

Forecasting the Beef Meat Prices in Erbil Using Box-Jenkins Models

Feink Mohammed Omer¹ & Wasfi Tahir Salih²

¹College of Administration & Economy, Salahaddin University, Erbil, Iraq

²Business and Management Department, Faculty of Administrative Sciences and Economics, Tishk International University, Erbil, Iraq

Correspondence: Wasfi Tahir Salih, Tishk International University, Erbil, Iraq.

Email: wasfi.kahwachi@ishik.edu.iq

Doi: 10.23918/ejmss.v1i1p1

Abstract: Foodstuff has a crucial role for everybody life in the world. In Iraq, beef meat is one of the important parts of the food basket of every household in Erbil. Prices of beef meat fluctuates over time, according to the market prices and the economic situation in the country. The data used were averages of prices of beef meat were originally based on two observations each month took from (Kurdistan Regional Office of Statistics in Erbil city) for the period from January 2008 to December 2018. The methodology used for the forecasting was Box-Jenkins modelling. The best model achieved was ARIMA (0,2,2). Minimum value of Root Mean Square Error (RMSE), (MAE) and WK1 were used to select the best model.

Keywords: Box-Jenkins, Autocorrelation, Root Mean Square Error, New Criterion, Mean Absolute Error

1. Introduction

Time series analysis deals with the problems of identification of basic characteristic features of time series. The applications of time series models in various fields attracted research towards adopting this methodology for forecasting. Some examples are; sales forecasting, weather forecasting, price of foodstuff, oil prices etc. In this paper, the method has been used in analysis of time series is Box-Jenkins, the method which is used to find the best model for forecasting and accurate result. This is a very important method to forecast future values of any data.

2. The Objectives of this Study

1. Using Box-Jenkins models to forecast for 12 months for price beef meat in Erbil.
2. Comparing a new criteria for residual WK1 (developed by Wasfi Kahwachi), with accuracy criteria such as RMSE and MAE based on minimum values for selecting the best model.

3. Data Collection

The data were collected from the Kurdistan Region Statistical Office. The data were on monthly bases from January 2008 until December 2018. The original data were on a bi-observation per month, averaged to get one single observation each month.

4. Time Series

A time series is a set of correlated observations measured sequentially through time. Time series data may be measured continuously or discretely. Auto Correlation Function (ACF) and Partial Autocorrelation Function are mainly used due to the behavior scheme of the data. According to the behavior of ACF and PACF, the suggested model for the time series data can be determined (Chatfield, 1975; McGee, 1999).

The main objectives of analyzing time series using Box-Jenkins models is to build the best model that best represents the data, and to forecast for future observations. Of course, the forecasting does not deliver the exact future values of data that the given time series will really have, but rather their estimates (Popovic, 2005). The data should be checked for stationarity around variance and mean, then analysing to achieve the best model. The resultant model for forecasting should have normal residuals (white noise). The most common Box-Jenkins models are Autoregressive models (AR), Moving Average Models (MA), and the mixed model Auto Regressive Integrated Moving Average models ARIMA (Chatfield, 1975). The criteria used to judge on the best model based on the residuals, are; Mean absolute error (MAE) or Mean absolute deviation (MAD), Root Mean Square Error (RMSE), and a new criterion called WK1 (Wasfi Kahwachi 1) that deals with the residual observations from the model. It was proven by Wasfi Kahwachi, and defined as follows:

$$WK1 = \sum \frac{ei+1}{ei} / n-1 \quad [1]$$

where, e_t = residual, $e_t \sim N(0, \sigma^2)$

5. The Data

The data is originally based on two observations of prices for each month took from Kurdistan Regional Office of Statistics. Averages of these observations were taken for each month, and then data analyses were made for the period of January 2008 to December 2018 monthly. The data analysis was performed using STATGRAPHICS program, and SPSS program.

5.1 Data Description

Beef meat: It is known as red meat. It is popular meat in most countries. It is used in different ways to cook food in different cultures. It has sources of vitamins.

