology:**
**Results and Interpretation:**

**Stationary Test:**

**Graphical Method:**
In graphical method we use correlogram and line graph.

**Empirical Method:**
Generally, in this method two tests are used which are Augmented Dickey Fuller Test (ADF Test) and Phillips Perron Test (PP Test).

**General Procedure to Check Stationarity of Data:**
To analyse the data first of all we check data is stationary or not. If not then make it stationary and make its possible model. The procedure is given below:

**Step 1:**
Hypothesis
- $H_0$: Data is not stationary.
- $H_1$: Data is stationary.

**Step 2:**
Level of significance $\alpha=0.05$

**Step 3:**
Test statistic
1. Unit root test or Augmented Dickey Fuller test (ADF Test)
2. Phillip Perron test (PP Test)

**Step 4:** Calculation
On E-views

**Step 5:**
Critical region On the basis of p-value. If the p-value less than level of significance reject $H_0$ otherwise don’t reject $H_0$.

**Step 6:**
Decision If reject we conclude series is stationary otherwise we say series is non-stationary.

**To Diagnose Stationary:**
When the variable has no change in mean and variance for a long time, it said to be stationary. For applying Box Jenkins methodology, variable must be stationary.

**Australia weekly data**
Figure 1.1 shows the time series plot for data (Australia percentage time). The graph depicts that there is fluctuations or an increasing trend in the data.

Unit Root Test:
Ho: The data is non-stationary.
H1: The data is stationary

Table 1.2:- Unit Root Test to check Stationarity at first difference.

| Augmented Dickey-Fuller test statistic | t-Statistic | Prob.* |
|----------------------------------------|------------|--------|
| Test critical values:                  |            |        |
| 1% level                               | -11.23856  | 0.0001 |
| 5% level                               | -3.507394  |        |
| 10% level                              | -2.584738  |        |

From the provided evidence as our p-value (0.0001) so we have weak evidence against Ho and conclude that the results are significant. So the data is stationary at 1st difference (Table 1.2).

Correlogram:

Figure 1.2 shows the autocorrelation and partial autocorelations at first difference. It seems from graph of autocorrelation that only one spikes at lag one is statistically significant but the rest are not. So, it identifies MA(1) model. Similarly, partial autocorrelation shows that only two spikes are statistically significant but the rest are not. So, it identifies AR(2) model.

Model Identification:

Table 1.3:- AIC and SBC of Different ARIMA Models

| Models     | AIC       | SBC       | MSE    |
|------------|-----------|-----------|--------|
| ARIMA(0,1,1)| 4.428636*| 4.484560*| 4.800* |
| ARIMA(2,1,0)| 4.911301 | 4.967997 | 4.958  |
| ARIMA(2,1,1)| 4.435418 | 4.489873 | 4.816  |

Table 1.3 demonstrates some possible ARIMA models. The results revealed that ARIMA (0,1,1) has the lowest value of AIC, SBC and MSE. That’s why suggested model is ARIMA (0,1,1) because it has least value of AIC and SBC. So, this model is recommended for forecasting.
The general model is:

\[ D(PT) = C + \alpha AR(p) + \beta MA(q) + \mu_i \]

Where \( D(PT) \) is 1st differenced series of data, \( C \) is an intercept, \( \alpha \) is coefficient of autoregressive lag values i.e. AR\( (p) \), \( \beta \) is a coefficient of moving average lag values i.e. MA\( (q) \) and \( \mu \) shows the residuals of model where residuals should be independently identically and normally distributed.

ARIMA\((0,1,1)\) model is used for prediction, so its estimated equation is:

\[ PT = 0.041722 - 0.725018 \]

Where \( PT \) denotes the percentage time that parts for industrial projects in Australia.

For checking the accuracy of forecasting we apply forecasting checks.

