Journal of Applied Sciences

ISSN 1812-5654

OPEN ACCESS Journal of Applied Sciences

ISSN 1812-5654
DOI: 10.3923/jas.2016.279.285

Research Article
A Robust Composite Model Approach for Forecasting Malaysian Imports: A Comparative Study

Mohamed A.H. Milad and Rose Irnawaty Ibrahim

Department of Mathematical Finance, Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia

Abstract
Objective: With the increasing importance of imports as one of the important factors of economic growth, the current study proposed techniques of more reliable and predictable Malaysian imports of crude material in the future. Specifically, this study proposes composite models for probabilistic imports of crude material forecasting in Malaysia. **Methodology:** In this study, the proposed composite models (With regression processing of heteroscedasticity), (With regression processing of heteroscedasticity and autocorrelation) were employed to extract information that assists in increasing accurate forecasting of the size of the Malaysian imports as well as forecasting engines and compare it with other commonly used models including regression models and ARIMA models. **Results:** The forecasting results of the study showed that the composite model (With regression processing of heteroscedasticity) approach provides more probabilistic information for improving forecasting of Malaysian imports of crude material. **Conclusion:** The results also showed two sets of benefits: The main benefit is that the composite model (Without regression processing) is capable of solving the problem of autocorrelation in residuals but it was unable to solve heteroscedasticity in the residuals. The second benefit is processing the problem of autocorrelation in the composite model in a case when it is not processed in the regression model. However, in the case of the emerging problem of the heteroscedasticity, it can be processed in the regression model prior to the composite model formation.

Key words: Malaysia imports forecasting, composite models (With regression processing of heteroscedasticity), (With regression processing of heteroscedasticity and autocorrelation)

Received: March 22, 2016 Accepted: April 15, 2016 Published: May 15, 2016

Citation: Mohamed A.H. Milad and Rose Irnawaty Ibrahim, 2016. A robust composite model approach for forecasting malaysian imports: A comparative study. J. Applied Sci., 16: 279-285.

Corresponding Author: Mohamed A.H. Milad, Department of Mathematical Finance, Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

To crystallize the problem of searching or finding a robust model that can increase the accuracy of the predictability of the size of imports in Malaysia in the future as well as forecasting engines, the absence of a robust model may affect planning the imports in the future. Forecasting future values of economic variables are some of the most critical tasks of a country.

Several studies have been conducted for developing a model to predict the Malaysia imports using a variety of statistics methods[1-7]. These studies used an Ordinary Least Square (OLS) regression method in which the response variable is the value of imports and the explanatory variables are sets of variables. Osman[8] carried out a study that aimed to determine the best fitted model among the methods of exponential smoothing. Shabri *et al.*[9] also developed a model for prediction of yields of rice imports in Malaysia using three methods of predictable of Artificial Neural Network (ANN), the statistical the autoregressive integrated moving average and the double exponential smoothing.

The overall purpose of this study was to propose robust composite models to forecasting the values of imports in Malaysia and the proposed approaches were compared empirically with each other and with other used methods in terms of the measurement criteria on the forecasting performance.

MATERIALS AND METHODS

Materials: This section describes the case study that demonstrates the effectiveness of the proposed approach through comparisons with other models. This case study is described in the following two sub-sections:

Data collection: The study used data on imports of (CM) in Malaysia to demonstrate the effectiveness and reliability of the proposed forecasting approach. The data covered a period of 23 years, starting from the 1st quarter of 1991 until the 3rd quarter in 2013 as shown in Fig. 1 and Table 1 along with data definitions and sources.

Evaluation indices for forecasting performance: For evaluation of the proposed approaches and other models, the current study used test significance of parameters and two statistical indices as a means to measuring the forecasting accuracy. The first statistical indices are known as Theil's inequality coefficients in which small values are indicative of the high forecast performance. This coefficient is also confined to the interval between 0 and 1 and its values of 0 and 1 indicate a perfect predictor and a perfect inequality, respectively. The coefficient illustrated as follows:

$$U = \sqrt{\frac{1}{n}\sum\left(y_t - \hat{y}_t\right)^2} \Big/ \sqrt{\frac{1}{n}\sum \hat{t}_y^{2} + \sqrt{\frac{1}{n}\sum y_t^2}}$$

where, y_t is the real data for a time period t, $\hat{y}_t$ the predicted value at the same time point and n is a number of periods[10].

