International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Modelling and Forecasting Malaysian Consumer Price Index

Azme Bin Khamis[1], Asran Nazmi Bin Atan[2], Siti Noor Asyikin Mohd Razali[3], Norhaidah Mohd Asrah[4], Sabariah Saharan[5]

Faculty of Applied Science and Technology, Universiti Tun Hussien Onn Malaysia

Abstract: ***Consumer price index is measure the average of people spent for consumer goods and services, such as transportation, food and medical care.CPI also can affect an inflation that will reduce the rate of economic growth in the country. The main objective of this research was to determine the appropriate model, to compare the appropriate model for Malaysian consumer price index using RMSE, MAD and MAPE and to forecast Malaysian consumer price index for short period. Therefore, to come up with a model and forecasts of CPI, Box and Jenkins methodology were used which consists of three main steps; Model Identification, Parameter Estimation and Diagnostic Checking. Assumptions about linear regression model are relationship between response and predicted is linear, mean of is 0. Third was variance are independent of each other, variables are identically and independently distributed following N (0, σ^2) and predicted variable is a nonstochastic variable. Therefore, ARIMA (1, 1, 0) was selected as an appropriate model which can fits well data, as well as to make also accurate forecast. Hence, the forecast was made for 12 months ahead of the year 2017, and the findings have shown that the CPI was likely to continue rising up with time.***

Keywords: Consumer Price Index, Box-Jenkins Model, Linear Regression

1. Introduction

Consumer price index (CPI) defines as the average cost of user to spend on goods and services that depend on many aspect such as the broad money, exports of goods and services, gross domestic product and household final consumption expenditure that have relationship with CPI. CPI also can affect an inflation that will reduce the rate of economic growth in the country. CPI will always increase over time, resulting in increased inflation will unfavourably affect economic growth and also produce poor living condition to the people [1, 2, 3].

CPI in Malaysia have firm rise pattern from 95.4% in 2000 to 122.8% in 2013. It mean consumer need spent a huge money in order to buy their daily need. It will cause an inflation in Malaysia because they need pay a high price to get the product that they need. But the introducing of Goods and Services Tax (GST) in 2015 is another way not to reduce the CPI but it will give more burden to the people in Malaysia. The most causes that lead to high CPI is food and non-alcoholic beverages that is common use in human daily life [4, 5, 6].

CPI also used to measure the average of people spent for consumer goods and services, such as transportation, food and medical care. The CPI is calculated by taking price changes for each item in the prearranged basket of goods and averaging of the goods are weighted according to their importance. Every changes in CPI are used to assess price changes associated with the cost of living [6, 7].

2. Methodology

2.1 Box-Jenkins model

First step for produce Box-Jenkins model was model identification. If the time series were not stationary, transformations or differencing the data before procced to another step. Chatfield and Prothero [8, 9] have studied that stationary of the variance can be determined by using transformations that would improve the forecast accuracy. For box-cox plot, the rounded value of λ must one, if less than one transformations must be take place. Differencing the data was the other way to make sure data will non-stationary. It was after the time series was become stationary and then estimating the sample autocorrelations (ACF) and partial autocorrelation (PACF) and compare the result. Hipel, Mcleod and Lennox [9] studied that second step was estimation the parameters from the model identification. The model could be estimated using least-squares or maximum likelihood. Maximum likelihood estimation will maximize the probability of observed value to find the value of parameters. For least-squares methods, it will produce a smaller sum of squares.

[10, 11] conducted a studied that the last step in modelling the Box-Jenkins model was model checking regarding to statistical suitability to detect whether the model was good or bad. First was check the autocorrelation in the residuals to detect if there is information that not been counted for in the model. It could help to improve the model that could affect the forecast accuracy. Then, Box-Ljung test also could be used to test the lack of fit of the model. If there were correlations between residuals, then there was information left in the residuals which should be used in computing forecasts. If the residuals have a mean other than zero, then the forecasts are biased.

