Malaysian Journal of Computing, 5(1): 374-389, 2020

eISSN: 2600-8238 online

A COMPARATIVE STUDY BETWEEN UNIVARIATE AND BIVARIATE TIME SERIES MODELS FOR CRUDE PALM OIL INDUSTRY IN PENINSULAR MALAYSIA

Pauline Jin Wee Mah and Nur Nadhirah Nanyan

Centre of Statistical and Decision Science Studies,
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA
40450 Shah Alam, Selangor, Malaysia
pauli874@fskm.uitm.edu.my

ABSTRACT

The main purpose of this study is to compare the performances of univariate and bivariate models on four-time series variables of the crude palm oil industry in Peninsular Malaysia. The monthly data for the four variables, which are the crude palm oil production, price, import and export, were obtained from Malaysian Palm Oil Board (MPOB) and Malaysian Palm Oil Council (MPOC). In the first part of this study, univariate time series models, namely, the autoregressive integrated moving average (ARIMA), fractionally integrated autoregressive moving average (ARFIMA) and autoregressive (ARAR) algorithm were used for modelling and forecasting purposes. Subsequently, the dependence between any two of the four variables were checked using the residuals' sample cross correlation functions before modelling the bivariate time series. In order to model the bivariate time series and make prediction, the transfer function models were used. The forecast accuracy criteria used to evaluate the performances of the models were the mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE). The results of the univariate time series showed that the best model for predicting the production was ARIMA (1,1,0) while the ARAR algorithm were the best forecast models for predicting both the import and export of crude palm oil. However, ARIMA (0,1,0) appeared to be the best forecast model for price based on the MAE and MAPE values while ARFIMA (0, 0.08903, 0) emerged the best model based on the RMSE value. When considering bivariate time series models, the production was dependent on import while the export was dependent on either price or import. The results showed that the bivariate models had better performance compared to the univariate models for production and export of crude palm oil based on the forecast accuracy criteria used.

Keywords: *Crude palm oil, ARIMA, ARFIMA, bivariate time series.*

Received for review: 05-09-2019; Published: 06-03-2020

1. Introduction

Malaysia is the second largest palm oil producer in the world after Indonesia (Sawe, 2018). As one of the largest producers and exporters of palm oil and palm oil products, Malaysia plays an important role in fulfilling the growing global demand for fats and oils (Malaysian Palm Oil Council, 2018). However, Malaysia has been suffering from low production efficiency for over 20 years (Lee, 2011). In 2009, crude palm oil production in Malaysia declined to 17.56 million tonnes from 17.73 million tonnes in 2008 due to the stress of the trees after heavy production in 2008 which registered a sharp increase of 12.1% from 2007 (Abdullah &Wahid, 2011). The crude palm oil production for Sabah in the third quarter of 2017, from July to September, declined 7.2% to 1.33 million tonnes from 1.43 million tonnes for the same period in 2016 (The Edge Financial Daily, 2017). According to the New Straits Times (2018), the production of Malaysia's palm oil was expected to drop to 19.8 million tonnes in 2018 as compared to 19.92 million tonnes recorded in 2017 due to lower yields. The scenario mentioned above would be worth exploring as crude palm oil industry is one of the main contributors to Malaysia's economy. The Nikkei Markets (2019) stated that the palm oil yields in Malaysia are on the downtrend due to management, lack of replanting and cutting of fertilizers. The trees covering 1.2 million hectares, or about 20% of the total oil palm in Malaysia aged 20 years or older. The price of palm oil is also expected to grow high in 2019 as the production of palm oil slows. The Star Online (2018) reported that the average price of the crude palm oil fell 1.4% month-on-month to RM2,183 per tonne in August 2018.This was due to concern over weak demand for palm oil partly caused by India's move to raise import duties on crude palm oil by 14% points to 44%, resulting in lower palm oil exports to the country and strong palm oil supplies in Indonesia.

As the production of Malaysia's palm oil was expected to drop in 2018 due to lower yields (New Straits Times, 2018), forecasting would be the main consideration in this study. Although many studies related to the crude palm oil industry were carried out, but forecasting was not considered in those studies. Some recent examples of the studies include the Life Cycle Assessment (LCA) (Andarani, Nugraha, Sawitri & Budiawan, 2018, Subramaniam *et al.*, 2010 and Vijaya, Ma, Choo & Sulaiman, 2008), an activity based costing method (Sembiring, Wahyuni, Sinaga & Silaban, 2018), a system dynamic tool called structural thinking, experiential learning laboratory with animation (STELLA) based on simulation (Otieno *et al.*, 2016) and the Cobb-Douglas (C-D) production function that fulfilled the C-D assumptions (Norhidayu, Syazwani, Radzil, Amin & Balu, 2017) as well as the time series analysis method which included Johansen cointegration technique, error correction model and Granger causality tests (Asari *et al.*, 2011). Since time series is a sequence of data points collected over time, the use of time series analyses is very crucial in order to obtain an understanding of the underlying forces and structure produced by the data, which would eventually be used in forecasting or monitoring (Farrelly, 2017).

Hence, time series analysis has become an important tool applied in different fields for the purpose of modelling and forecasting like some recent studies in the field of agriculture (Mah, Zali, Ihwal & Azizan, 2018; Mah, Buhary, Abdullah & Saad, 2018). Among the most effective approaches used in analysing time series data is the model introduced by Box and Jenkins, i.e., autoregressive integrated moving average (ARIMA) (Lazim, 2011). As there were not many studies done on crude palm oil production, price, import and export using univariate time series models, the ARIMA, fractionally integrated autoregressive moving average (ARFIMA) and autoregressive (ARAR) algorithm will be considered in this study. Therefore, the purpose of this study is to find suitable univariate time series models that could predict the production, price, import and export of crude palm oil industry in Peninsular Malaysia. Besides univariate time series, bivariate time series is also considered as alternative models because the dependence among the variables may affect the crude palm oil industry in Peninsular Malaysia as well. Therefore, bivariate time series modelling using transfer function model will also be explored in this study.

The organisation of this paper is as follows. The data set and time series modelling procedures will be included in the Methodology of Section 2. The results of the study will be presented and discussed in Section 3 while the conclusions are contained in Section 4.

2. Methodology

2.1 The Data Set

This study utilized data containing 81 complete sets of monthly observations from January 2012 to September 2018. The data used in this study were obtained from the Malaysian Oil Palm

Statistics 2012-2018 which was launched by the Malaysian Palm Oil Board (MPOB) and Malaysian Palm Oil Council (MPOC). The data consisted of four variables which are the production (in tonnes), price (in RM/tonnes), import (in tonnes) and export (in tonnes) of crude palm oil. These four variables were chosen because the price of crude palm oil usually would depend on the production, import and export of crude palm oil. The flow of the study is as displayed in Figure 1.