I. RMSE (Root Mean Square Error)
II. MAE (Mean Absolute Error)
III. MAPE (Mean Absolute Percentage Error)

We select the model which has minimum RMSE, MAE, MAPE.

| Models      | MAPE   | MAE     | RMSE   |
|-------------|--------|---------|--------|
| ARIMA\((0,1,1)\) | 2.5224* | 2.0459* | 2.5630* |
| ARIMA \((2,1,0)\) | 4.7497  | 3.8350  | 4.5923  |
| ARIMA \((2,1,1)\) | 2.8097  | 2.2696  | 2.8888  |

ARIMA\((0,1,1)\) is selected because it has less MAPE, MAE and RMSE.

Diagnostic Checking:

In Box-Jenkins Methodology, the model is adequate if there is no existence of correlation between residuals and the residuals are independently identically normally distributed (IID) i.e. random. To test whether estimated results of residuals are white noise or not (when residuals shows white noise it means that the model is just right) ACF and PACF of results are plotted.

Correlogram for Residuals:

![Figure 1.3: ACF and PACF of Residuals](image-url)
Based on the Figure 1.3 it appears that none of the correlations for the autocorrelation and partial autocorrelation function of the residuals are significant. So, the model meet the assumption that the residuals are independent, random and white noise.

**Forecasting:**

**Table 1.4:-** Forecasting for next 20 weeks.

| Weeks     | Forecast |
|-----------|----------|
| 09/02/19  | 85.0956  |
| 16/02/19  | 85.1392  |
| 23/02/19  | 85.1827  |
| 02/03/19  | 85.2263  |
| 09/03/19  | 85.2699  |
| 16/03/19  | 85.3135  |
| 23/03/19  | 85.3571  |
| 30/03/19  | 85.4007  |
| 06/04/19  | 85.4442  |
| 13/04/19  | 85.4878  |
| 20/04/19  | 85.5314  |
| 27/04/19  | 85.5750  |
| 04/05/19  | 85.6186  |
| 11/05/19  | 85.6622  |
| 18/05/19  | 85.7057  |
| 25/05/19  | 85.7493  |
| 01/06/19  | 85.7929  |
| 08/06/19  | 85.8365  |
| 15/06/19  | 85.8801  |
| 22/06/19  | 85.9237  |

*Table 1.4:-* shows the forecasting for next 20 weeks. It shows that the percentage time that parts for industrial projects in Australia will increase gradually.

**Conclusion:**

This paper attempts to forecast percentage time that parts for industrial projects in Australia. Firstly data should be stationary and by observing plot of data, a minor trend is found in data set and to remove trend 1st difference is applied which makes the data stationary. Secondly, for model identification ACF and PACF are plotted. Since, ARIMA (0,1,1) has the lowest value of AIC, SBC and MSE. So, this model is recommended as best for forecasting. Moreover, on the basis of MAPE, MAE, RMSE the precision of the forecasted values are also examined and ARIMA(0,1,1) gives the more precise results as compare to other proposed models. It can be concluded that the percentage time that parts for industrial projects in Australia will increase gradually in the upcoming years.

**References:**

1. Hudson, D., and Ethridge, D. (1999). Export taxes and sectoral economic growth: evidence from cotton and yarn markets in Pakistan. Agricultural Economics, 20(3), 263
2. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.
3. Dutta, D., and Ahmed, N. (2004). Trade liberalization and industrial growth in Pakistan: a cointegration analysis. Applied Economics, 36(13), 1421-1429.
4. Zhang, G. P., and Qi, M. (2005). Neural network forecasting for seasonal and trend time series. European journal of operational research, 160(2), 501-514.
5. Kayacan, E., Ulutas, B., and Kaynak, O. (2010). Grey system theory-based models in time series prediction. Expert systems with applications, 37(2), 1784-1789.
6. Thomassey, S. (2010). Sales forecasts in clothing industry: The key success factor of the supply chain management. International Journal of Production Economics, 128(2), 470-483.
7. Stambuli, B. B. (2013). Price and Income Elasticities of Oil Demand in Tanzania: An Autoregressive Approach. Business Management Dynamics, 3 (1), 75-83.
8. Laura, B. A. M. (2015). Effects of coconut oil on human health. USA: Script. 7(1), 20-30.

