Palm Oil to Biodiesel: Determination Factors on Production and Export Volume

Noraini Abdullah, H. J. Zainodin, and Haider F. Abdul Amir

Abstract—In this study, determination factors that affect the production volume of oil palm fruits and the export volume of biodiesel are identified using the multiple regression technique, and results using model transformation method are compared. For this purpose, two models are developed, namely, Model I is for the production volume of oil palm fruits, while Model II is for the export volume of biodiesel. Data variables are transformed using the ladder of power transformation method. There are five independent variables in Model I and four independent variables in Model II. The four phases in multiple regression model-building are carried out on both Model I and II respectively. Model I will use the model transformation method to change the data from non-normality to normality. The best model obtained for Model I by model transformation is M72.2.5. The main factor is the total workers employed during last pay period, and the interaction factors up to order two are: harvested area interact with yield per hectare, harvested area interact with local delivery average price, harvested area interact with total workers employed during last pay period, yield per hectare interact with local delivery average price, harvested area interact with local delivery average price interact with total workers employed during last pay period and yield per hectare interact with local delivery average price interact with total workers employed during last pay period. The MAPE value for best model on model transformation is 2.62%. Therefore, the best model using the model transformation method is said to be acceptable to forecast for the production volume. The best model M72.2.5 is then used to obtain Model II which is given by M2.0.0. The main factor is the export volume of biodiesel.

Index Terms—Determination factors, multiple regression, model transformation, MAPE, production volume, export volume.

I. INTRODUCTION

Oil palm is one of the most important Malaysia's export commodities and it has been contributing to the Malaysia's economy income from the nineteenth century till present. Palm oil can be used to produce biodiesel which is biodegradable and non-toxic fuel. This is a new market for palm oil which has the potential to dramatically increase the global demand for this commodity. Biodiesel has become

Manuscript received March 13, 2015; revised June 10, 2015.

more attractive recently because of its environmental benefits and the fact that it is made from renewable resources. Biodiesel production is the process of producing the biofuel, biodiesel, through the chemical reactions transesterification and esterification. This involves vegetable or animal fats and oils being reacted with short-chain alcohols. According to [1], in the beginning of 2004, the increase in the petroleum prices have provided the motive force for turning vegetable oils into biodiesel as transportation fuel. It converts the cheapest palm oil, rapeseed oil or soybean oil into biodiesel and sells these oils at the price of petroleum diesel to earn a gross margin.

The production of oil palm fruits in the international market is increasing from time to time. There are many factors at present that are able to influence the production of oil palm fruits. This research will discuss about the relationship of the planted area, harvested area, yield per hectare of oil palm, local delivery average price and the total number of workers employed during the last pay period of the oil palm with the production of the oil palm. An increase in the planted area is expected to increase the production of oil palm fruits. Even though the expansion of the palm oil plantation helps to increase the total palm oil production, it does not reflect the technological and scientific advances made [2].

The production will keep increasing in the future due to the increase in the planted areas that have entered the maturity stage. Oil palm fruits when reaches their maturity stage would depend on many factors, which are technological problems, geology problems and also environmental problems. Yield of oil palm fruits is normally assumed to depend on the technological advancement of its production. Besides that, the discovery of new high yielding varieties, improvements in planting techniques and the increasing use of fertilizers and chemicals will hence increase the yield. Supposedly when the oil palm production increases, the palm oil price will also be reduced. The price of palm oil started to reduce in 2005 due to an increase in its production and a reduction in the price of its substitute goods; namely soya bean oil. Hence, there is negative relationship between the price of palm oil and its production in Malaysia [2]. Increase in the production of oil palm will cause the total number of workers employed to be increased. Therefore, the total number of workers employed is also one of the factors that can influence the production of oil palm. If there is a lack of workers in the industry, not just the process of harvesting and processing the fruits, even the process of extracting and generating the oil will also become inefficient.

In this study, multiple regressions are used to model the production volume of oil palm fruits and the export volume of the biodiesel. The processes of trans-esterification and esterification from palm oil into biodiesel are not dealt with in

Noraini Abdullah is with the Mathematics with Economics Program, Faculty of Science and Natural Resources of University Malaysia Sabah, Kota Kinabalu, Sabah 88400, Malaysia (e-mail: norainiabdullah.ums@gmail.com).

H. J. Zainodin is with the Mathematical and Statistical Applications (MASA) Research Group of University Malaysia Sabah, Kota Kinabalu, Sabah 884000, Malaysia (e-mail: zainodin@gmail.com).

Haider F. Abdul Amir is with the Physics with Electronics Program, Faculty of Science and Natural Resources of University Malaysia Sabah, Kota Kinabalu, Sabah 88400, Malaysia (e-mail: haider@.ums.edu.my).

this paper. Model Transformation method will be used to change the non-normality data into normality for production model of oil palm fruits in Model I. The best model can then be used in Model II for forecasting the export volume according to the Mean Absolute Percentage Error (MAPE).