The second one is known as the predicted R-square, which has an indication of how well the model is capable of predicting responses to new observations. The higher value of predicted R^2 indicates that the developed model is more capable of prediction. It ranges between 0 and 1 and is calculated. If R^2 (Pred) = 0, the predictive performance is badness, if U = 1, there is a perfect fit[11]. The formula is:

$$R^2(\text{Pred}) = 1 - \left[\text{PRESS} \Big/ \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2\right]$$

Where:

$$\text{PRESS} = \sum_{i=1}^{n} e_i^2$$

Methods: This section introduces the proposed methods and the individual methods, including the multiple linear

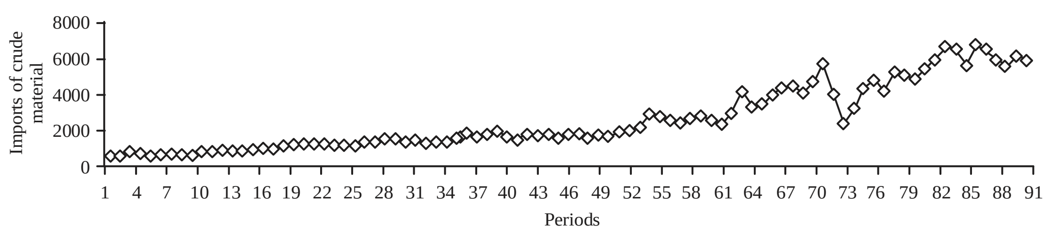

Fig. 1: Time series of imports of crude material, for the period, Q11991-Q32013

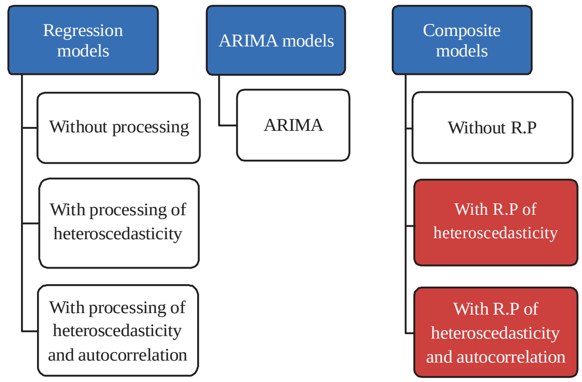

Fig. 2: Overall for forecasting models

Table 1: Data definition and sources

Variable	Definition	Source
Value of imports of (CM)	Value of imports of (CM) from Malaysia, unit (Million RM)	Department of Statistics Malaysia
Value of exports of (CM)	Value of export of (CM) from Malaysia, unit (Million RM)	Department of Statistics Malaysia
PPI for the domestic economic	The PPI (producer price index) for the domestic economic is a composite index based on the price data derived from that of local production and import price indices of (CM), 2005 = 100	Department of Statistics Malaysia
GDP	Gross domestic product, 2005 = 100	Department of Statistics Malaysia
Exchange rate	Exchange rate, against the USD	Bank Negara Malaysia
Tariff tax	The average of tariff tax on imports of (CM)	Royal Malaysian customs department
Sales tax	The average sales tax of imports of (CM)	Royal Malaysian customs department

regression, the ARIMA models and the composite model. Then, it provides a description of the operating process of the proposed composite models. Moreover, the forecasting models and proposed approaches in this study are presented in Fig. 2 where, the proposed models are shown by the red squares and those identified models are marked by the white squares in the case study of (C.M).

Multiple linear regression method: Multiple linear regression is a generalization of linear regression that takes into account more than one independent variable. On practice, the most commonly used model is the general multiple regression model that has multiple explanatory variables. Nevertheless, OLS estimation of regression weights in multiple regressions can be influenced by the emergence of outliers, non-normality, heteroscedasticity and multicollinearity. The model of multiple linear regression can be represented as follows:

$$Y = \beta_0+\beta_1X_{1t}+\beta_2X_{2t}+......+\beta_kX_{kt}+E_t$$

where, Y is the development trend of profession (expressed by demand generally), $X_{1t}+X_{2t}$,......, X_{pkt} are the influence factors of profession development trend (explanatory variables); β_0, β_1, β_2, ... , β_k are the regression coefficients, E_t is the random error, its mean is zero and constant variance σ and E_t is also assumed to be uncorrelated[12,13].