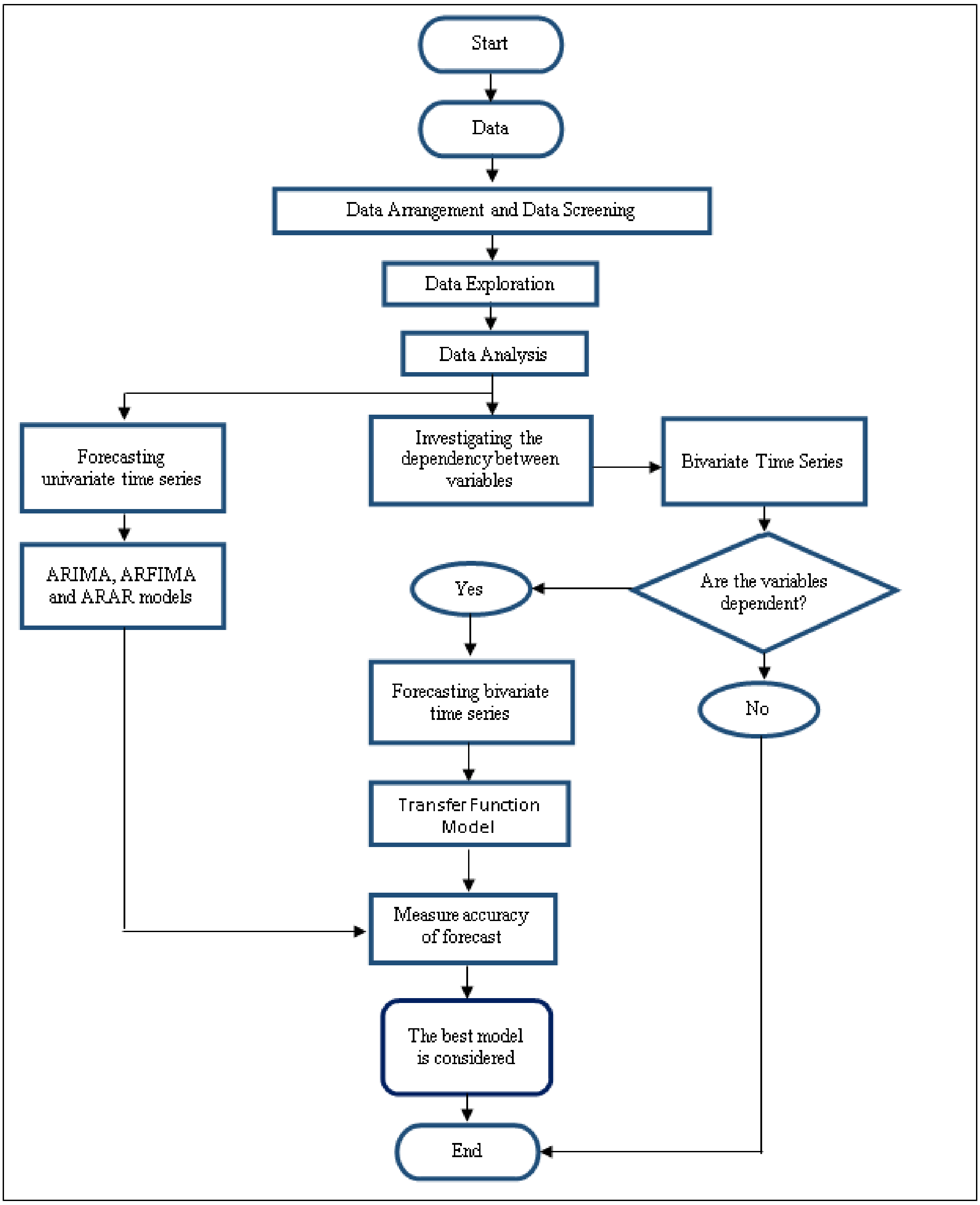


Figure 1. The flowchart of the study.

2.2 Time Series Models

In this section, the univariate time series models used, namely the ARIMA (*p*, *d*, *q*) and ARFIMA (*p*, *d*, *q*) model as well as the ARAR algorithm will be given followed by the discussion of the time series modelling procedure for both the univariate and bivariate time series as well as the transfer function models.

2.2.1 The ARIMA (*p, d, q*) Model

An ARIMA (*p*, *d*, *q*) model is a generalization of an ARMA (*p*, *q*) model. A stationary ARMA (*p*, *q*) model is defined as a sequence of random variables$\{X_t\}$ with p and q (order selection) and is given by Eq. (1).

$$X_t-\phi_1 X_{t-1}-\ldots-\phi_p X_{t-p}=Z_t+\theta_1 Z_{t-1}+\ldots+\theta_q Z_{t-q} \quad (1)$$

where $\{Z_t\}$ is a white noise series with zero mean and constant variance.A process $\{X_t\}$ is called an ARIMA (*p*, *d*, *q*) process if the order of differencing (*d*) is a nonnegative integers such that $(1-B)^d X_t$ is a causal ARMA (*p*, *q*) process with B being the backward shift operator ($B^j X_t = X_{t-j,} B^j Z_t = Z_{t-j,}\ j = 0, \pm 1, \ldots$). The time series $\{X_t\}$ is said to be an autoregressive process of order p (AR(*p*)) if$\theta(B) \equiv 1$, and a moving average process of order q (MA(*q*)) if$\phi(B) \equiv 1$.The ARIMA (*p*, *d*, *q*) processes satisfy the difference equation as in Eq. (2),

$$\phi(B)X_t=\phi(B)(1-B)^d X_t=\theta(B)Z_t\,,\ \{Z_t\} \sim \mathrm{WN}(0,\sigma^2), \quad (2)$$

where $\phi(B)$ and $\theta(B)$ are the p^{th} and q^{th} degree polynomials in Eq. (3) and (4), respectively,

$$\phi(B)=1-\phi_1 B-\ldots-\phi_p B^p \quad (3)$$

and

$$\theta(B)=1+\theta_1 B+\ldots+\theta_q B^q, \quad (4)$$

with$\phi(B) \neq 0$ for$|B| \leq 1$. The polynomial$\phi(B)$has a zero of order $d$at $B = 1$. The process $\{X_t\}$ is stationary if and only if $d = 0$, in which case it reduces to an ARMA (*p*, *q*) process (Brockwell & Davis, 2002).

2.2.2 The Integrated ARFIMA (*p, d, q*) Model

A long memory process or a fractionally integrated ARMA, ARFIMA (*p*, *d*, *q*) processes with $0 < |d| < 0.5$ is a stationary process with much more slowly decreasing autocorrelation function $\rho(k)$ at lag k as $k \to \infty$ which satisfies the property of $\rho(k) \sim Ck^{2d-1}$. The ARFIMA processes satisfy the difference equation given by Eq. (5),

$$(1-B)^d \phi(B)X_t= \theta(B)Z_t \quad (5)$$

where $\{Z_t\}$ is a white noise series with zero mean and constant variance, denoted as $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2), \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$satisfying $\phi(B) \neq 0$ and$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$, satisfying $\theta(B) \neq 0$ for all B such that $|B| \leq 1$, and B is the backward shift operator. The operators$(1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j$ with $n_0 = 1$ and $\pi_j = \prod_{0<k\leq j} \frac{k-1-d}{k}$are for $j = 0,1,2, \ldots$. (Brockwell & Davis, 2002).