II. MATERIALS AND METHODS

Study area: This research was conducted based on the production volume of oil palm fruits and export volume of biodiesel in Malaysia. The data of the oil palm includes the production of oil palm fruits (*P*1), planted area (A), harvested area (H), yield per hectare (Yd), local delivery average price (Ap) and total workers employed during the last pay period (Tw) (viz. the day you get your exact amount pay), whereas the data on biodiesel would include the export volume of biodiesel (EVol), production volume of oil palm (P1), export value of biodiesel (EVal), gross domestic product (GDP) and gross national income (GNI).

Data collection: A total of 31 data of oil palm from 1982 till 2011 were collected from [3], and through the internet official website of [4]. In this research, 26 data sets would be used for modelling while only 5 data sets selected at random would be used later for validation of the model. The data on oil palm would made up the five variables for Model I which would thus include the production of oil palm fruits, planted area, harvested area, yield per hectare, local delivery average price and total workers employed during the last pay period (the day you get your exact amount of pay), whereas the data on biodiesel volume would made up the four variables for Model II which would include the production volume of oil palm from Model I, the export volume and value of biodiesel obtained from the official website of Malaysian Palm Oil Board from 2007 to 2011. Other data two would include the gross domestic product and gross national income, taken from the [4] from 2007 till 2011.

III. METHODOLOGY

In this research, the hierarchical multiple regression technique is used to relate a dependent variable Y, to several independent variables $W_1, W_2, ..., W_k$. The regression model equation can be written in this form [5]:

$$Y = \Omega_0 + \Omega_1 W_1 + \Omega_2 W_2 + \dots + \Omega_{k-1} W_{k-1} + \Omega_k W_k + u$$
(1)

With Ω_0 denotes as the constant term, Ω_j is the *j*-th coefficient of the independent variable, W_j for (j=1, 2, ..., k), *k* is the number of independent variables, '*k*+1' is the number of parameters, and *u* denotes the residuals of the regression model equation. In this research, the independent variables would take in the form of transformed variables using ladder transformation of non-normal variables to normality. Basic assumptions of the multiple regression models can be referred to as in [5].

Normality Test: In statistics, normality tests are used to determine if a data set is well-modeled by a normal distribution, and to compute how likely it is for a random variable underlying the data set to be normally distributed

visually supported by the frequency distribution (or histogram) and the Q-Q plot (quantile-quantile plot). Both these methods were supporting evidences, while the numerical statistical tests used, depending on the number of observations (*n*) were the Kolmogorov-Smirnov (for n>50) and Shapiro-Wilk (for n<50). If the significance level is larger than 0.05, then the alternative hypothesis is rejected and therefore normality is assumed. On the other hand, if the significance level is less than 0.05, then the null hypothesis is rejected and thus normality is not assumed.

Model Transformation: In this research, data transformation is needed on the non-normal data so as to more closely approximate a theoretical distribution with statistical properties. Transforming the data can spread the data more evenly and hence, make the distributions more symmetrical. Besides, the relationships between variables are more linear and the data have constant variance.

According to [6], a power transformation is a type of transformation that statisticians have found useful for straightening a plot. A power is first selected, and each original value is raised to that power to obtain the corresponding transformed value. The power of 1 corresponds to no transformation, while the power of 0 will transform every value to 1, which is certainly not informative. Therefore, statisticians use the logarithmic transformation in its place in the ladder of transformations. Table I displays a "ladder" of the most commonly used power transformations as used in this research.

Ladder-Power	Transformed Value	Name
3	(Original value) ³	Cube
2	(Original value) ²	Square
1	Original value	No transformation
1/2	(Original value) ^{1/2}	Square root
$\frac{1}{3}$	(Original value) ^{1/3}	Cube root
0	log(Original value)	Logarithm
- 1/2	$-\frac{1}{(\text{Original value})^{1/2}}$	Reciprocal root
-1	$-\frac{1}{(\text{Original value})}$	Reciprocal
-2	$-\frac{1}{(\text{Original value})^2}$	Reciprocal square

TABLE I: COMMONLY USED TRANSFORMATIONS

Fig. 1 shows the four phase model-building summaries carried out in this research as in [7].



Fig. 1. Model building summaries (Source: [7]).

According to [8]-[10], the number of possible models of Phase 1 is calculated as shown by the formula:

$$N = \sum_{i=1}^{q} j({}^{q}C_{j}) \tag{2}$$

In equation (2), *N* is the number of all possible models, and '*q*' is the number of single quantitative vaiables. The interactions of the single independent variables (X_1X_2 — first order; $X_1X_2X_3$ — second order; $X_1X_2X_3X_4$ — third order) denote the product of these variables, and can be rewritten as X_{12} , X_{123} and X_{1234} respectively [5]. All the possible models would undergo through the Global test as shown by [11]. The selected models of Phase 2 will then undergo two tests, namely, the correlation based multicollinearity test and the coefficient test at $\alpha = 5\%$ significant level. Each variable removal due to multicollinearity, and elimination of insignificant variables from the model using the coefficient test would be denoted in the model labelling [12] as shown in Fig. 2.