ARIMA (p, d, q) model: It means an autoregressive integrated moving average model. The ARIMA process consists of three processes: AR(p) process that accounts for the memory past event, an integrated process I(d) that accounts for making the data stationary and MA(q) that accounts for a finite sum of forecasting error terms. Seasonal and non-seasonal ARIMA models: The general non-seasonal model is known ARIMA (p, d, q) and the general seasonal model is known as ARIMA (p, d, q) (P, D, Q)s. Where (p, d, q) represents the non-seasonal part of the model and (P, D, Q)s represents the seasonal part of the model, when s is the number of periods per season. In this model, seasonal differencing of appropriate order[14] is used to remove non-stationary from the series. A first order seasonal difference is the difference between an observation and the corresponding observation from the previous year. For monthly time series S = 12 and for quarterly time series S = 4. This model is generally termed as the SARIMA (p, d, q) (P, D, Q)s. An ARIMA model can be defined as:

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{1-q}$$

Model coefficients ϕ_1, ... ϕ_p for AR (p) and θ_1,......θ_q for MA(p). If we write the model in the equation above using the backshift operations, the model is given by Ali[15]:

$$\left(1-\varnothing_1 B - \varnothing_2 B^2 - \varnothing_p B^p\right) y = c + (1-\theta_1 B - \theta_2 B^2 - \theta_q B^q) e_t$$

where, C is the constant term, $\varnothing_i$ = jth is the autoregressive parameter, θ_i = jth is the moving average parameter, e_t is the error at time t and B_k = kth is the order backward shift operator. Seasonal ARIMA (P, D, Q) parameters may also be identified for specific time series data. These are the seasonal autoregressive (P), the seasonal differencing (D) and the seasonal moving average (Q). The general expression of the seasonal ARIMA model (p, d, q) (P, D, Q) is given by the following:

$$\varnothing_{AR}(B)\,\varnothing_{SAR}(B^s)\,(1\text{-}B)^d\,(1\text{-}B_s)^D\,y_t = \theta_{MA}(B)\theta_{SMA}(B^s).e_t$$

where, S is No. of periods in season, $\varnothing_{AR}$ is non-seasonal autoregressive parameter and θ_{AM} is non-seasonal. The moving average parameter is the seasonal moving average parameter.

Composite model: The composite model refers to a combination of forecasts from regression and ARIMA models together. The advantage of the composite model is that, in the most cases, it outperforms any of the individual forecasts. The composite model (Combining-regression and ARIMA models) has been well documented in previous research[16]. It can also be explained as follows: Suppose that we forecasted the variable y_t, this would include those independent variables as follows:

$$Y_t = \beta_0+\beta_1X_1+\beta_2X_2+.....+\beta_pX_p+\varepsilon_t$$

This equation has an additive error term that accounts for unexplained variance in, that is, it accounts for that part of the variance of which is not explained by X_1, X_2, ..., X_p. The equation above can be estimated using a regression analysis as one source of forecast error would come from the additive noise term whose future values cannot be predicted. One effective application of time series analysis is constructing an ARIMA model for the residual series of this regression. This is followed by substituting the ARIMA model for the implicit error term in the original regression equation. It would also enable us to forecast the error term using the ARIMA model. The ARIMA model provides some information about what future values of are likely to be. For instance, it helps to "Explain" the unexplained variance in the regression equation. The combined regression- time series model is:

$$Y_t = \beta_0+\beta_1X_1+\beta_2X_2+.....+\beta_p X_p+\varnothing^{-1}(B)\theta(B)\eta_t$$

where, Y_t is dependent variable, X_1, X_2, ..., X_p are independent variables, β_0, $\beta_1\beta_2$,......., β_p are regression parameters, ø, θ are AR and MA parameters and η_t is error random variable.

RESULTS

In this study, all the different models were compared and the general steps followed in conducting a comparison among the models in the present study were testing the significance of the estimated parameters and measuring the forecasting error.

First step: Testing the significance of the estimated parameters: This stage focused on testing the significance of the estimated coefficients of Malaysia's imports of (CM) models and results showed that they are significant at difference levels. Table 2-8 present the for the estimated coefficients of all Malaysia's imports of (MC). The p-values for the estimated coefficients of ARIMA models, regression model (With processing of heteroscedasticity) and the composite model (With R.P of heteroscedasticity) is less than 0.05, thus indicating that they are highly significant. However, the p-value for at least one of the estimated coefficient of the regression model (Without processing), regression model (With processing of heteroscedasticity and autocorrelation), composite model (Without R.P), composite model (With R.P of heteroscedasticity and autocorrelation) was higher than 0.05. This value indicates that they are not significant at a level of 0.05. Therefore, we had to drop or exclude them from the next phase.