5.2 Data Statistics

Some general descriptive Statistics of the collected data are shown below in Table 1.

Table 1: Descriptive statistics of data

Food stuff	N	Minimum	Maximum	Mean	Std. Deviation
Beef meat	132	10500	17000	14813.45	1693.858

Analyzing Time Series: To obtain the best model, the following steps were followed:

1. Data Plot, time series plots for each time series original to identify some of its initial properties. It was noted that there are fluctuations in beef meat prices plots and trend; it means the time series is not stationary.

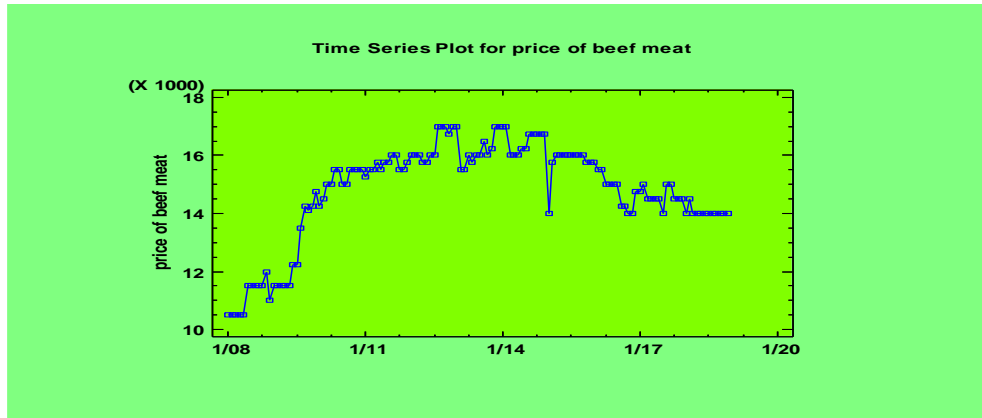


Figure 1: Time series plot for price of beef meat

2. ACF & PACF plots: The ACF and PACF for beef meat prices were examined to know the behavior of the data, from which the model and order of model is obtained. It is shown in Figure (2) and Figure (3), that the ACF behavior is exponential, and the PACF behavior is cut-off.

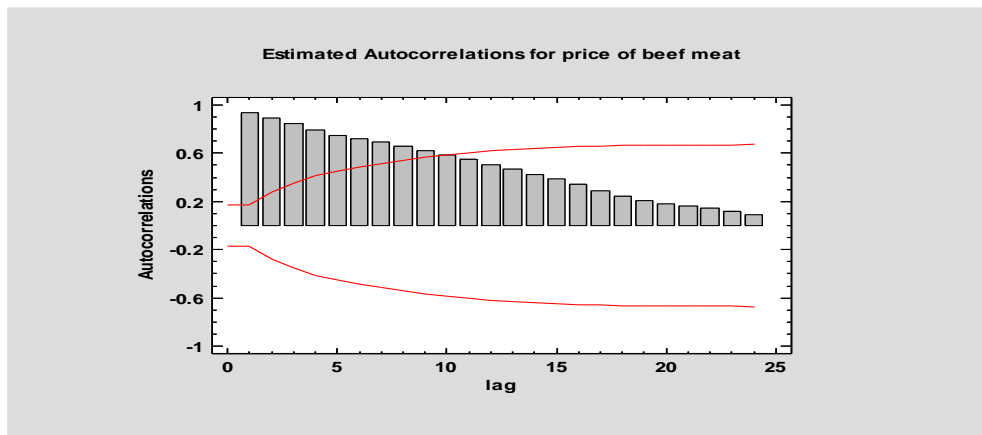


Figure 2: Estimated Autocorrelations for price of beef meat is exponential behavior