Fig. 2. Model labelling in multiple regression.

From the selected models, the best model of Phase 3 will be chosen with the minimum values based on the eight selection criteria (8SC) as shown by [8], [10]. The best model will then undergo through the goodness-of-fit tests in Phase 4, namely, the randomness test using the scatter plot, and the normality test on the standardized residuals. These model-building processes are performed on all the models in this research.

The mean absolute percentage error (MAPE) is commonly used in quantitative forecasting methods as it produces a measure of the relative overall fit. The absolute values of all the percentage errors are summed up and the average is calculated. The lower MAPE value, the best the model can be. It can then be used in forecasting or evaluating the missing values. The MAPE of the best model is thus calculated so that its efficiency in forecasting the future export volume of biodiesel can therefore be made. The acceptable range for the MAPE test is less than 25% [11], where the smaller the value of MAPE would a better prediction. The MAPE formula is given as in [11]:

MAPE=
$$\frac{1}{m} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \times 100\%$$
 (3)

where A_t is the actual value, F_t is the forecasted value, and 'm' is the number of randomly selected observations reserved for verification of the model.

IV. RESULTS AND DISCUSSIONS

Using equation (2), Table II shows the number of all

possible models for Model I (80 models) with its corresponding number of variables. Model I contains one dependent variable (P1) and five independent variables (A, H, Yd, Ap and Tw).

TABLE II: ALL POSSIBLE MODELS WITH FIVE INDEPENDENT VARIABLES (MODEL D)

Number of	Single	Ord	Total					
variables	Single	1st	2nd	3rd	4th	Total		
1	5	-	-	-	-	5		
2	10	10	-	-	-	20		
3	10	10	10	-	-	30		
4	5	5	5	5	-	20		
5	1	1	1	1	1	5		
Total	31	26	16	6	1	80		

Table III shows the normality tests (Shapiro-Wilk: df=26) on the variables before and after transformation using the ladder-power transformation method. It can be seen that all the non-normal variables (*Yd*, *Ap* and *Tw*) have turned to normal after the appropriate transformation with all their *p* values > 0.05 (Table III) and supported by the normality plots as depicted in Fig. 3.

TABLE III: NORMALITY TESTS BEFORE AND AFTER TRANSFORMATION (MODEL I)

	Bef	ore Trans	formation	After Transformation		
Model 1	Statistics	P-value			Statistics	P-value
<i>P</i> 1	0.948	0.210	Normal	<i>Y</i> = <i>P</i> 1	0.210	0.210
Α	0.936	0.105	Normal	$X_1=A$	0.105	0.105
Η	0.941	0.145	Normal	$X_2=H$	0.145	0.145
Yd	0.898	0.014	Not normal	$X_3 = (Yd)^3$	0.934	0.094
Ap	0.878	0.005	Not normal	$X_4 = (Ap)^{-1/2}$	0.974	0.716
Tw	0.879	0.005	Not normal	$X_5 = (Tw)^{-1/2}$	0.943	0.155

TABLE IV: ALL POSSIBLE MODELS WITH FOUR INDEPENDENT VARIABLES (MODEL II)

(
Number of	Single	Order of	Total					
variables	Single	1 st	2 nd	3 rd	Total			
1	4				4			
2	6	6			12			
3	4	4	4		12			
4	1	1	1	1	4			
Total	15	11	5	1	32			

Similarly, using equation (XX), Table IV shows the number of all possible models for Model II (32 models) with its corresponding number of variables.

Table IV shows the number of possible models for Model II which contains one dependent variable (*EVol*) and four independent variables (*P*1, *EVal*, GDP and GNI). Table V shows the normality tests (Shapiro-Wilk: df=5) on these variables. Thus, no transformation is needed on the variables for Model II.



Fig. 3. Normality plots before and after transformations.

	Shapiro-Wilk ($Df=5$)							
Model II	Statistic	P-value	Normality					
EVol	0.917	0.509	Normal					
P1	0.964	0.839	Normal					
EVal	0.797	0.077	Normal					
GDP	0.985	0.959	Normal					
GNI	0.980	0.936	Normal					

Phase 2 of model-building is then carried out on the possible models of the transformed variables. The multicollinearity test is carried out on all the models as illustrated by [7], and followed by the coefficient test as shown by [9]. This has resulted in 34 selected models being obtained for Model I. Table VI below shows some selected models with their eight selection criteria values. It can be seen from Table VI that the best model which has the minimum

values of all the eight selection criteria (8SC) is given by model M72.2.5. Model M72.2.5 would then undergo finally the goodness-of-fit tests for validation to be used for forecasting.