Second step: Measuring the forecast error: Table 9 shows the results concerning the comparative forecasting performance of the different investigated models obtained from the first phase of ARIMA models, regression model (With processing of heteroscedasticity) and the composite model (With R.P of heteroscedasticity) are presented. These results include U-tial test and predicted-R^2. The results obtained from the robustness evaluation of the different

Table 2: Test of significance for estimated parameters of regression model (Without processing)

Model	Parameters	Estimate	Standard error	t-value	Significance value
Regression model without processing	β_0	-1047.5	150.9	-6.94	0.000
	β_2	6.947	3.708	1.87	0.064
	β_3	0.021	0.002	14.00	0.000
	β_4	0.254	0.083	3.06	0.003
	β_6	-65763	20032	-3.28	0.001

Source: Own data calculations

Table 3: Test of significance for estimated parameters of regression model (With processing of heteroscedasticity)

Model	Parameters	Estimate	Standard error	t-value	Significance value
Regression model	β_0	10.174	3.745	2.716	0.008
(With processing of	β_2	-969.16	140.663	-6.890	0.000
heteroscedasticity)	β_3	0.0205	0.001	13.938	0.000
	β_4	0.153	0.070	2.189	0.031
	β_6	-61327	19529	-3.140	0.002

Source: Own data calculations

Table 4: Test of significance for estimated parameters of regression model (With processing of heteroscedasticity and autocorrelation)

Model	Parameters	Estimate	Standard error	t-value	Significance value
Regression model	β_0	5.61	3.071	1.826	0.071
(With processing of	β_2	-954.710	164.700	-5.800	0.000
heteroscedasticity	β_3	0.021	0.002	12.700	0.000
and autocorrelation)	β_4	0.209	0.074	2.840	0.006
	β_6	-56271.0	22520	-2.500	0.014

Source: Own data calculations

Table 5: Test of sgnificance for estimated parameters of ARIMA models

Model	Parameters	Estimate	Standard error	t-value	Significance value
ARIMA models (0, 1, 0) (1, 1, 0)	Seasonal θ_1	-0.535	0.091	-5.907	0.000

Source: Own data calculations

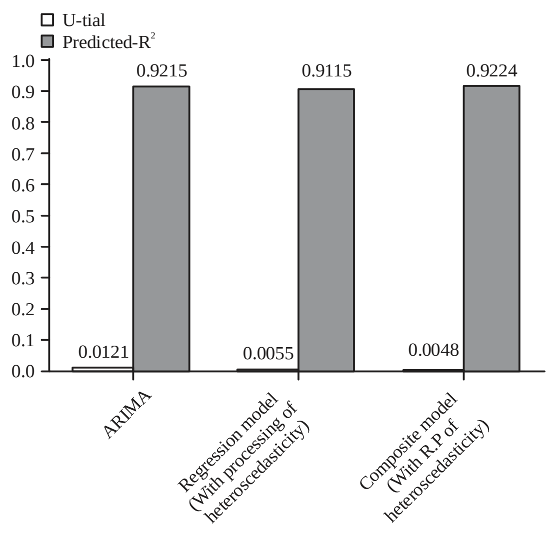

Fig. 3: Comparison of forecasting performance of different models

methods are shown in Fig. 3 where, the forecasting performances of these different models are further compared and where each bar is indicative of the number of best forecast yielded by the corresponding model in terms of a specified accuracy measure.

After the proposed approaches were illustrated through the experiments, we conducted more analyses of the relevant issues in the following section.

DISCUSSION

This section discusses the forecasting performance of the proposed approaches in more details. Such performance was initially analyzed based on the results obtained from the previously mentioned experiments in last section and then, deep insights into the performance were provided using the composite model.

Experimental design and methodologies used were intended to make the experimental forecasts of imports of (CM). This was followed by assessing or evaluating the forecasting performances by testing the significance of the estimated parameters and two main measurements criteria.