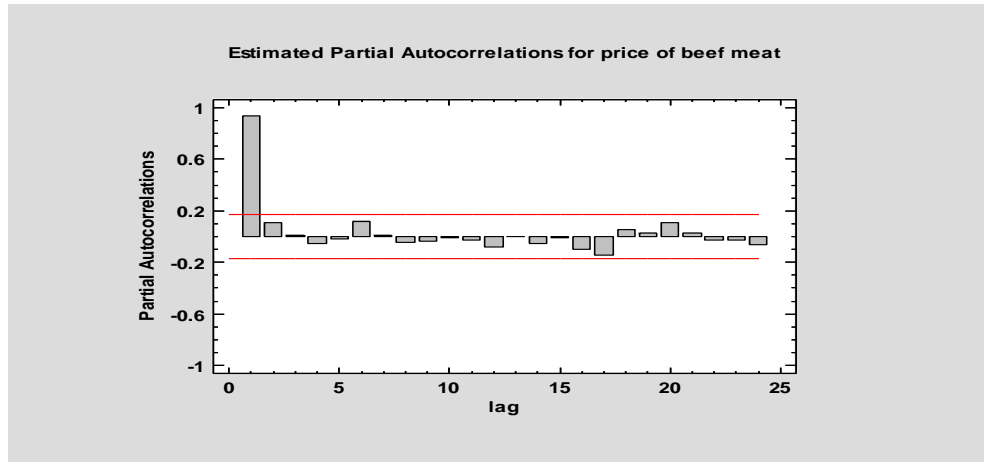


Figure 3: Estimated Partial Autocorrelations for price of beef meat is cut off (1) behavior

3. Randomness test: After plotting the ACF & PACF of price of beef meat, it is needed to test of randomness for the data, because the time series data should be proven random in order to find the best model, for this purpose, Box- pierce test was used. The null hypotheses states that the times series are random, versus the alternative hypotheses that they are not random, in order to know that the data has a certain behavior than having a random behaviour. If p-value is less than or equal to 0.05, that means that the null hypothesis is rejected (i.e. the time series is not random). The hypothesis follows as explained below.

H_0 : The data series is random.

H_1 : The data series is not random.

Table 2: Hypotheses of randomness test, and Box-Pierce test p-value for time series

Series	Hypothesis	P- value
Price of Beef meat	H_0 : the data series is random	0.000
	H_1 : the data series is not random	

After making sure that price of beef meat is not random, because it has p- value less than 0.05, it is needed to achieve stationarity around variance and around mean. Natural logarithm transformation was applied for the data, due to the apparent fluctuations, then to achieve stationarity around mean, a first order difference was taken. After the achieving stationarity, it is needed to check for randomness again, the random series means that all other effects on the data series is removed, in other words, the external effects on data series were unmasked, what remained is the data itself without any sort of effects. Table (3) shows the results.

Table 3: Hypotheses of randomness test, and p-value for data of time series

Series	Hypothesis	P- value
Price of Beef meat	H_0 : the data series is random H_1 : the data series is not random	0.757557

The p-value of Box-pierce test is greater than 0.05 for the data, this means the time series is random. To examine the series after transformation, drawing the data is needed. The properties of the series are not changed with the change of observations over time, which means that there is no trend indicating non-stationary series.