Model M72 is given as in equation (1) by equation (3) below:

$$M72.0.0: Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{25} X_{25} + \beta_{34} X_{34} + \beta_{35} X_{35} + \beta_{45} X_{45} + \beta_{234} X_{234} + \beta_{235} X_{235} + \beta_{245} X_{245} + \beta_{345} X_{345} + u$$
(3)

After the execution of the multicollinearity and coefficient tests with the removal of variables (X_2 , X_3 , X_4 , X_{35} , X_{45} , X_{234} , and X_{235}), model equation of M72.2.5 is thus obtained as in (4):

M72.2.5:
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_5 X_5 + \hat{\beta}_{23} X_{23} + \hat{\beta}_{24} X_{24} + \hat{\beta}_{25} X_{25} + \hat{\beta}_{34} X_{34} + \hat{\beta}_{245} X_{245} + \hat{\beta}_{345} X_{345}$$
 (4)

TABLE VI: BEST MODEL USING 8SC

Selected	Selected							
model	AIC	FPE	GCV	НQ	RICE	SCHWARZ	SGMASQ	SHIBATA
	5.366×	5.367×	5.399×	5.517×	5.437×		4.984×	5.308×
M1.0.0	1012	1012	1012	1012	1012	5.911×10 ¹²	1012	1012
:	:	:	:	:	:	:	:	:
	1.174×	1.176×	1.191×	1.225×	1.212×		1.054×	1.148×
M14.0.0	1013	1013	1013	1013	1013	1.358×1018	1013	1013
	1.110×	1.111×	1.126×	1.157×	1.146×		9.962×	1.085×
M15.0.0	1013	1013	1013	1013	1013	1.283×10 ¹⁸	1012	1013
:	:	:	:	:	:	:	:	:
	2.540×	2.546×	2.608×	2.685×	2.697×		2.206×	2.441×
M42.3.0	10 ¹²	1012	1012	1012	1012	3.082×10 ¹²	1012	10 ¹²
	1.087×	1.093×	1.134×	1.166×	1.203×		9.163×	1.025×
M45.1.1	1012	1012	1012	1012	1012	1.385×10 ¹²	1011	1012
:	:	:	:	:	:	:	:	:
	4.933×	5.083×	5.775×	5.592×	8.023×		3.776×	4.178×
M55.1.1	1011	1011	1011	1011	1011	7.626×10 ¹¹	1011	1011
	3.617×	3.692×	4.078×	4.043×	5.082×		2.823×	3.158×
M56.1.2	1010	1010	1010	1010	1010	5.327×1010	1010	1010
:	:	:	:	:	:	:	:	:
	4.738×	4.837×	5.342×	5.297×	6.658×		3.699×	4.136×
M71.2.5	1011	1011	1011	1011	1011	6.978×10 ¹¹	1011	1011
	1.418×	1.447×	1.599×	1.585×	1.992×		1.107×	1.238×
M72.2.5	1010	1010	1010	1010	1010	2.088×10 ¹⁰	1010	1010
	4.471×	4.732×	5.763×	5.212×	1.247×		3.325×	3.541×
M73.13.2	1010	1010	1010	1010	1011	7.613×10 ¹⁰	1010	1010
	2.771×	2.794×	2.951×	3.012×	3.243×		2.270×	2.552×
M74.9.1	10**	10**	10**	10**	10**	3.704×1011	10**	10**
	1.889×	1.969×	2.311×	2.171×	3.793×		1.422×	1.548×
M78.2.4	1010	1010	1010	1010	1010	3.064×1010	1010	1010
	1.418×	1.447×	1.599×	1.585×	1.992×		1.107×	1.238×
MINIMUM	1010	1010	1010	1010	1010	2.088×1010	1010	1010

 TABLE VII: COEFFICIENT TABLE FOR MODEL M72.2.5

	Coefficients	Standard Error	t Stat	P-value
Intercept	3.158×10 ⁷	1.061×10 ⁶	29.76338	9.220×10 ⁻¹⁷
X5	-9.204× 10 ⁹	3.107×10 ⁸	-29.62371	1.002×10 ⁻¹⁶
X ₂₃	1.671×10 ⁻³	1.985×10 ⁻⁵	84.17172	8.004×10 ⁻²⁵
X ₂₄	2.737×10 ²	9.827×10 ⁰	27.84922	2.978×10 ⁻¹⁶
X ₂₅	1.290×10 ³	1.289×10 ²	10.01133	8.777×10 ⁻⁹
X34	-1.648× 10 ⁵	4.646×10^{3}	-35.48400	4.097×10 ⁻¹⁸
X245	-4.895× 10 ⁴	5.214×10 ³	-9.38858	2.337×10 ⁻⁸
X345	4.835×10^{7}	1.444×10 ⁶	33.47994	1.149×10 ⁻¹⁷



Table VII shows the coefficient table of the best model M72.2.5 after the coefficient test has been carried out. It can be seen that the remaining factors have all the *p*-values less than 0.05, indicating that they are significant.