2.2.3 The ARAR Algorithm

The time series modelling procedure by the ARAR algorithm used was an adaptation from Parzen's ARARMA algorithm. The ARAR algorithm is a version of which the ARMA fitting step is replaced by the fitting of a subset AR model to the transformed data (Brockwell & Davis, 2002). The ARAR algorithm need not deal with the stationarity of the time series like the ARIMA and ARFIMA models and Eq. (6) gives the ARAR algorithm as,

$$X_t = \phi_1 X_{t-1} + \phi_{l_1} X_{t-l_1} + \phi_{l_2} X_{t-l_2} + \phi_{l_3} X_{t-l_3} + Z_t. \tag{6}$$

where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ for the given lags, l_1, l_2 and l_3 with the coefficients ϕ_j.The value σ^2 are found in Eq. (7) and (8), the Yule-Walker equations

$$\begin{bmatrix} 1 & \hat\rho(l_1-1) & \hat\rho(l_2-1) & \hat\rho(l_3-1) \\ \hat\rho(l_1-1) & 1 & \hat\rho(l_2-l_1) & \hat\rho(l_3-l_1) \\ \hat\rho(l_2-1) & \hat\rho(l_2-l_1) & 1 & \hat\rho(l_3-l_2) \\ \hat\rho(l_3-1) & \hat\rho(l_3-l_1) & \hat\rho(l_3-l_2) & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_{l_1} \\ \phi_{l_2} \\ \phi_{l_3} \end{bmatrix} = \begin{bmatrix} \hat\rho(1) \\ \hat\rho(l_1) \\ \hat\rho(l_2) \\ \hat\rho(l_3) \end{bmatrix} \tag{7}$$

and

$$\sigma^2 = \hat\gamma(0)\left[1 - \phi_1\hat\rho(1) - \phi_{l_1}\hat\rho(l_1) - \phi_{l_2}\hat\rho(l_2) - \phi_{l_3}\hat\rho(l_3)\right], \tag{8}$$

where $\hat\gamma(i)$ and $\hat\rho(i)$, $i = 0, 1, 2, \ldots$, are the sample autocovariances and autocorrelations of the series $\{X_t\}$.

2.3 Time Series Modelling Procedures

All analyses in this study were done using the Interactive Time Series Modelling software, ITSM 2000 version 7.0 (Brockwell & Davis, 2002). For the purpose of time series modelling, the total of 81 sets of observations were divided into two parts where the first 72 sets of observations (from January 2012 to December 2017) were used to fit the time series models while the remaining 9 sets (from January to September 2018) were used to check the accuracy of the forecasted values for each model.The accuracy of the forecasts in this study were evaluated using the mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) as given in Eq. (9), (10) and (11) respectively

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}_i| \tag{9}$$

$$\mathrm{RMSE} = \frac{1}{n}\left(\sum_{i=1}^{n}(x_i - \hat{x}_i)^2\right)^{1/2} \tag{10}$$

$$\mathrm{MAPE} = \frac{1}{n}\left(\sum_{i=1}^{n}\left|\frac{x_i - \hat{x}_i}{x_i}\right|\right) \tag{11}$$

where x_i are the actual observed values and $\hat{x}_i$ are the predicted values while n is the number of predicted values.

2.3.1 Univariate Time Series

Since modelling of ARIMA (*p*, *d*, *q*) and ARFIMA (*p*, *d*, *q*) requires a zero-mean stationary time series with a constant variance, a box-cox transformation was first performed to stabilise the variance before differencing the series to remove the trend and seasonality in it. When the stationary condition was achieved, the mean of the series was subtracted before the model was fitted using the 'auto fit' option in the software where the best model was chosen based on the smallest AICC statistic value. The AICC is a version of Akaike Information Criterion (AIC) that has a correction for small sample sizes with p and q in a specified range. An example of a zero-mean stationary times series plot for the crude palm oil production in Peninsular Malaysia from January 2012 to December 2017 is as displayed in Figure 2. For the ARAR algorithm, a memory shortening filter was applied to the data before fitting a subset autoregressive model with nonzero coefficients at 4 lags by minimising the white noise variance at a maximum of 13 lags. Based on Eq. (7) and (8), lag 13 is chosen for a slower procedure that maximises the Gaussian Likelihood of the observations by ITSM 2000 version 7.0 (Brockwell & Davis, 2002).

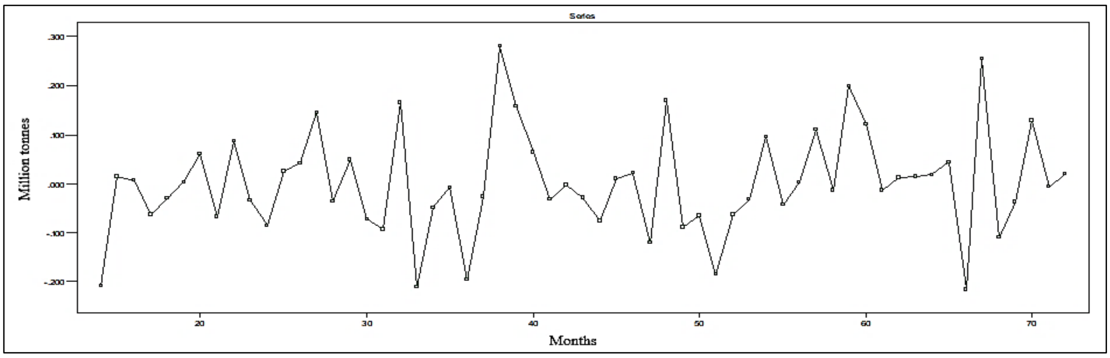


Figure 2. The transformed and differenced time series plot of crude palm oil from January 2012 to December 2017.

2.3.2 Bivariate Time Series

The bivariate time series modelling was used to investigate the dependence of two stationary time series variables like production and price, production and import, production and export, price and import, price and export, and, export and import. When modelling bivariate time series, we began by plotting the series and the sample cross correlation functions of each of the two variables considered. Just like the univariate times series, a stationary bivariate AR model had to be obtained first (through differencing) before further checking on the dependence between the two variables were conducted. Figure 3 and Figure 4 show an example of the stationarised bivariate time series and their sample cross correlation functions respectively.

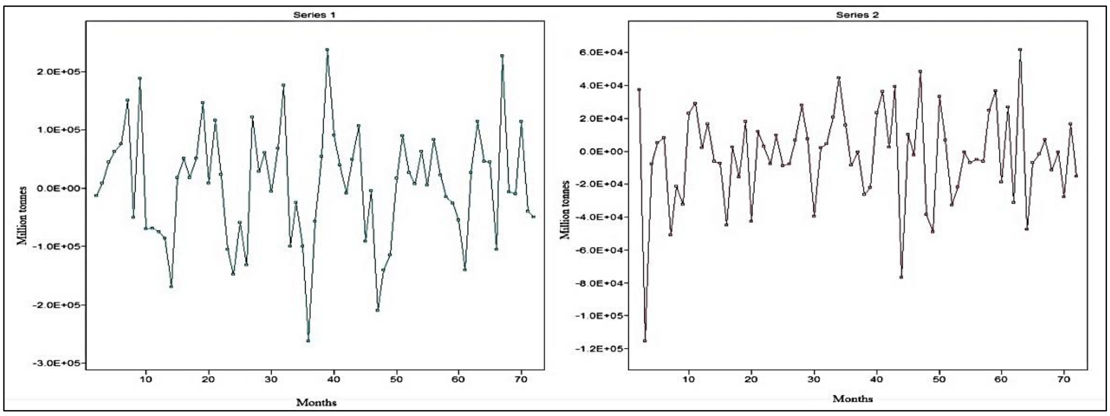


Figure 3. The stationarised plot of the bivariate time series for the production and import of crude palm oil in Peninsular Malaysia.