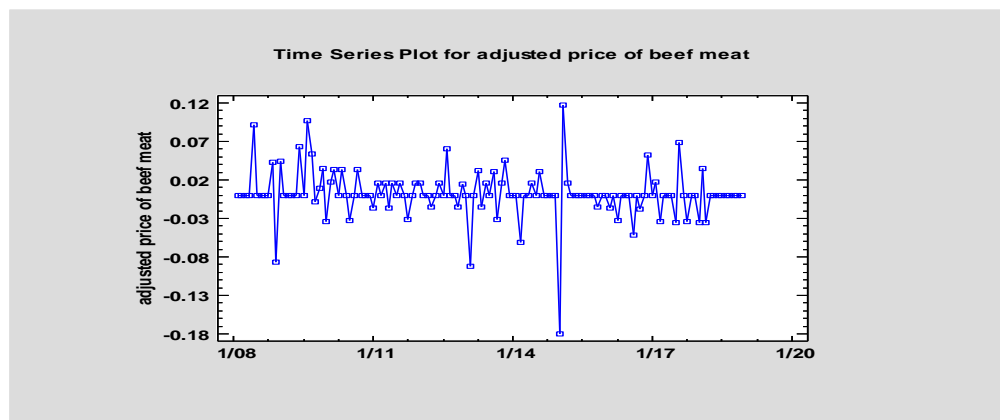


Figure 4: The new price for beef meat plot after taking after taking the natural logarithm and one non-seasonal differencing.

It is not enough to look at the data series plot to determine the stability of the series, so invoking the values and parameters of the ACF & PACF, of the transformed series will be good. It is noted that the values of ACF in second series all values are lying within interval confidence:

$$-1.96 S(r_k) \leq \rho_k \leq 1.96 S(r_k)$$

k: shift time.

$S(r_k)$: The standard deviation of the estimated autocorrelation.

Figure (5) and figure (6) indicate all series are stationary around mean.

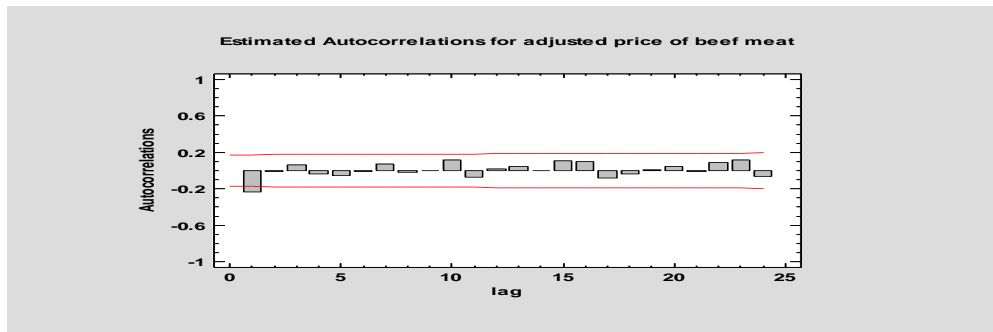


Figure 5: Estimated Autocorrelations for adjusted price of Beef meat.

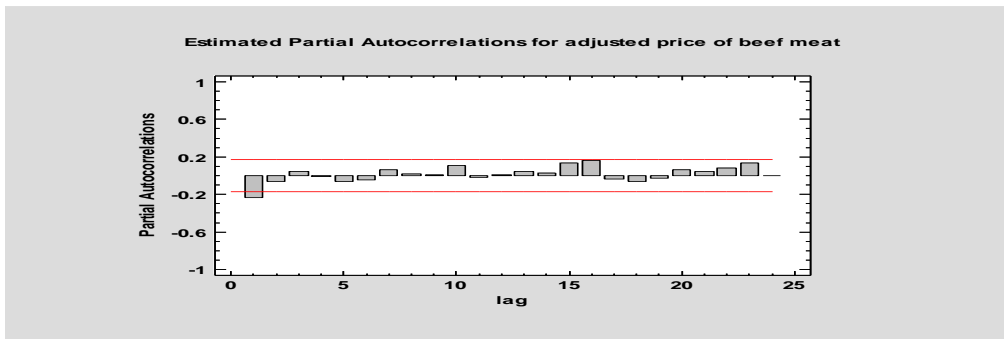


Figure 6: Estimated Partial Autocorrelations for adjusted price of Beef meat

The new resulted data series is taken into account for determining the model. Table 4 shows the estimated autocorrelations between values of adjusted price of beef meat. The lag k of autocorrelation coefficient measures the correlation between values of adjusted price of beef meat at time t and time t-k. In this case, 1 of the 24 autocorrelation coefficients is a statistically significant correlation at lag k at the 95.0% confidence level, implying that the

time series may not be completely random. The lag k partial autocorrelation coefficient measures the correlation values of adjusted price of beef meat price at time and t-k having accounted for the correlations at all lower lags.