TABLE VIII: PEARSON CORRELATION MATRIX FOR MODEL M72.2.	.5
---	----

	Y	X5	X ₂₃	X ₂₄	X ₂₅	X ₃₄	X ₂₄₅	X345
Y	1.0000	-0.9801	0.9853	0.9277	0.8968	-0.2883	0.6067	-0.7165
X5	-0.9801	1.0000	-0.9485	-0.8729	-0.8439	0.3851	-0.4941	0.7805
X ₂₃	0.9853	-0.9485	1.0000	0.9120	0.8581	-0.1686	0.5987	-0.6261
X ₂₄	0.9277	-0.8729	0.9120	1.0000	0.9353	-0.1002	0.8347	-0.5571
X ₂₅	0.8968	-0.8439	0.8581	0.9353	1.0000	-0.2493	0.8333	-0.6328
X ₃₄	-0.2883	0.3851	-0.1686	-0.1002	-0.2493	1.0000	0.1752	0.8659
X ₂₄₅	0.6067	-0.4941	0.5987	0.8347	0.8333	0.1752	1.0000	-0.1904
X345	-0.7165	0.7805	-0.6261	-0.5571	-0.6328	0.8659	-0.1904	1.0000

Table VIII shows the correlation matrix for best model M72.2.5 after the multicollinearity test and the coefficient test have been carried out. The table shows the correlation between each significant determining factor (X_5 , X_{23} , X_{24} , X_{25} , X_{34} , X_{245} and X_{345}) with the production of oil palm fruits, *Y*. It can be seen that the correlation between X_5 and *Y* is 0.9801, between X_{23} and *Y* is 0.9853, between X_{24} and *Y* is 0.9277, between X_{25} and *Y* is 0.8968, between X_{34} and *Y* is 0.2883, between X_{245} and *Y* is 0.6067, between X_{345} and *Y* is 0.7165. The factor which is the most highly correlated with the production of oil palm fruits is X_{23} and the factor which is the X_{34} .

Referring to the coefficient table in Table VII, by putting the values of the coefficients into the best model of M72.2.5 regression equation as in (4), equation (5) is thus obtained:

M72.2.5 :
$$\hat{Y} = 3.158 \times 10^7 - 9.204 \times 10^9 X_5 + 1.671 \times 10^{-3} X_{23} + 2.737 \times 10^2 X_{24} + 1.290 \times 10^3 X_{25} - 1.648 \times 10^5 X_{34} - (5)$$

4.895 × 10⁴ X₂₄₅ + 4.835 × 10⁷ X₃₄₅



Fig. 4. The goodness-of-fit tests for M72.2.5.

The Randomness test of Phase 4 can also be tested through the scatter plot as a supporting evidence. Fig. 4 shows the scatter plot for best model M72.2.5. From the scatter plot, almost all the standardized residuals are within the range of ± 3 . Hence the standardized residuals for the best model are said to be randomly distributed. In addition, the goodness of a regression model is also verified by the normality of the standardized residuals. The normality test on the standardized

residuals based on Shapiro-Wilk statistics (df=26) gives a value of 0.956 with a significant p-value of 0.327, hence, greater than at 0.05. The standardized residuals are thus said to be normally distributed. The bell-shaped normality plot as shown in Fig. 4(b), depicts that model M72.2.5 is normally distributed.

TABLE IX: RESERVED DATA FOR VALIDATION

Y	X_1	X_2	X_3	X_4	X_5
87749835	3947763	3915924	20.18	2777.5	500817
66766810	4082124	3477438	19.20	2236.5	502229
64282738	4202381	3565321	18.03	2701.0	446368
71384441	4302283	3625416	19.69	3219.0	421465
69834479	4385241	3696902	18.89	2764.0	435763

Table IX shows the reserved data for validation of the model. By substituting the values into equation (5), the forecasted Y values can thus be obtained. Table X hence shows the actual Y and forecasted Y values of the reserved data.

TABLE X: RESERVED DATA USED FOR MAPE (MODEL TRANSFORMATION)

(At) (Ft) (At-Ft)/At APE x100	
87749835.0000 79633121.5502 0.0925 9.2498	
<u>66766810.0000</u> <u>66637226.4694</u> 0.0019 0.1941	
64282738.0000 62918626.7757 0.0212 2.1220	
71384441.0000 71364957.3186 0.0003 0.0273	
69834479.0000 68796599.1205 0.0149 1.4862	
Total 13.0795	
MAPE 2.6159	

By substituting all the observations of actual Y and forecasted Y values into equation (5), the MAPE value is hence obtained as shown in Table X. The MAPE value is found to be 2.62% to 3%. Since the MAPE value is less than 25%, this is thus considered to be an acceptable model that can be used for forecasting.