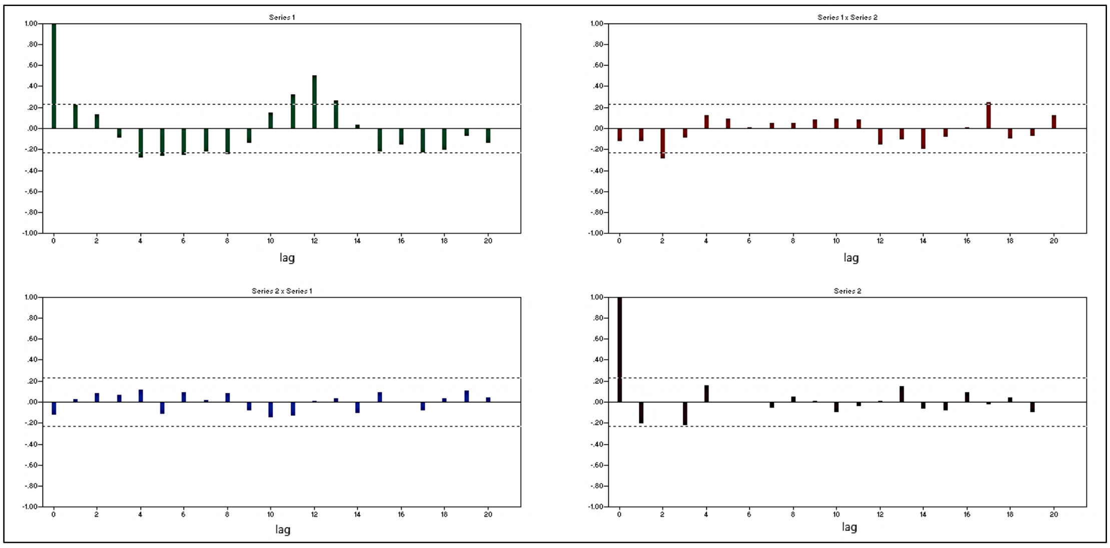


Figure 4. The plot of the sample cross correlation functions for the bivariate time series.

When no further differencing was necessary, the bivariate AR model was fitted using the Yule-Walker Estimation with the smallest AICC value. The fitted model is given by

$$\mathbf{X}_t - \hat{\mathbf{\Phi}}_0 - \hat{\mathbf{\Phi}}_1\mathbf{X}_{t-1} - \cdots - \hat{\mathbf{\Phi}}_1\mathbf{X}_{t-p} = \mathbf{Z}_t$$

where $\{\mathbf{Z}_t\}\sim WN(0,\mathbf{\Sigma}^2)$ with the fitted coefficients $\hat{\mathbf{\Phi}}_j$.If the upper right component of the coefficient matrices is close to zero, $\{X_{t1}\}$can be effectively modelled independently of$\{X_{t2}\}$ and the model provides an adequate fit to the univariate series $\{X_{t1}\}$. If the bottom left component of the coefficient matrices is close to zero, $\{X_{t2}\}$can be effectively modelled independently of $\{X_{t1}\}$ and the model provides an adequate fit to the univariate series $\{X_{t2}\}$. However, if there is a relation between $\{X_{t1}\}$ and $\{X_{t2}\}$, the preliminary model of $\{X_{t2}\}$ which is dependent on $\{X_{t1}\}$can be expressed as

$$X_{t2} = T(B)X_{t1} + N_tW_t, \{W_t\}\sim WN(0,\hat{\sigma}^2)$$

where $T(B) = \frac{\theta(B)}{\phi(B)}$ and $N_t = \frac{1}{\phi(B)}W_t$ with $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$satisfying $\phi(B) \neq 0$ and$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$satisfying $\theta(B) \neq 0$ for all B such that $|B| \leq 1$, and B is the backward shift operator.Besides, the sample cross correlation of the residual vectors $\hat{\mathbf{Z}}_t$ as given in Figure 5 can also be used to determine the dependence between the two variables. If the spikes were nearly all within the $\pm 1.96/\sqrt{n}$ boundaries, then two residual series (and hence the two-original series) are uncorrelated or independent. Thus, the univariate model was a good fit. If the two series between the variables are correlated (dependence), we would proceed to do the prediction by using the transfer function model.

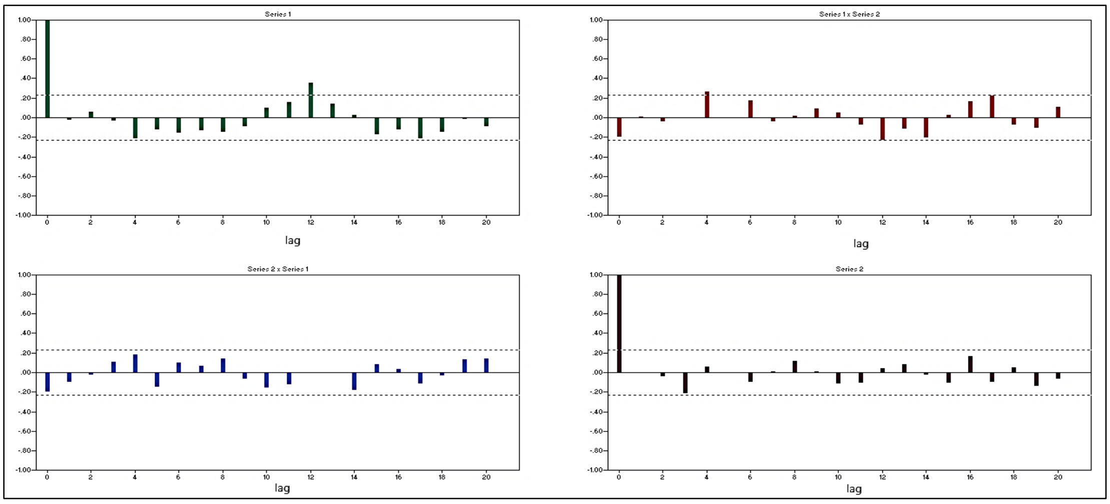


Figure 5. The Plot of the Sample Cross Correlations of the Residual Vectors, $\hat{\mathbf{Z}}_t$.