2.2 Simple Linear Regression

The estimated regression equation describing a straight-line relationship between an independent variable x and a dependent variable y is written as [12],

$$\hat{y} = b_0 + b_1 X$$

Where b_1 and b_0 are unknown parameter and can be estimated using maximum likelihood estimator, which is by minimizing the sum of squares of errors. After that, compute testing of hypothesis. The analyst wishes to control the risk of a Type I error at a = 0.05. The conclusion H_a could be reached at once by referring to the 95 percent confidence interval for b_1, since this interval does not include 0.An explicit test of the alternatives is based on the test statistic:

$$t^* = \frac{b_1}{s\{b_1\}}$$

Where b_1 is the slope and s(b_1) is the standard error of the slope. The decision rule with this test statistic for controlling the level of significance at a is:If $|t^*| \leq t\ (1-\alpha/2;n-2)$, conclude H_0 true, otherwise If $|t^*| > t\ (1-\alpha/2;n-2)$, *conclude* H_a. If the value of t^* is bigger than *t-value* from analysis of variance, we conclude H_a, that $b_1 \neq 0$ that there is linear association between response variable and predictor variable.

2.3 Forecast Accuracy

According to [13], root mean square error (RMSE), mean absolute deviation (MAD) and mean absolute percentage error (MAPE) could be used to measure forecast accuracy that involve time period, *t*. The measurements were number of period forecast, *n*, actual value in time period at time, *t*, *Yt*and forecast value at time period, *t*, *Ft*.

$$RMSE = \sqrt{\frac{\sum(Yt-Ft)^2}{n}}$$

$$MAD = \frac{\sum|Yt-Ft|}{n}$$

$$MAPE = \frac{\sum\left|\frac{Yt-Ft}{Yt}\right|}{n} x100$$

3. Results and Discussions

Malaysian Consumer Price Index (CPI) was collected around 536 data of the and represents 536 months data from January 1972 to August 2016 are used to model fitting and those from September 2016 until August 2017 were used to forecast the future CPI.

3.1 Box-Jenkins Method

Figure 3.1 shows the time series plot for CPI for the period January 1972 to June 2012. It is show that the pattern of CPI was increased trend over the time, which explains also that the mean is not constant and depends on time. It mean that the series is not stationary that the series must be transformed to obtain stationarity by using box-cox plot that the value of λ must one. But if the λ value is less that one, transformation must be take place.