For Model II, there is only one selected model which is M2.0.0. The best estimated export model is shown below in (6):

$$M2.0.0 = \hat{Y} = \hat{\beta}_0 + \hat{\beta}_2 X_2 \tag{6}$$

TABLE XI: COEFFICIENT TABLE FOR EXPORT VOLUME M2.0.0

	Coefficients	Standard Error	t Stat	P-value
Intercept	-1.863X10 ³	2.136X104	-0.08723	0.93598
X ₂	3.411X10 ²	5.014X10	6.80277	0.00650

By referring to the coefficient table in Table XI, and putting the values of the coefficients into the best model M2.0.0, the forecasted export volume equation model is given as in (7):

$$M2.0.0 = \hat{Y} = -1.863x10^3 + 3.411x10^2 X_2 \tag{7}$$

The export volume model is also subjected to the goodness-of-fit test for its robustness for forecasting and prediction as shown in Fig. 5.







The scatter plot for Model M2.0.0 is used as supporting evidence to support that the model is randomly distributed. From the scatter plot in Fig. 5(a) above, almost all the standardized residuals are within the range of ± 3 . Hence, the standardized residuals for the export volume model M2.0.0 are said to be randomly distributed. The normality test on the standardized residuals based on Shapiro-Wilk statistics (*df*=5) gives a value of 0.995 with a significant *p*-value of 0.995. Since it is greater than at 0.05, therefore, the standardized residuals are be said to be normally distributed.

From the forecasted model M72.2.5 in equation (5), it is thus found that the significant variables that have contributed to the production of oil palm fruits are one single independent variable, X_5 and six interaction variables (X_{23} , X_{24} , X_{25} , X_{34} , X_{245} and X_{345}). It is realized that these interaction variables have significant contribution to the production model, and they cannot occur as single independent variables.

The forecasted model M72.2.5 can be described as that the total workers employed during the last pay period (X_5) is the only main factor that gives a negative contribution to the production of oil palm fruits. It has a strong correlation (0.9801) with the production of oil palm fruits. Besides that, the harvested area also interacts with the oil palm yield per hectare (X_{23}) showing a positive contribution to the production of oil palm fruits up to the first order interaction factor. It also has a strong correlation (0.9853) with the production of oil palm fruits. The harvested area that interacts with the local delivery average price (X_{24}) also shows a positive contribution to the production of oil palm fruits. It too has a strong correlation (0.9277) with the production of oil palm fruits. On the other hand, the harvested area also interacts with the total workers

employed during last pay period (X_{25}) showing a positive contribution to the production of oil palm fruits up to first order interaction. This also has a strong correlation (0.8968) with the production of oil palm fruits. Besides that, the yield per hectare that interacts with the local delivery average price (X_{34}) shows a negative contribution to the production of oil palm fruits up to first order interaction factor. It has a weak correlation (0.2883) with the production of oil palm fruits. In addition, the harvested area that interacts with the local delivery average price and with the total workers employed during last pay period (X_{245}) shows a negative contribution and medium correlation (0.6067) to the production of oil palm fruits up to the second order interaction. Lastly, the yield per hectare that interacts with the local delivery average price and with the total workers employed during last pay period (X_{345}) shows that there is a positive contribution and medium correlation (0.7165) to the production of oil palm fruits up to second order interaction.

The value of 3.158×10^7 which is the intercept of the model indicates the production of oil palm fruits in tonnes when all the other independent variables from the model are equal to zero. According to [13], this is a mechanical interpretation of the intercept term. In regression analysis, however, such literal interpretation of the intercept term may be meaningless. Paradoxically, while the value is generally meaningless, it is crucial to include the constant term in most of the regression models. It can guarantee that the residuals have a mean of zero, and additionally, it can also avoid the regression coefficients and predictions of the models to be biased.

In this research, from best model in Model I, model M2.0.0 is shown to be the forecasted model in Model II that can be used to indicate the export volume of biodiesel. It is found out that the main significant factor that contributes to the export volume of biodiesel is the export value of biodiesel. The export value of the biodiesel (X_2) gives a positive contribution, and has a strong correlation (0.9691) with the export volume of biodiesel. The value of -1.863×10^3 which is the intercept of the model indicates the export volume of biodiesel in tonnes when all the independent variables from the model are equal to zero. Since it is impossible for the export volume to have a negative value, hence, the intercept term is meaningless.

The best model M72.2.5 for Model I has undergone the MAPE test for forecasting purposes. The MAPE value for best model M72.2.5 using the model transformation method is found to be 2.62%. Since the value is less than 25%, model M72.2.5 is hence suitable for use in forecasting the export volume of biodiesel.