2.3.3 Transfer Function Model

A transfer function model is a statistical model describing the relationship between an output variable and one or more input variables. This model is one of the popular approaches in the time series modelling for forecasting. When there is an output series which is inclined by an input series, the objective of the transfer function model is to identify the role of input series in determining the variable of interest (Arumugam & Anithakumari, 2013). In this study, the transfer function models were only used to make predictions for bivariate time series data which were dependent where the modelling and forecasting procedures were done using the 'transfer' option in ITSM 2000 software (Brockwell & Davis, 2002).

3. Results

The results in this section will include the monthly forecasts from January to September 2018 of the respective univariate time series models for the production, import, export and price of crude palm oil in Peninsular Malaysia. For each variable considered, only the models that were able to *predict well* will be discussed and presented. A model predicts well when all the actual values of the variable were within the 95% forecast boundaries like the plot displayed in Figure 6. Besides univariate, the forecasts of the bivariate time series using transfer function models will also be included. Finally, a comparison between the performance of the univariate and bivariate time series based on the smallest errors will be presented.

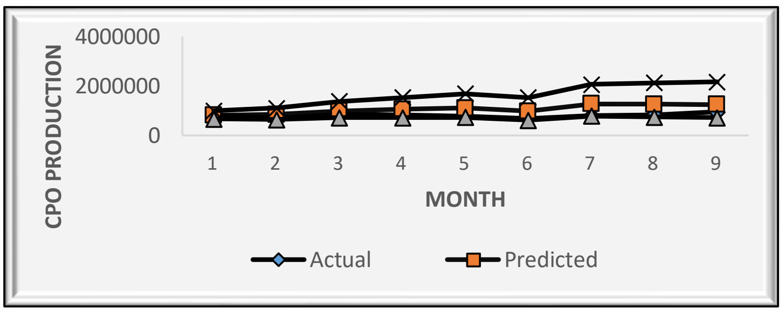


Figure 6. The actual, predicted and forecast boundaries of the monthly crude palm oil production from January to September 2018.

3.1 Univariate Time Series

The ARIMA (*p*, *d*, *q*) and ARFIMA (*p*, *d*, *q*) models as well as the ARAR algorithm were fitted with the monthly production, import, export and price of crude palm oil in Peninsular Malaysia. The models that were able to predict the respective variables well are tabulated in Table 1. The table shows that all the three models considered were good in predicting both the import and price of crude palm oil. However, the ARIMA model and the ARAR algorithm were the only suitable models to predict the production and export of crude palm oil, respectively. Therefore, the post forecast accuracy check is necessary in order to find the best predicting models for import and price of crude palm oil. The values of the MAE, RMSE and MAPE were calculated and are presented in Table 2.

Table 1. The models which predict the respective variables well.

Variable	ARIMA (p, d, q)	ARFIMA (p, d, q)	ARAR algorithm
Production	ARIMA $(1, 1, 0)$	Not available	Not available
Import	ARIMA $(1, 1, 1)$	ARFIMA $(0, -0.2651, 0)$	ARAR
Export	Not available	Not available	ARAR
Price	ARIMA $(0, 1, 0)$	ARFIMA $(0, 0.08903, 0)$	ARAR

Table 2. The MAE, RMSE and MAPE values for the production, import and export and price of crude palm oil.

Variable	Model	MAE	RMSE	MAPE
Production	ARIMA $(1, 1, 0)$	258229.33	293016.94	32.18%
Import	ARIMA $(1, 1, 1)$	37686.89	42562.06	66.28%
	ARFIMA $(0, -0.2651, 0)$	38426.78	43333.98	67.69%
	ARAR	**15939.33**	**19161.84**	**29.69%**
Export	ARAR	196097.11	207664.00	15.19%
Price	ARIMA $(0, 1, 0)$	**85.32**	96.34	**3.71%**
	ARFIMA $(0, 0.08903, 0)$	87.80	**91.81**	3.73%
	ARAR	169.26	188.56	7.34%

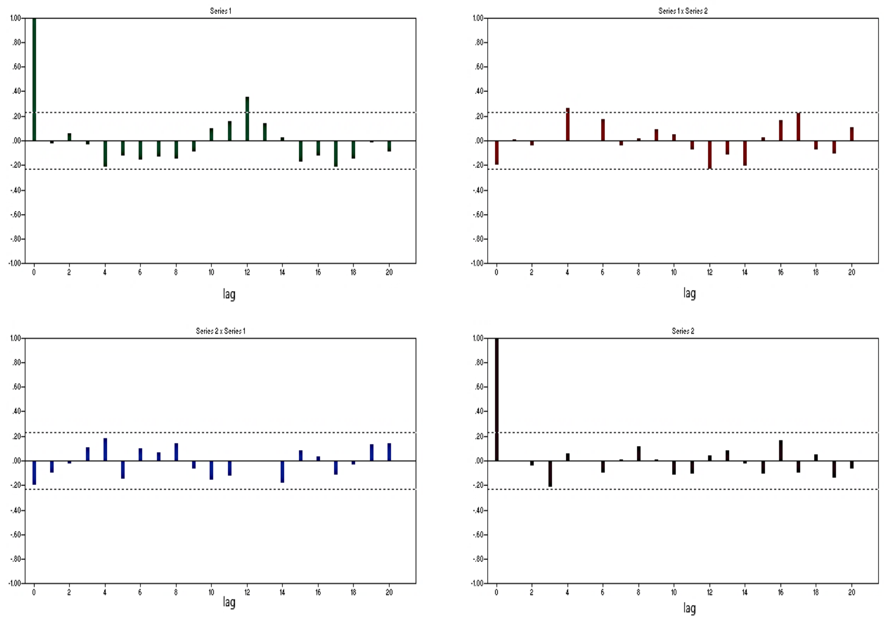
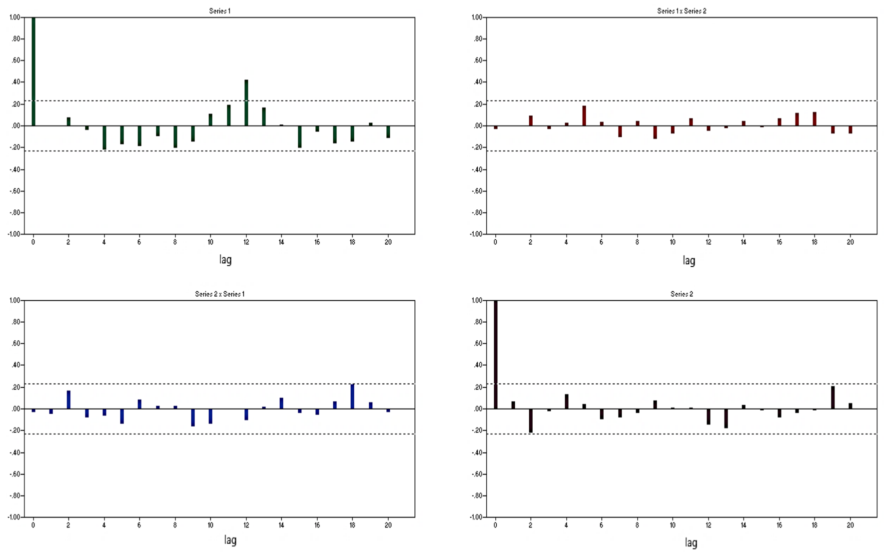
Since the best predicting model is the model that has the lowest MAE, RMSE and MAPE values, it is clear from Table 2 that the best model for predicting the import of crude palm oil is the ARAR algorithm (highlighted in bold). As for the price of crude palm oil, the ARIMA model has the best predicting ability when considering the MAE and MAPE values. If the RMSE is taken as the criterion for model selection, the ARFIMA emerged the best model.