Table 4: Estimated Autocorrelations for adjusted price of beef meat

Lag	Autocorrelation	Standard Error	Lower 95.0% Prob. Limit	Upper 95.0% Prob. Limit
1	-0.232036	0.0873704	-0.171243	0.171243
2	-0.00772542	0.0919542	-0.180227	0.180227
3	0.0624317	0.0919592	-0.180237	0.180237
4	-0.0323901	0.0922822	-0.18087	0.18087
5	-0.0503009	0.0923689	-0.18104	0.18104
6	-0.0113019	0.0925778	-0.181449	0.181449
7	0.0721691	0.0925883	-0.18147	0.18147
8	-0.0185423	0.0930167	-0.18231	0.18231
9	0.00245785	0.093045	-0.182365	0.182365
10	0.115417	0.0930455	-0.182366	0.182366
11	-0.0682121	0.094132	-0.184496	0.184496
12	0.0140925	0.0945085	-0.185234	0.185234
13	0.0468407	0.0945246	-0.185265	0.185265
14	-0.00206065	0.0947016	-0.185612	0.185612
15	0.105473	0.094702	-0.185613	0.185613
16	0.0952043	0.0955945	-0.187362	0.187362
17	-0.0843394	0.0963155	-0.188775	0.188775
18	-0.0334493	0.0968776	-0.189877	0.189877
19	0.00936803	0.0969658	-0.19005	0.19005
20	0.0459825	0.0969727	-0.190063	0.190063
21	-0.0126733	0.097139	-0.190389	0.190389
22	0.0887011	0.0971516	-0.190414	0.190414
23	0.118089	0.0977679	-0.191622	0.191622
24	-0.0648673	0.0988507	-0.193744	0.193744

Table 5: Estimated Partial Autocorrelations for adjusted price of beef meat

Lag	Partial Autocorrelation	Standard. Error	Lower 95.0% Prob. Limit	Upper 95.0% Prob. Limit
1	-0.232036	0.0873704	-0.171243	0.171243
2	-0.0650695	0.0873704	-0.171243	0.171243
3	0.048213	0.0873704	-0.171243	0.171243
4	-0.0067879	0.0873704	-0.171243	0.171243
5	-0.0596516	0.0873704	-0.171243	0.171243
6	-0.0447757	0.0873704	-0.171243	0.171243
7	0.062893	0.0873704	-0.171243	0.171243
8	0.019974	0.0873704	-0.171243	0.171243
9	0.00562503	0.0873704	-0.171243	0.171243
10	0.112018	0.0873704	-0.171243	0.171243
11	-0.0138251	0.0873704	-0.171243	0.171243
12	0.00684671	0.0873704	-0.171243	0.171243
13	0.0444302	0.0873704	-0.171243	0.171243
14	0.029175	0.0873704	-0.171243	0.171243
15	0.133989	0.0873704	-0.171243	0.171243
16	0.161949	0.0873704	-0.171243	0.171243
17	-0.0340225	0.0873704	-0.171243	0.171243
18	-0.0657277	0.0873704	-0.171243	0.171243
19	-0.0251435	0.0873704	-0.171243	0.171243
20	0.0600093	0.0873704	-0.171243	0.171243
21	0.046549	0.0873704	-0.171243	0.171243
22	0.0829492	0.0873704	-0.171243	0.171243
23	0.136633	0.0873704	-0.171243	0.171243
24	0.00222822	0.0873704	-0.171243	0.171243

To recognize range of stationary time series around variance, after differencing of order one and natural logarithm for original series, each series is divided 11 parts to calculate means for each series, after that for each 11 parts (years) ranges were obtained. Then ranges vs means were plotted for the data series. If the data are approximately randomly scattered, then it can be claimed that data series is stationary around variance.