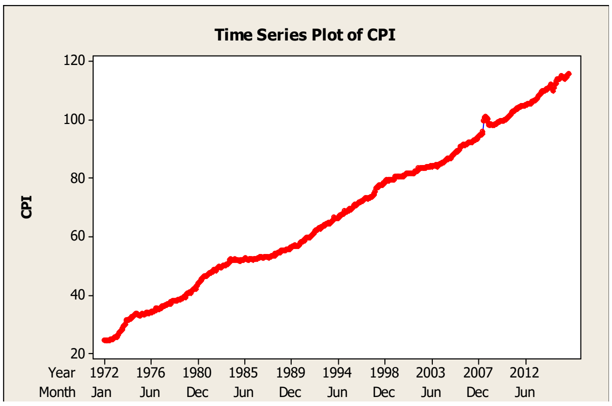


Figure 3.1: Time series plot of CPI

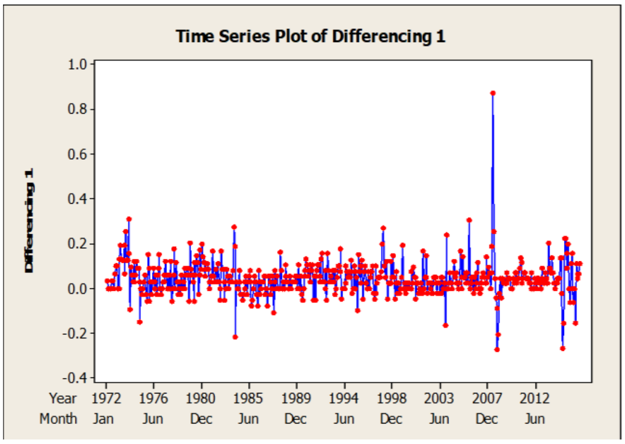
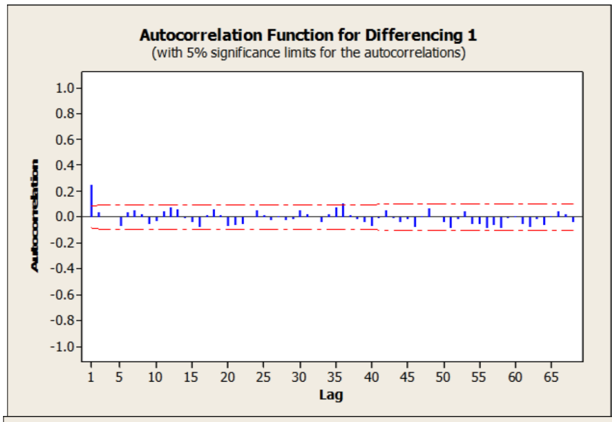


Figure 3.2: Time series plot of Differencing 1

Figure 3.2 above show the time series plot after first differencing that it shows the stationarity on the data that can be used to plot another Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).



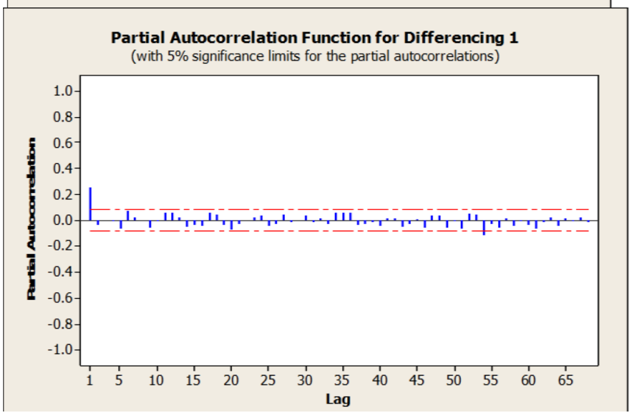


Figure 3.3: ACF and PACF plot for Differencing 1

Figure 3.3 show the ACF and PACF plots were used to come up with the order of the model to identify an appropriate model. Result from ACF and PACF are both have cuts off after lag 1. The model are AR (1) or MA (1) and the appropriate model ARIMA (1, 1, 0) or ARIMA (0, 1, 1).

Second step was estimation the parameters from the model identification. The selecting a potential model are ARIMA (1, 1, 0) or ARIMA (0, 1, 1) and estimating its parameters. The estimation for the coefficient of AR (1) is 0.2614, *t*-test =6.25 with *p*-value=0.000 that is less than 0.05 which is significant. Therefore the AR(1) in the model is significant.

Paper ID: SR20719075537 DOI: 10.21275/SR20719075537

The estimation for the coefficient of MA (1) is -0.2608, *t*-test =-6.23 with *p*-value=0.000 that is less than 0.05 which is significant. Therefore the MA (1) in the model is significant. The result from Box-Ljung test show no sufficient evidence to reject the null hypothesis of no serial correlation in the first 48 lags for both model. Hence we conclude that the fitted model is adequate.