3.2 Bivariate Time Series

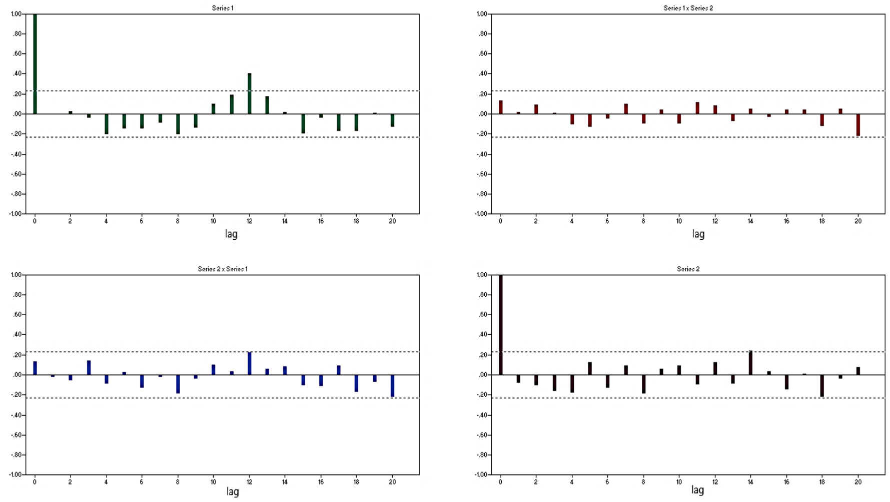
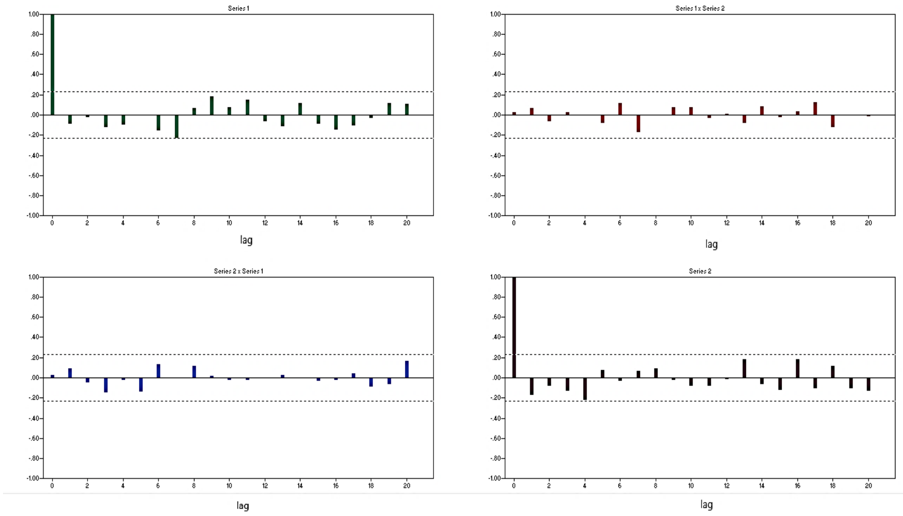
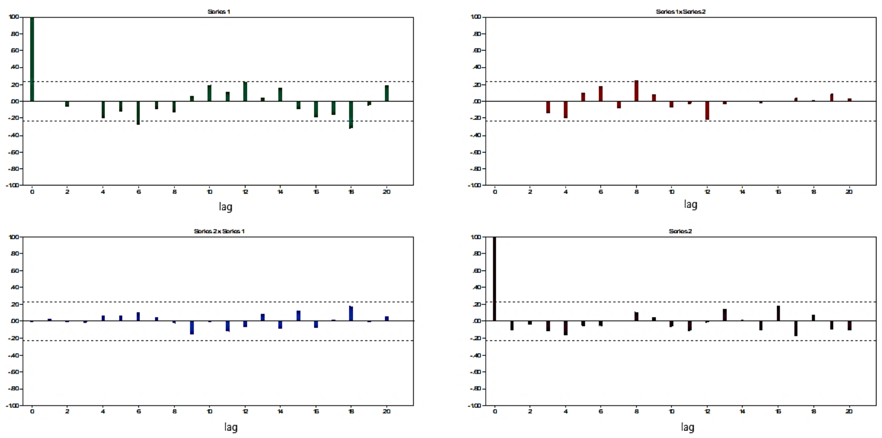
To fit a bivariate time series model, the dependence of two stationary time series had to be investigated first. Hence, the sample cross correlation functions of the residuals were used for this purpose. The sample cross correlation of the residuals is useful in the examination of relationship between the two-time series. If the sample cross correlation of the residuals between two series are within the bounds of $\pm 1.96/\sqrt{n}$, i.e., the two horizontal lines in the plots of Figure 5, then two series are uncorrelated (independent). Thus, the univariate model is a good fit. Otherwise, the two series are said to be dependent. Therefore, only dependent bivariate series will be discussed in this section. The residuals' sample cross correlation functions for the respective bivariate series are displayed in Table 3. Note that $\{X_{t1}\}, \{X_{t2}\}, \{X_{t3}\}$ and

$\{X_{t4}\}$denote the production, import, price and export of crude palm oil, respectively. From the plots in Table 3, it is obvious that there are three sets of dependent variables which are the production and import,$\{X_{t1}, X_{t2}\}$, export and import,$\{X_{t4}, X_{t2}\}$, and export and price,$\{X_{t4}, X_{t3}\}$.The coefficients of the three bivariate series are tabulated in Table 4 while Table 5 shows the bivariate time series models after deleting the values of coefficients which are small and negligible.

Table 3. The residuals sample cross correlation functions for the respective bivariate series

Variables	**Sample Cross Correlation Functions**
$\{X_{t1}, X_{t2}\}$	
$\{X_{t1}, X_{t3}\}$	

(Continuation from Table 3.)

$\{X_{t1}, X_{t4}\}$	
$\{X_{t3}, X_{t2}\}$	
$\{X_{t4}, X_{t2}\}$	

(Continuation from Table 3.)