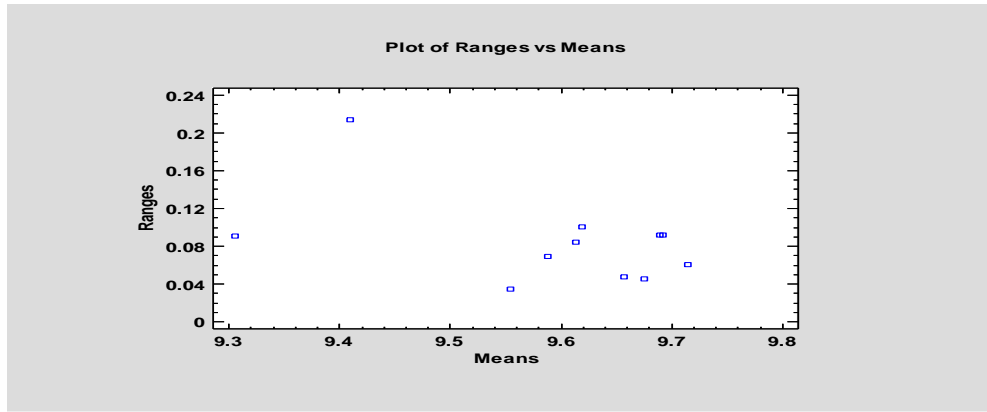


Figure 7: Price of Beef meat for dispersion

5.3 Choosing Appropriate Model

After checking stationarity of the data series around mean and variance, the model identification process is used to determine the best model that describes the data. Model selection was made by depending on the ACF & PACF, but sometimes it is not clear correctly to rely on them for obtaining an adequate model. To suggest the best adequate model for the data series we depend on minimum value of root mean square error (RMSE), (MAE) and WK1.

Table 6: The suggested model for price of beef meat

Model	RMSE	MAE	WK1
ARIMA(1,1,0)	457.562	261.109	0.007687
ARIMA(1,0,0)	467.483	247.29	0.007687
ARIMA(0,2,2)	451.311	272.879	0.007626
ARIMA(2,1,0)	457.67	267.765	0.007687
ARIMA(0,1,2)	458.084	267.216	0.007732

5.4 Model Identification

In this process, the ACF and PACF for the stationary data for the data series were examined. In accordance the best adequate model depending on RMSE, MAE, and WK1 values were obtained. The suggested models were ARIMA (0,2,2) for price of beef meat.

5.5 Estimation

For the best model for the price of beef meat, the parameters are estimated as in Table 7.

Table 7: Function Model and parameter estimate for price of beef meat

Series	Model	Function Model	Parameter estimate			
			ϕ_1	ϕ_2	θ_1	θ_2
Price of Beef meat	ARIMA (0,2,2)	$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$			1.2827	-0.27083

5.6 Model Diagnostic Checking

For a good forecasting model, the residual for fitted model should be simply white noise. Therefore, the ACF and PACF of the residuals were calculated and plotted. Non-significant ACF and PACF indicated the best-fitted model. Besides all that, the fitted model should reflect the most adequate forecasting.