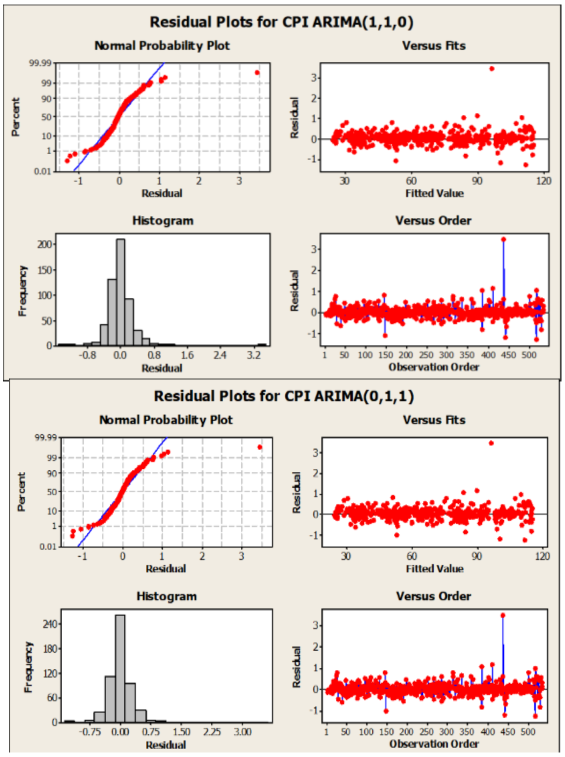


Figure 3.4: Residual Plot of CPI ARIMA (1, 1, 0) and ARIMA (0, 1, 1)

Figure 3.4 show that both model have the residuals are almost exactly normally distributed, and it can be also confirmed that there is no correlation in the residuals which means there is no left information in the residuals to be used in fitting a model. Therefore, the ARIMA (1, 1, 0) and ARIMA (0, 1, 1) was successfully selected as a potential model to be used for forecasting.

3.2 Simple Linear Regression

Figure 3.5 above show the fitted line plot for CPI against year. From the figure above it show that CPI and year have straight line relationship that mean if year increase CPI also will increase. From the graph also, it have the value of R^2 that is 99.3% that mean CPI and year have strong relationship between them that will give straight line relationship. The regression equation is CPI = - 3842 + 1.96**Year*.

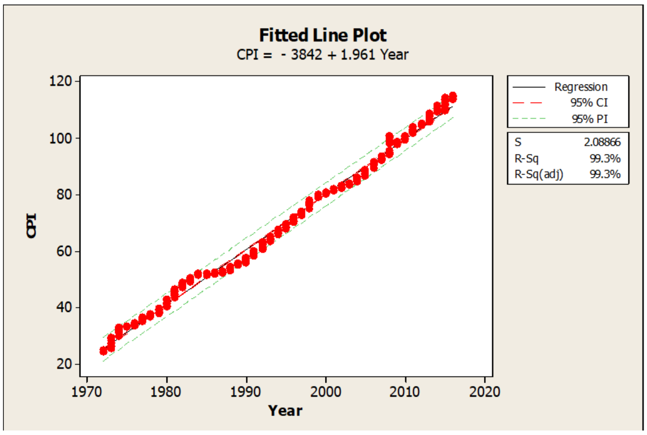


Figure 3.5: Fitted Line Plot of CPI

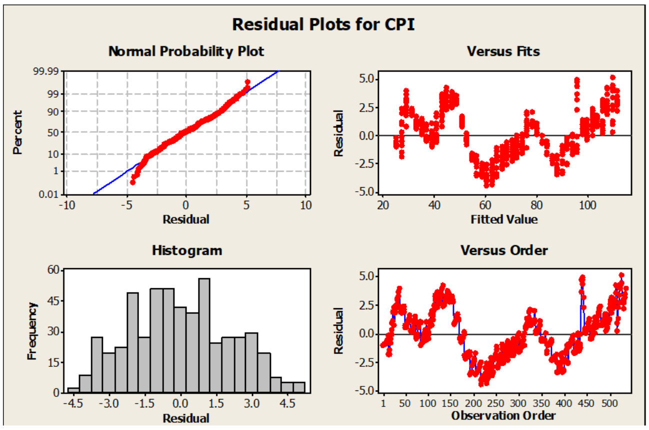


Figure 3.6: Residual Plot for CPI

Figure 3.6 show that the normal probability plot and histogram plot of residuals are almost exactly normally distributed, and it can be also confirmed that there is no correlation in the residuals which means there is no left information in the residuals to be used in fitting a model. Therefore, the model from simple linear regression was successfully selected as a potential model to be used for forecasting.

To compare which one is the best model, we used the values of MSE (Mean Square Error), mean absolute deviation (MAD) and mean absolute percentage error (MAPE).