The results in Table 9 show that the U-tial and predicted-R^2 values of composite model (With R.P of heteroscedasticity) are 0.0048 and 0.9224 for time series of imports of (CM), respectively. While the first value is evidently lower than that of other methods, the second value is higher than that of other methods. In comparing among all models of the current study, the performance of the composite model (With R.P of heteroscedasticity) was the best as indicated by its better fit, while in an earlier study, it was found that the best performance was achieved by the composite model (With R.P)[11]. The reason for this is that this earlier study[12] aimed to compare between the composite model (Without R.P) and the composite model (With R.P). However, the composite model (Without R.P) is able to solve the problem of autocorrelation in results and it can also increase the accuracy in forecasting. This finding is in agreement with findings of some previous studies[16-21]. Because these studies have indicated that the composite model (Without R.P) contributes to the increased forecasting accuracy.

Table 6: Test of significance for estimated parameters of composite model (Without R.P)

Model	Parameters	Estimate	Standard error	t-value	Significance value
Composite model (without R.P)	β_0	-1047.5	150.9	-6.94	0.000
	β_2	6.947	3.708	1.87	0.064
	β_3	0.021	0.002	14.00	0.000
	β_4	0.254	0.083	3.06	0.003
	β_6	-65763	20032	-3.28	0.001
	MA, θ_1	-0.364	0.108	-3.359	0.001
	AR, seasonal, $\emptyset_1$	-0.304	0.149	-2.042	0.044
	MA, seasonal, θ_2	0.576	0.139	4.142	0.000

Source: Own data calculations

Table 7: Test of significance for estimated parameters of composite model (With R.P of heteroscedasticity)

Model	Parameters	Estimate	Standard error	t-value	Significance value
Composite model (With R.P of heteroscedasticity)	β_0	10.174	3.745	2.716	0.008
	β_2	-969.16	140.663	-6.890	0.000
	β_3	0.0205	0.001	13.938	0.000
	β_4	0.153	0.070	2.189	0.031
	β_6	-61327	19529	-3.140	0.002
	AR, $\emptyset_1$	0.393	0.124	3.182	0.002
	MA, θ_1	0.994	0.477	2.083	0.040
	MA, seasonal, θ_2	0.776	0.087	8.918	0.000

Source: Own data calculations

Table 8: Test of significance for estimated parameters of composite model (With R.P of heteroscedasticity and autocorrelation)

Model	Parameters	Estimate	Standard error	t-value	Significance value
Composite model (With R.P of heteroscedasticity and autocorrelation)	β_0	5.61	3.071	1.826	0.071
	β_2	-954.710	164.700	-5.800	0.000
	β_3	0.021	0.002	12.700	0.000
	β_4	0.209	0.074	2.840	0.006
	β_6	-56271	22520	-2.500	0.014
	MA, θ_1	-0.240	0.108	-2.216	0.029
	MA, seasonal, θ_2	0.797	0.078	10.243	0.000

Source: Own data calculations

Table 9: Statistical measures of forecast error for the Malaysia imports of (CM) models

	ARIMA	Regression model (With processing of heteroscedasticity)	Composite model (With R.P of heteroscedasticity)
U-tial	0.0121	0.0055	0.0048
Predicted-R^2	0.9215	0.9115	0.9224

Source: Own data calculations

From Table 2-9, the results underlie several interesting conclusions. First, it can be concluded that the composite model (Without R.P) is able to solve the problem of autocorrelation in residuals but it cannot solve heteroscedasticity in the results. Secondly, when this method is used as in the case of the emerging problem of autocorrelation, such problem cannot be processed in the regression model but it can be processed only through the composite model. However, in case of the emergence of the heteroscedasticity problem, it is possible to be processed in the regression model prior to construction of the composite model.

CONCLUSION AND FUTURE RECOMMENDATIONS

This study, on the basis of composite model, also proposed two approaches called composite model (With R.P of heteroscedasticity) and composite model (With R.P of heteroscedasticity and autocorrelation) for imports of (CM) in Malaysia. Empirical illustration and comparison of these approaches to other methods were made based on time series of (CM) at Malaysia. Finally, some relevant issues are discussed and conclusions are drawn.

The study contributes to previous research in that it presents the first work using the composite model (With R.P of heteroscedasticity) and (With R.P of heteroscedasticity and autocorrelation) in this research area. Such composite models are proved to be robust forecasting methods aiming at improving or increasing the accuracy of the predictability of the value of imports of (CM) and forecasting engines in the context of Malaysia. In this study, it was observed that when such methods were applied, the followings should be considered especially in the case of the emergence of the problem of autocorrelation. First, this problem cannot be processing in the regression model but only by employing the composite model (Without R.P). However, such emerging problem of heteroscedasticity is processed in the regression model prior to forming the composite model.