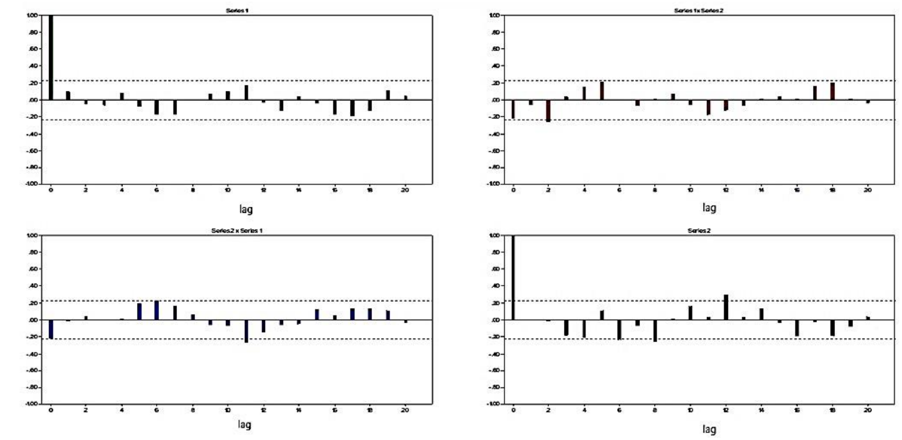
$\{X_{t4}, X_{t3}\}$	

Table 4. The coefficient matrices of the bivariate AR models

	$\{X_{t1},X_{t2}\}$	$\{X_{t4},X_{t2}\}$	$\{X_{t4},X_{t3}\}$
$\hat{\phi}_0$	$\begin{bmatrix}-633.074\\-3573.07\end{bmatrix}$	$\begin{bmatrix}-633.074\\-3573.07\end{bmatrix}$	$\begin{bmatrix}-8.728874\\-2957.01\end{bmatrix}$
$\hat{\phi}_1$	$\begin{bmatrix}0.151030 & -0.565373\\-0.005910 & -0.223255\end{bmatrix}$	$\begin{bmatrix}-0.842516 & -0.542150\\0.003506 & -0.861242\end{bmatrix}$	$\begin{bmatrix}0.213354 & 0.000269\\-329.198 & -0.257180\end{bmatrix}$
$\hat{\phi}_2$	$\begin{bmatrix}0.071951 & -0.991650\\0.030448 & -0.039741\end{bmatrix}$	$\begin{bmatrix}-0.534947 & -0.649138\\0.045337 & -0.621853\end{bmatrix}$	
$\hat{\phi}_3$		$\begin{bmatrix}-0.344548 & -2.236159\\0.026127 & -0.596533\end{bmatrix}$	
$\hat{\phi}_4$		$\begin{bmatrix}-0.367207 & -1.389035\\0\ .003847 & -0.237453\end{bmatrix}$	

Table 5. The bivariate AR models.

Variables	**Bivariate AR Models**
$\{X_{t1},X_{t2}\}$	$X_{t1}=\frac{B(-0.565373-0.991650B)}{(1-0.151030B)}X_{t2}+\frac{W_t}{(1-0.151030B)}$, $\{W_t\}\sim WN(0, 8.05332\times10^9)$ $X_{t2}=-0.2268\ B-0.09604\ B^2-0.2943B^3+Z_t$, $\{Z_t\} \sim WN(0, 7.65058\times10^8)$

(Continuation from Table 5.)

$\{X_{t4},X_{t2}\}$	$X_{t4}=\frac{(-0.542150B-0.649138B^2-2.236159B^3-1.389035B^4)}{(1+0.842516B+0.534947B^2+0.344548B^3+0.367207B^4)}X_{t2}+\frac{W_t}{(1+0.842516B+0.534947B^2+0.344548B^3+0.367207B^4)}$, $\{W_t\}\sim WN(0, 2.41793\times10^{10})$ $X_{t2}=-0.2268\ B-0.09604\ B^2-0.2943B^3+Z_t$, $\{Z_t\}\sim WN(0, 7.65058\times10^8)$
$\{X_{t4},X_{t3}\}$	$X_{t4}=\frac{-329.198B}{(1+0.257180B)}X_{t3}+\frac{W_t}{(1+0.257180B)}$, $\{W_t\}\sim WN(0, 2.08300\times10^{10})$ $X_{t3}=(1+0.1688B)Z_t$, $\{Z_t\}\sim WN(0, 2.32867\times10^4)$

3.3 Transfer Function Model

In order to make predictions for bivariate time series, transfer function models were used. For the bivariate series, $\{X_{t1}, X_{t2}\}$, the output and input series are the $\{X_{t1}\}$ and $\{X_{t2}\}$ respectively. For $\{X_{t4}, X_{t2}\}$ and $\{X_{t4}, X_{t3}\}$, the output variable for both series is export while the input series are import and price, respectively. The least squares models of the transfer function models are given in Table 6. All the three bivariate series were able to predict the respective variables well.

Table 6. The transfer function models.

Variables	Transfer Function Models
$\{X_{t1},X_{t2}\}$	$X_{t1}=\frac{B(-0.6154-0.9917B)}{(1-0.3010B)}X_{t2}+(0.1490B+W_t)$, $\{W_t\}\sim WN(0, 9.05628\times10^9)$ $X_{t2}=-0.2268\ B-0.09604\ B^2-0.2943B^3+Z_t$, $\{Z_t\}\sim WN(0, 7.65058\times10^8)$
$\{X_{t4},X_{t2}\}$	$X_{t4}=\frac{(-0.4422B-0.2491B^2-2.136B^3-1.189B^4)}{(1+0.4925B+0.4850B^2+0.3446B^3+0.03279B^4)}X_{t2}+(-0.1925B+0.1150B^2-0.04455B^3-0.3672B^4+W_t)$, $\{W_t\}\sim WN(0, 2.08242\times10^{10})$ $X_{t2}=-0.2268\ B-0.09604\ B^2-0.2943B^3+Z_t$, $\{Z_t\}\sim WN(0, 7.65058\times10^8)$

(Continuation from Table 6.)

$\{X_{t4},X_{t3}\}$	$X_{t4}=\frac{-337.1B}{(1+0.2572B)}X_{t3}+(-0.2428B+W_t)$, $\{W_t\}\sim WN(0, 2.22423\times10^{10})$ $X_{t3}=(1+0.1688B)Z_t$, $\{Z_t\} \sim WN(0, 2.32867\times10^4)$

3.4 Comparison between Univariate and Bivariate Models

The comparison of the forecast accuracy between the univariate and bivariate models will be evaluated based on the MAE, RMSE and MAPE values. From the results in Section 3.3, the variables involved in the comparison here are the production and export of crude oil. The MAE, RMSE and MAPE value for both the univariate and bivariate models are as tabulated in Tables 7 and 8 respectively. The values in Tables 7 and 8 show that when import is the input variable for production, the forecast errors reduced compared to forecasting using the univariate time series. As for the export of crude palm oil, the bivariate model can predict better than the univariate model when the input variable is import instead of price.