Appendix 1:-
https://datamarket.com/data/set/22qk/time-that-parts-for-industrial-project-available-when-needed-weekly#!ds=22qk&display=line

| Week  | % time that parts for industrial project available when needed (weekly) | Week  | % time that parts for industrial project available when needed (weekly) | Week  | % time that parts for industrial project available when needed (weekly) |
|-------|------------------------------------------------|-------|------------------------------------------------|-------|------------------------------------------------|
| 0001 W01 | 80.4 | 0001 W31 | 84.4 | 0002 W09 | 79.4 |
| 0001 W02 | 83.9 | 0001 W32 | 83.7 | 0002 W10 | 77.9 |
| 0001 W03 | 82.6 | 0001 W33 | 84.5 | 0002 W11 | 80.4 |
| 0001 W04 | 77.9 | 0001 W34 | 84.6 | 0002 W12 | 79.4 |
| 0001 W05 | 83.5 | 0001 W35 | 85.2 | 0002 W13 | 83.2 |
| 0001 W06 | 80.8 | 0001 W36 | 85.2 | 0002 W14 | 81 |
| 0001 W07 | 78.5 | 0001 W37 | 80.1 | 0002 W15 | 81.7 |
| 0001 W08 | 79.3 | 0001 W38 | 86.5 | 0002 W16 | 81.2 |
| 0001 W09 | 81.2 | 0001 W39 | 81.8 | 0002 W17 | 79.1 |
| 0001 W10 | 81.1 | 0001 W40 | 84.4 | 0002 W18 | 80 |
| 0001 W11 | 78.2 | 0001 W41 | 84.2 | 0002 W19 | 81.5 |
| 0001 W12 | 80.9 | 0001 W42 | 84.1 | 0002 W20 | 83.8 |
| 0001 W13 | 77.9 | 0001 W43 | 83.2 | 0002 W21 | 82.2 |
| 0001 W14 | 81.5 | 0001 W44 | 83.9 | 0002 W22 | 82.4 |
| 0001 W15 | 81.4 | 0001 W45 | 86 | 0002 W23 | 79.9 |
| 0001 W16 | 78.9 | 0001 W46 | 82.2 | 0002 W24 | 82.3 |
| 0001 W17 | 76.2 | 0001 W47 | 81.2 | 0002 W25 | 83.2 |
| 0001 W18 | 79.4 | 0001 W48 | 83.7 | 0002 W26 | 81.3 |
| 0001 W19 | 81.4 | 0001 W49 | 82.7 | 0002 W27 | 82.4 |
| 0001 W20 | 80 | 0001 W50 | 84.8 | 0002 W28 | 82.2 |
| 0001 W21 | 79.9 | 0001 W51 | 81.2 | 0002 W29 | 82 |
| 0001 W22 | 80.5 | 0001 W52 | 83.8 | 0002 W30 | 83.7 |
| 0001 W23 | 79.7 | 0002 W01 | 86.4 | 0002 W31 | 84.6 |
| 0001 W24 | 81.4 | 0002 W02 | 81.6 | 0002 W32 | 85.7 |
| 0001 W25 | 82.4 | 0002 W03 | 83.6 | 0002 W33 | 85.1 |
| 0001 W26 | 83.1 | 0002 W04 | 85.9 | 0002 W34 | 84.5 |
| 0001 W27 | 80.4 | 0002 W05 | 79.8 | 0002 W35 | 85.6 |
| 0001 W28 | 82.9 | 0002 W06 | 80.8 | 0002 W36 | 84.7 |
| 0001 W29 | 82.7 | 0002 W07 | 78.7 | 0002 W37 | 79.9 |
| 0001 W30 | 84.9 | 0002 W08 | 80.6 | 0002 W38 | 88.9 |