Therefore, future studies would gain better benefits by focusing the investigation on the application of other methods to imports of crude material in forecasting by utilizing data from a wider sample of imports in the context of Malaysia and comparison of them to the composite model.

ACKNOWLEDGMENTS

The authors are grateful to the Faculty of Science and Technology, University Science Islam Malaysia (USIM) in providing the facilities to carry out the research. They also thank and appreciate the efforts by the Journal editors and reviewers in advance.

REFERENCES

1. Alias, M.H., 1978. A demand equation for West Malaysian imports. Akademika, 12: 53-64.
2. Semudram, M., 1982. A macromodel of the Malaysian economy: 1959-77. Dev. Econ., 20: 154-172.
3. Abu Bakar, N., 2000. Time series estimation of Malaysia's export and import demand: A dynamic old method. Analisis, 7: 61-77.
4. Tang, T.C. and M. Nair, 2002. A cointegration analysis of Malaysian import demand function: reassessment from the bounds test. Applied Econ. Lett., 9: 293-296.
5. Applanaidu, S.D.A.P., F.M. Arshad, M.N. Shamsudin and A.A. Abdel Hameed, 2011. An econometric analysis of the link between biodiesel demand and Malaysian palm oil market. Int. J. Bus. Manage., 6: 35-45.
6. Prambudia, Y. and M. Nakano, 2012. Exploring Malaysia's transformation to net oil importer and oil import dependence. Energies, 5: 2989-3018.
7. Mohamad, J., 2012. The impact of tariff reductions on real imports in Malaysia from 1980-2010: An empirical study. J. Graduate School Asia-Pac. Stud., 24: 181-199.
8. Osman, L.I.B., 2012. Best fitted model to forecast the trade balance of Malaysia. Faculty of Computer Science and Mathematics, Shah Alam, Selangor. http://www.termpaperwarehouse.com/essay-on/Best-Fitted-Model-To-Forecast-The/103307
9. Shabri, A., R. Samsudin and Z. Ismail, 2009. Forecasting of the rice yields time series forecasting using artificial neural network and statistical model. J. Applied Sci., 9: 4168-4173.
10. Stekler, H.O., 1968. Forecasting with econometric models: An evaluation. Econometrica, 36: 437-463.
11. Milad, M.A.H., R.I. Ibrahim and S. Marappan, 2015. A comparison among two composite models (without regression processing) and (with regression processing), applied on Malaysian imports. Applied Math. Sci., 9: 5757-5767.
12. Milad, M.A.H., R.I. Ibrahim and S. Marappan, 2015. Regression analysis to forecast Malaysia's imports of crude material. Int. J. Manage. Applied Sci., 1: 121-130.
13. Ul-Saufie, A.Z., A.S. Yahaya, N.A. Ramli and H.A. Hamid, 2012. Performance of multiple linear regression model for long-term PM_{10} concentration prediction based on gaseous and meteorological parameters. J. Applied Sci., 12: 1488-1494.
14. Adhikari, R. and R.K. Agrawal, 2013. An Introductory Study on Time Series Modeling and Forecasting. Lap Lambert Academic Publishing, Germany, ISBN: 9783659335082, Pages: 76.
15. Ali, S.M., 2013. Time series analysis of Baghdad rainfall using ARIMA method. Iraqi J. Sci., 54: 1136-1142.
16. Quintos, C.E., 1998. Analysis of cointegration vectors using the GMM approach. J. Econometrics, 85: 155-188.
17. Shamsudin, M.N. and M.A. Fatimah, 1990. Composite models for short term forecasting for natural rubber prices. Petanika, 12: 283-288.
18. Islam, A., 2007. Explaining and forecasting investment expenditure in Canada: Combined structural and time series approaches, 1961-2000. Applied Econometrics Int. Dev., 7: 75-89.
19. Khin, A.A., E.C.F. Chong, M.N. Shamsudin and Z. Mohammed, 2008. Natural rubber price forecasting in the world market. Proceedings of the International Conference on Agriculture Extension, June 15-19, 2008 Putrajaya, Malaysia, pp: 92-104.
20. Khin, A.A., M. Zainalabidin and S.M. Nasir, 2011. Comparative forecasting models accuracy of short-term natural rubber prices. Trends Agric. Econ., 4: 1-17.
21. Khin, A.A., M. Zainalabidin, C.A.N. Malarvizhi and S. Thambiah, 2013. Price forecasting methodology of the Malaysian palm oil market. Int. J. Applied Econ. Finance, 7: 23-36.