Table 7. The MAE, RMSE and MAPE values for the Univariate Models

Variable	MAE	RMSE	MAPE
Production	258229.33	293016.94	32.18%
Export	196097.11	207664.00	15.19%

Table 8. The MAE, RMSE and MAPE values for the Bivariate Models

Output Variable	Input Variable	MAE	RMSE	MAPE
Production	Import	179386.67	195175.42	23.28%
Export	Import	180552.67	200312.76	14.26%
Export	Price	196619.11	237607.62	16.16%

4. Conclusions

The focus of this study is to model the crude palm oil production, price, import and export in Peninsular Malaysia using time series models. Therefore, in the first part of our study, we considered fitting univariate time series models using the ARIMA, ARFIMA and ARAR algorithm. It was found that the best model in predicting the production was ARIMA (1,1,0) while the ARAR algorithm were the best forecast models for both the import and export. As for the price, ARIMA (0,1,0) appeared to be the best forecast model based on the MAE and MAPE values while ARFIMA $(0, 0.08903, 0)$ is considered the best model based on RMSE value. It is obvious that the selection criterion used in model selection is important when choosing the best forecast model. Subsequently, the dependence between two variables were investigated before modelling the bivariate time series. For this purpose, the residuals' sample cross correlation functions were used and the results showed that the production is dependent on import while the export is dependent on either price or import. In order to model the bivariate time series and make predictions, the transfer function models were used. Finally, the forecast ability between the univariate and bivariate models based on the MAE, RMSE and MAPE values was carried out. It was found that the bivariate models generally had better performance compared to the univariate models. In this study, the variables considered were related to crude palm oil in Peninsular Malaysia. In the calculation of errors, the large MAE and RMSE values could be due to the small sample sizes used. Therefore, we suggest that larger samples sizes be considered in future. It is also hoped that this analysis be applied to other variables in the field of agriculture as well as other related fields. Since the models considered were some univariate and bivariate time series, other types of time series models like spectral analysis and multiple time series models can be considered for future studies. Artificial neural network (ANN) can be further explored too.

References

Abdullah, R. & Wahid, M. B. (2011). World Palm Oil Supply, Demand, Price and Prospects: Focus on Malaysian and Indonesian Palm Oil Industry. Malaysian Palm Oil Board (MPOB). *Oil Palm Industry Economic Journal* 11(2): 13-25

Andarani, P., Nugraha, W. D., Sawitri, D. &Budiawan, W. (2018). Life-Cycle Assessment of Crude Palm Oil Produced at Mill J, PT XYZ, Sumatera Island using Eco-indicator 99. Retrieved September 14, 2018 fromhttps://www.matec-conferences.org/articles/matecconf/pdf/2018/18/matecconf_ijcaet-isampe2018_01028.pdf

Arumugam, P. & Anithakumari, V. (2013). Seasonal Time Series and Transfer Function Modelling for Natural Rubber Forecasting in India. *International Journal of Computer Trends and Technology* (*IJCTT*), 4 (5): 1366-1370.

Asari, F. F. A. H., Rahman, N. H. A., Razak, E. A. A., Ahmad, B. A. S., Harun, N. F. A. & Jusoff, K. (2011). A Time Series Analysis of the Relationship Between Total Area Planted, Palm Oil Price and Production of Malaysian Palm Oil.*World Applied Sciences Journal*, 12, 34-40.

Brockwell, P. J., & Davis, R. A. (2002). Introduction to Time Series and Forecasting (2nd ed.). New York: Springer-Verlag.

Farrelly, C. (2017). What is the importance and usage of time series. Retrieved Jun 21, 2017, from https://www.quora.com/What-is-the-importance-and-usage-of-Time-Series

Lazim, M. A. (2016). Introductory Business Forecasting-a practical approach (3rd ed.). Shah Alam, Selangor: UiTM Press.

Lee, Y. N. (2011). Increase Malaysia Palm Oil Production Efficiency. Retrieved September 24, 2018 fromhttp://bora.uib.no/bitstream/handle/1956/6151/91529167.pdf;sequence=1

Mah P.J.W., BuharyF. N., AbdullahN. H. & Saad S. A. M. (2018). A Comparative Study Of Univariate Time Series Modelling For Natural Rubber Production In Malaysia. *Malaysian Journal of Computing*, 3 (2): 108–118.

Mah P.J.W., Zali N.N.M., Ihwal N.A.M.& Azizan N.Z. (2018). Forecasting Fresh Water And Marine Fish Production In Malaysia Using ARIMA And ARFIMA Models. *Malaysian Journal of Computing*, 3 (2): 81–92.

Malaysian Palm Oil Board (MPOB). (n.d). Economic and Industry Development Division. Retrieved September 14, 2018 fromhttp://bepi.mpob.gov.my/index.php/en/

Malaysian Palm Oil Council (MPOC). (n.d). Retrieved September 14, 2018 from http://www.mpoc.org.my/default.aspx

New Strait Times. (2018). Malaysia's Palm Oil Production to Drop in 2018 on Lower Yields. Retrieved October 19, 2018 from https://www.nst.com.my/business/2018/08/405827/malaysias-palm-oil-production-drop-2018-lower-yields

Nikkei Market. (2019). Palm Oil Prices Likely to Rise in 2019 as Output Slows. Retrieved July 5, 2019 fromhttps://asia.nikkei.com/Business/Markets/Nikkei-Markets/Palm-oil-prices-likely-to-rise-in-2019-as-output-slows

Norhidayu, A., Syazwani, N. M., Radzil, R. H. M., Amin, I. Z. M. & Balu, N. (2017). The Production of Crude Palm Oil in *Malaysia. Int. Journal of Economics and Management*, 11 (53), 591-606.

Otieno, N. E., Dai, X., Barba, D. D., Bahman, A., Smedbol, E., Rajeb, M., & Jaton, L. (2016). Palm Oil Production in Malaysia: An Analytical Systems Model for Balancing Economic Prosperity, Forest Conservation and Social Welfare. *Agricultural Sciences*, 7, 55-69. doi: 10.4236/as.2016.72006.

Sawe, B. E. (2018). Top Palm Oil Producing Countries In The World. Retrieved August 19, 2019 from https://www.worldatlas.com/articles/top-palm-oil-producing-countries-in-the-world.html

Sembiring, M. T., Wahyuni, D., Sinaga, T. S. & Silaban, A. (2018). Study of Activity based Costing Implementation for Palm Oil Production using Value-Added and Non-Value-Added Activity Consideration in PT XYZ Palm Oil Mill. Materials Science and Engineering, 309, 012059, doi:10.1088/1757-899X/309/1/012059.

Subramaniam, V., Choo, Y. M., Muhamad, H., Hashim, Z., Yew, A.T. & Puah, C.W. (2010). Life-Cycle Assessment of the Production of Crude Palm Oil (Part 3). *Journal of Oil Palm Research*. Vol. 22, p. 895-903.

The Edge Financial Daily. (2017). Sabah Planters will continue to feel the Heat from El Nino. Retrieved September 24, 2018 fromhttp://www.theedgemarkets.com/article/sabah-planters-will-continue-feel-heat-el-nino

The Star Online. (2018). Better Outlook for Planters. Retrieved July 8, 2019 fromhttps://www.thestar.com.my/business/business-news/2018/09/11/better-outlook-for-planters/

Vijaya, S., Ma, A. N., Choo, Y. M. & Sulaiman, N. M. N. (2008). Life-Cycle Inventory of the Production of Crude Palm Oil- A Gate Case Study of 12 Palm Oil Mills. *Journal of Oil Palm Research*. Vol. 20, June 2008, p. 484-494.