V. CONCLUSION

This research is carried out to determine the factors that would contribute to the production of oil palm fruits, i.e. the production model, and ultimately, this production model is further used to forecast the export volume of biodiesel after the processes of trans-esterification and esterification. The best model M72.2.5 using the model transformation method implicates that the main factor that influences the production of oil palm fruits is the total workers employed during last pay period. The negative coefficient of the total workers employed during last pay period shows a negative contribution to the production of oil palm fruits. In other words, for every single increase (or decrease) in the total number of workers employed during the last pay period would cause a decrease (or increase) in the production of oil palm fruits for approximately 9.204×10^9 tonnes. There are six interaction factors exist in model M72.2.5. They are harvested area (X_2) interacts with the yield per hectare (X_3) , Harvested area (X_2) interacts with local delivery average price (X_4) , harvested area (X_2) interacts with total workers employed during last pay period (X_5) , yield per hectare (X_3) interacts with local delivery average price (X_4) , harvested area (X_2) interacts with local delivery average price (X_4) interacts with total workers employed during last pay period (X_5) and yield per hectare (X_3) interacts with local delivery average price (X_4) interacts with total workers employed during last pay period (X_5) .

This research also manages to identify the factor that influences the export volume of biodiesel. The model M2.0.0 indicates that the main factor that influences the export volume is the export value of biodiesel. The positive coefficient of the export value of biodiesel in the model shows the positive contribution to the export volume of biodiesel. This means that the export value of biodiesel would increase the export volume of biodiesel when the price value of biodiesel has attained an economic and competitive global market pricing. In other words, positive quantum efforts have to made by oil producing countries, environmentally concerned governments and bodies, green technologies companies, and etc. to help in stabilising and regulating the price, and hence, consequently the production of this sustainable green energy.

APPENDIX

All Possible 80 models OF model i M1: $Y = \beta_0 + \beta_1 X_1 + u$ M2: $Y = \beta_0 + \beta_2 X_2 + u$ M3: $Y = \beta_0 + \beta_3 X_3 + u$ M4: $Y = \beta_0 + \beta_4 X_4 + u$ M5: $Y = \beta_0 + \beta_5 X_5 + u$ M6: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$ M7: $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + u$ M8: $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_4 + u$ M9: $Y = \beta_0 + \beta_1 X_1 + \beta_5 X_5 + u$ M10: $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + u$ M11: $Y = \beta_0 + \beta_2 X_2 + \beta_4 X_4 + u$ M12: $Y = \beta_0 + \beta_2 X_2 + \beta_5 X_5 + u$ M13: $Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 + u$ M14: $Y = \beta_0 + \beta_3 X_3 + \beta_5 X_5 + u$ M15: $Y = \beta_0 + \beta_4 X_4 + \beta_5 X_5 + u$ M16: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$ M17: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + u$

M18:
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_5 X_5 + u$$

M19: $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4 + u$
M20: $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_3 X_5 + u$
M21: $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_4 + \beta_5 X_5 + u$
M22: $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$
M23: $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_5 X_5 + u$
M24: $Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$
M25: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$
M27: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_5 X_5 + u$
M28: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_5 X_5 + u$
M29: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$
M30: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$
M31: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$
M32: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$
M33: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$
M33: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$
M33: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$
M34: $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_{13} X_{13} + u$
M35: $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_5 + \beta_{15} X_{15} + u$
M36: $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_2 X_{23} + u$
M37: $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_5 + \beta_{25} X_{25} + u$
M39: $Y = \beta_0 + \beta_3 X_3 + \beta_3 X_5 + \beta_{25} X_{25} + u$
M41: $Y = \beta_0 + \beta_3 X_3 + \beta_3 X_5 + \beta_{25} X_{25} + u$
M42: $Y = \beta_0 + \beta_3 X_3 + \beta_3 X_5 + \beta_{35} X_5 + u$
M41: $Y = \beta_0 + \beta_3 X_1 + \beta_2 X_2 + \beta_3 X_5 + \beta_{25} X_{25} + u$
M42: $Y = \beta_0 + \beta_3 X_1 + \beta_2 X_2 + \beta_3 X_5 + \beta_{25} X_{25} + u$
M44: $Y = \beta_0 + \beta_3 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_3 X_{13} + \beta_{23} X_{23} + u$
M42: $Y = \beta_0 + \beta_3 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_3 X_{13} + \beta_{23} X_{23} + u$
M44: $Y = \beta_0 + \beta_3 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_3 X_{31} + \beta_{33} X_{33} + \alpha_{33} X_{34} + u$
M44: $Y = \beta_0 + \beta_3 X_1 + \beta_3 X_3 + \beta_3 X_3 + \beta_4 X_4 + \beta_{12} X_{12} + \beta_{33} X_3 + u$
M45: $Y = \beta_0 + \beta_3 X_1 + \beta_3 X_2 + \beta_3 X_3 + \beta_{33} X_3 + \beta_{33} X_3 + \beta_{33} X_3 + \beta_{33} X_{34} + \beta_{33} X_{34} + \beta_{33} X_{34} + \beta_{32} X_{34} + \beta_{33} X_{34} +$

$$+\beta_{14}X_{14} + \beta_{24}X_{24} + \beta_{124}X_{124} + u$$

M60: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_5 X_5 + \beta_{12} X_{12}$

M60:
$$+\beta_{15}X_{15} + \beta_{25}X_{25} + \beta_{125}X_{125} + u$$

ACKNOWLEDGEMENT

The authors would like to thank Ms. Chen Lee Lee for part of her statistical analysis in this research.