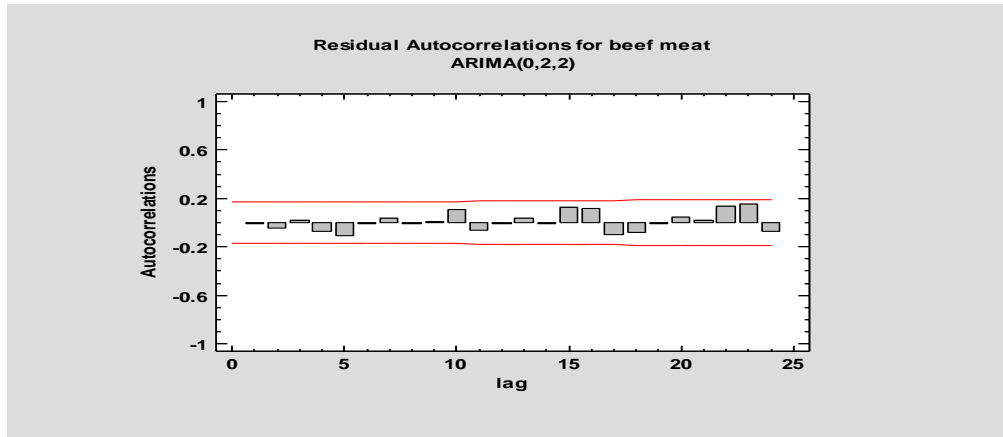


Figure 8: ACF of the residual of price of beef meat.

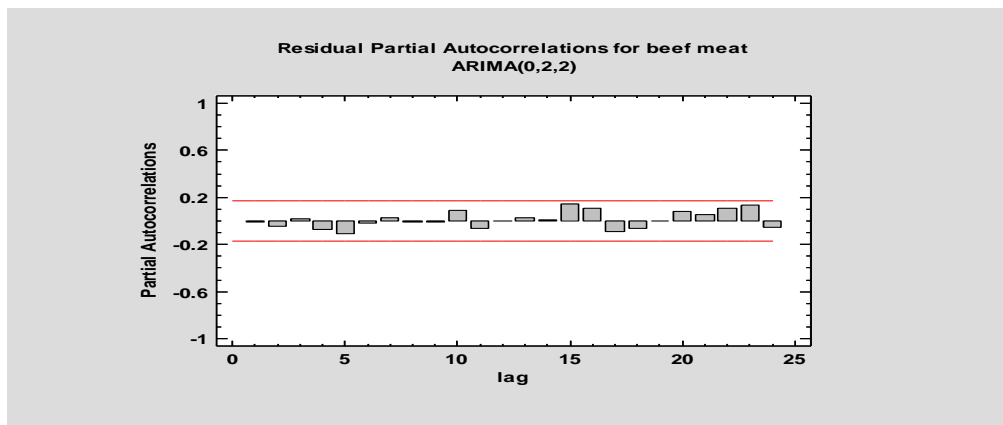


Figure 9: PACF of the residual of price of beef meat.

5.7 Goodness of Fit Test

It is probably possible to know the adequate model ARIMA (0,2,2), using Box-Pierce test. If $P\text{-value} \geq 0.05$, the null hypothesis is not rejected, which means the model above is adequate, meaning that the residuals are random, and the forecasting will be adequate and reliable.

Table 8: Hypothesis and P-value for residual series model

Series	Model	Hypothesis	P-value
Price of Beef meat	ARIMA (0,2,2)	H_0 : Adequate Model H_1 : not Adequate Model	0.766066

5.8 Forecasting

In the last step, after a model has passed the entire diagnostic test, it is adequate to be used for forecasting. To choose the final model for forecasting the accuracy of the models must be higher than that all competing models. The accuracy then, encourages forecasting of all the series 12 months. Prices of forecasts of beef meat series for the 12 months using ARIMA (0,2,2) are shown in Table 9.

Table (9) forecast for price of beef meat

Period	Forecast	Lower 95.0% Limit	Upper 95.0% Limit
1/19	13994.6	13101.4	14887.8
2/19	13990.6	12891.4	15089.7
3/19	13986.6	12719.7	15253.5
4/19	13982.6	12572.5	15392.7
5/19	13978.6	12442.8	15514.5
6/19	13974.7	12326.6	15622.7
7/19	13970.7	12221.1	15720.2
8/19	13966.7	12124.6	15808.8
9/19	13962.7	12035.7	15889.8
10/19	13958.7	11953.3	15964.2
11/19	13954.8	11876.7	16032.8
12/19	13950.8	11805.3	16096.3