Table 3.2: Compare the MSE, MAD and MAPE

Model	ARIMA (1, 1, 0)	ARIMA (0, 1, 1)	Simple Linear Regression
RMSE	1.7336	1.7648	3.4059
MAD	1.3433	1.3789	3.1635
MAPE	1.13%	1.16%	2.66%

From the table above, all the model give the highly accurate of forecast that have small value of RMSE and MAD and also all the value of MAPE were less than 10% that mean the forecast will highly accurate. But model ARIMA (1, 1, 0) is more suitable to be used for forecast the future CPI because have the smallest value of RMSE, MAD and MAPE value was less than 10% that give the forecast highly accurate.

Table 3.3: Forecast for ARIMA (1, 1, 0)

Month	Forecast	Actual	Month	Forecast	Actual
September 2016	115.857	115.3	March 2017	116.912	119.6
October 2016	116.050	115.7	April 2017	117.083	119.3
November 2016	116.227	116.9	May 2017	117.254	119.1
December 2016	116.399	116.6	June 2017	117.425	118.9
January 2017	116.571	118.2	July 2017	117.596	118.8
February 2017	116.741	119.7	August 2017	117.766	119.9

4. Conclusions

ARIMA and liner regression model has been considered in this study and comparative study has been carried out. By using RMSE, MAD and MAPE as measured the accuracy of the model, it was found that ARIMA (1, 1, 0) produced high accuracy compare to others model. Therefore, this model can be used as alternative model in order to forecast CPI index for short term period.

References

[1] S Venkadasalam, (2015). The Determinant of Consumer Price Index in Malaysia, *Economics, Business and Management*, 3(12), pp. 1115-1119.

[2] Adams, L.S.O., Awujola, A. and Alumgudu, A. I. (2014). Modelling Nigeria's Consumer Price Index using ARIMA Model. *Development and Economic Sustainability*, *2(2)*, pp. 37-47.

[3] Bédia, F. Aka and Pieretti, P. (2008). Consumer Price Index Dynamics in a Small Open Economy: A Structural Time Series Model for Luxembourg. *Applied Economics 5(1)*, pp. 1-13.

[4] Sulaiman, Y, Mat, N. N. Kand Ghani, A. N. H(2014). Constructing a Consumption Model for Malaysian Consumers, *Marketing Management*, 2(1), pp. 49-60.

[5] Mohamed Mahmoud, L. O. (2015). Consumer Price Index and Economic Growth: A Case Study of Mauritania 1990-2013. *Empirical Research*, *5(2)*, pp. 16-23.

[6] Molebatsi, K. and Raboloko, M. (2016). Time Series Modelling of Inflation in Botswana usingMonthly Consumer Price Index. *Economics and Finance*, *8(3)*, pp.15- 22.

[7] Wee Mei, L. T. (2011). The Relationship between Consumer Price Index (CPI) and Producer Price Index (PPI) in Malaysia, : Technical Report, University Malaysia Sarawak

[8] Chatfield, C.and Prothero, D.L.(1973). Box-Jenkins Seasonal Forecasting: Problems in a Case Study, *Royal Statistical Society*, 136(1), pp. 295-336.

[9] Hipel, K. W., Mcleod, A. I. and Lennox. W. C. (1977). Model Construction, Advances in Box-Jenkins Modelling, *Statistics*, 13(3), pp. 567-574.

[10] Doguwa, S. I. and Alade, S. O. (2013). Short-Term Inflation Forecasting Models for Nigeria. *Applied Statistics*, *4(2)*, pp. 1-29.

[11] Hyndman, R. J., and Athanasopoulos, G. (2014). *Forecasting Principles and Practice.* Australia: OTexts.com.

[12] M. H Kutner, C. J. Nachtsheim, J. Neter, and W. Li, (2005). *Applied Linear Statistical Model*, Fifth ed. New York: McGraw-Hill Companies.

[13] Kock, A.B. (2013). Forecasting the Finish Consumer Price Inflation using Artificial Neural Network Models and Three Automated Model Selection Techniques. *Finnish Economic 26(1)*. Pp 13-24.

Paper ID: SR20719075537 DOI: 10.21275/SR20719075537