REFERENCES

- B. Yusof, "Biodiesel impact on the palm oil industry," *Global Oils and Fats Business Magazine*, vol. 5, no. 3, pp. 1-7, 2008.
- [2] F. A. H. A. Fadli, N. A. R. Hayati, A. A. R. Errie, A. S. A. Bashir, N. A. H. Fahana, and K. Jusoft, "A time series analysis of the relationship between total area planted, palm oil price and production of Malaysian palm oil," *World Applied Sciences Journal*, vol. 12, pp. 34-40, 2011.
- [3] Department of Statistics Malaysia (DOSM). (2011). [Online]. Available: http://www.statistics.gov.my/main/main.php
- [4] Malaysian Palm Oil Board (MPOB). (2013). [Online]. Available: http://www.mpob.gov.my/
- [5] H. J. Zainodin, N. Abdullah, and G. Khuneswari, "Consumer behavioral buying patterns on the demand for detergents using hierarchically multiple polynomial regression model," *Science International*, vol. 2, no. 3, pp. 64-71, 2014.
- [6] P. Roxy, O. Chris, and D. Jay, Introduction to Statistics and Data Analysis, 4th edition, Cengage Learning, United States, 2011.
- [7] A. Noraini and H. J. Zainodin, "Multicollinearity remedial techniques in model-building," *MATEMATIKA*, vol. 29, pp. 107-115, 2013.
- [8] N. Abdullah, H. J. Zainodin, and A. Ahmed, "Sustainable urban forest using multiple regression models," *Research Journal of Forestry*, vol. 6, no. 1, pp. 1-15, 2012.
- [9] N. Abdullah, A. Ahmed, and H. J. Zainodin, "An improved volumetric estimation using polynomial regression," *Journal of Science & Technology (UTHM)*, vol. 3, no. 2, pp. 29-42, 2011.
- [10] A. Noraini, and H. J. Zainodin, "A polynomial stem volume estimation on cinnamon wood for cinnamomum extracts of aldehyde content," *Science International*, vol. 2, no. 2, pp. 44-50, 2014.
- [11] H. J. Zainodin and G. Khuneswari, "A case study on determination of house selling price model using multiple regression," *Malaysian Journal of Mathematical Sciences*, vol. 3, pp. 27-44, 2009.

- [12] H. J. Zainodin, N. Abdullah, and S. J. Yap, "An alternative multicollinearity approach in solving multiple regression problem," *Trends in Applied Sciences Research*, vol. 6, no. 11, pp. 1241-1255, 2011.
- [13] D. N. Gujarati, and D. C. Porter, *Essentials of Econometrics*, 5th edition, New York: McGraw-Hill, 2009.



Noraini Abdullah was born on December 3, 1959 in Malaysia. Noraini Abdullah is a senior lecturer with the Mathematics and Economics Programme in the Faculty of Science and Natural Resources, Universiti Malaysia Sabah. She obtained her bachelor of science degree from Universiti Kebangsaan Malaysia in 1999, the master of science degree in industrial mathematical modelling from Loughborough University of Technology, United Kingdom, 2003 and obtained her

doctorate (PhD) degree in mathematics with economics from Universiti Malaysia Sabah, 2013. At present, she is the vice-chairman of the Mathematical and Statistical Application Research Group (MASA) of Universiti Malaysia Sabah. Examples of her research projects are: 1) Modelling of the Insects Population Dynamics for the Fisheries Industries, 2) A Study of Community's Communication Strategies and Modelling of the Community Transformation, 3) Consumer-Centric Study on the Usage Pattern of Television-Receive-Only (TVRO) in Sabah. Her current research interests are multiple regression models, multiple binary logistics models, multinomial logistics regressions, parameter estimation and mathematics with Economics. Three of her recent and related publications are: 1). Noraini Abdullah, H. J. Zainodin and L. B. Rick, "Risk factors determination on UGIB patients in Kota Kinabalu, Sabah, Malaysia," Journal of Medical Sciences, vol. 13, no. 7, pp. 526-536, 2013. 2). Noraini Abdullah, Rohana Tair, and M. Harun Abdullah, "Heavy metals concentration relationship with Perna viridis physical properties in Mengkabong Lagoon, Sabah, Malaysia," Pakistan Journal of Biological Sciences, vol. 17, no. 1, pp. 62-67, 2014. 3). Noraini Abdullah and H. J. Zainodin, "A polynomial stem volume estimation on cinnamon wood for cinnamomum extracts of aldehyde content," Science International, vol. 2, no. 2, pp. 44-50, 2014.