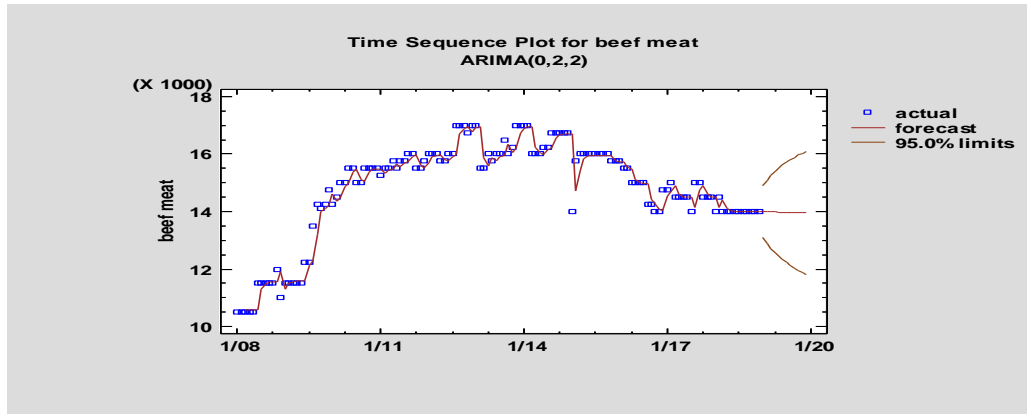


Figure 10: original prices of Beef meat, forecasted values and confidence limits the 2008-2019, forecast (12 Month).

5.9 Test of Normality for Residuals

Table 10 illustrates the normality test for residuals of the forecasts of the beef meat prices using three goodness of fit tests: Kolmogorov-Smirnov test, Anderson-Darling test, and Chi-Squared test. It appears from all these tests that for the data series is distributed normally at $\alpha = 0.01$, and 0.05. which means that the null hypothesis could not be rejected.

H_0 = Residual is normal distribution

H_1 = Residual is not normal distribution

Table 10: Test of normality for residual prices of beef meat

Details	Different Kinds of Normality Test (Goodness of Fit Test)					
	Kolmogorov-Smirnov		Anderson-Darling		Chi-Squared	
Statistic	0.08497		1.0114		4.836	
P-value	0.57295		Not Available		0.43622	
A	0.05	0.01	0.05	0.01	0.05	0.01
Critical Value	0.14868	0.1784	2.5018	3.9074	11.07	15.086
Reject H_0 ?	No	No	No	No	No	No

6. Conclusion

1. According to the results above, forecasts obtained (for 2019) are reliable and can be used for further and several means.
2. The new criterion WK1, has served as a means of determining the residual, besides the (RMSE), (MAE) and WK1, to judge on the accuracy of the chosen model. It depended on the minimum value of these criteria.

References

- Chan, N. H. (2002). Time series applications to finance. Chicago: Wiley-Interscience.
- Chatfield, C. (1975). Time-series forecasting. First Edition. London: Chapman & Hall/Crc.
- Fuller, W. A. (1996). Introduction to statistical time series. Second Edition Ed. New York: A Wiley-Interscience.
- Jürgenwolters, G. K. (2007). Introduction to modern time series analysis. Berlin: Springer.

- Mcgee, R. A. Y. (1999). Introduction to time series analysis and forecasting with applications of SAS and SPSS. New York: Academic Press, Inc.
- Metcalfe, P. S. C. (2009). Introductory time series with R. London: Springer.
- Popovic, A. K. (2005). Computational intelligence in time series forecasting. London: Springer.
- Sollis, R. H. (2003). Applied time series modelling and forecasting. London: Wiley.
- Stoffer, R. H. (2010). Time series analysis and its applications. Third Edition. New York: Springer.
- Tsay, R. S. (2005). Analysis of financial time series. Second Edition. New York: Wiley-Interscience.
- William W. (2006). Time series analysis univariate and multivariate methods. Second Edition Ed. New York: Pearson Addison Wesley.