

## LOG PRODUCTION PREDICTION MODEL: A COMPARISON BETWEEN MALAYSIA AND INDONESIA USING MULTIPLE REGRESSION TECHNIQUE

Noraini Abdullah  
Mathematics with Economics Programme  
Universiti Malaysia Sabah, 88400  
Kota Kinabalu, Sabah, Malaysia  
E-mail norainiabdullah.ums@gmail.com

Annie Kiu Lee Lee  
Mathematics with Economics Programme  
Universiti Malaysia Sabah, 88400  
Kota Kinabalu, Sabah, Malaysia  
Email: anniekiu91@gmail

Walter Lintangah  
International Tropical Forestry Programme  
Universiti Malaysia Sabah, 88400  
Kota Kinabalu, Sabah, Malaysia  
Email: walterjl@ums.edu.my

---

### ABSTRACT

Timber production and its related products have played a vital role in the development and economic growth of Malaysia as well as Indonesia. Logging activities are hence common to both countries pertaining to the revenues generated from these activities. Thus, this paper aspires to determine the trends on log production of Malaysia and Indonesia, and do a comparison between them. Modelling procedures using the multiple regression technique are illustrated, where the best models were selected after the four model building phases, incorporating the multicollinearity and coefficient tests, the eight selection criteria (8SC) and the goodness-of-fit tests. Prediction capability is evaluated through the calculated mean absolute percentage error (MAPE) whereby values less than 25% indicate an excellent accuracy measurement in predicting the log production. The log production model from Indonesia is found to give a better prediction with an MAPE of 2.39% compared to Malaysia (19.48%).

Key words: Log production, prediction, multiple regression, modelling procedures, best model.

---

### Introduction

Logging activity is the main activity in forestry which will bring income to the governments of Malaysia and Indonesia. The amount of timber production can affect the countries' economies, development and economic growth via the export of timber and timber products. Besides being a common activity in both countries, these logging activities will also bring many benefits to the society and one of them would be job opportunities. The timber industry is one of the sectors that enables to boost both nation's source of income through its high price of timber and timber products, hence leading to the flow of foreign currencies.

According to Thang (2009), in order to ensure the effectiveness of forest management and the implementation of the National Forestry Policy in Malaysia, various forestry enactments and ordinances were formulated and enforced by the respective state authorities since the early 1900s. Through the adoption of the National Forestry Act 1984, those enactments and ordinances were strengthened in the aspect of forest management planning and operations. This act was legislated to protect the forest resources from illegal forest harvesting activities. In Peninsular Malaysia, the maximum volume allowed to be harvested is 85m<sup>3</sup>/ha while in Sabah, only trees that have the size of 60cm at diameter-at-breast height (dbh) and above can be removed. In Sarawak, the cutting limits for dipterocarp and non-dipterocarp species being 60cm and 45cm at dbh and above respectively (Forestry Department of Peninsular Malaysia, Sabah and Sarawak, 2009).

Over the last three decades, Indonesia's forest resources had been the principle engine of national economic development. Forest resources had also contributed to foreign exchange revenues, employment, regional development and economic growth. According to the Forestry Department of Indonesia, Indonesian forest area comprises of 112.27 million hectares of permanent forest and 8.08 million hectare of forest area that are designated for conservation to other land uses. Permanent forest consists of 3 categories which are area protection forest, conservation or fully protected forest and reservation forests, and production forest (Indonesia Forestry Department, 1997).

### Literature Reviews

Sopipan *et al.* (2012) had done a research on forecasting SET50 Index with multiple regression based on the principal component analysis. SET50 Index means the stock price of the top 50 listed companies on SET (Stock Exchange of Thailand). The researchers stated that the multiple regression analysis is one of the most common used methodologies for expressing the

dependence of a response variable on several independent variables. All the data collected was in the period 4/01/2007 through 30/03/2011. The data set was obtained from the Stock Exchange of Thailand. The researchers considered the existence of a high correlation between the explanatory variables where multicollinearity problem exist. Principal component analysis (PCA) had been used to reduce the number of variables and to get rid of the problem of multicollinearity. In this study, the daily closed prices of SET50 Index at time  $t$  was the dependent variable. The explanatory independent variables were the differences between the daily closed price factors. As the result, 97.4% of variation in SET50 could be explained by the first three PCs. 98.4% of SET50 explained by the PCA1, PCA2, PCA9, and PCA13. 99.4% of variation in SET50 can be explained by all the PCAs.

Noraini *et al.* (2012) had done a research on sustainable urban forest using multiple regression models. Urban forests sustainability can be mathematically modelled using multiple regression because the green canopy is directly related to the bole and tree volume. There are three volumetric formulas used to calculate the tree stem volume, namely Newton, Huber, and Smalian equations. The data were collected and categorized according to different sizes which were small (S), medium (M), and large (L). There were six independent variables based on measureable variables and five categorical variables based on location samples. Spearman correlation coefficient matrix was used to identify the bivariate relationship. Insignificant variables and high multicollinearity variables were removed during the analysis. Three selected best volumetric models from the equations were chosen based on the eight selection criteria (8SC). As a result, the Newton's multiple regression model on all the three sizes (S, M, L) were found to be the best model to represent the measurement growth factors which affect the sustainability of the urban forests.

Chen *et al.* (2008) had also done a research on comparing forecasting models in tourism. The objective of this research was to implement basic, intermediate, and advanced extrapolative methods to the USNPS data. The second objective was to evaluate the advantages and disadvantages of using basic intermediate and advanced time series methods. This research had focussed on examining the different statistical techniques and applying them to the National Park Service data in the United States. The data was taken from year 1960 to 2000 from selected national parks. The selected national parks had included the Great Smoky Mountains National Park, Yellowstone National Park, and the Yosemite National Park. Park forecasting models consisted of three stages which were specification, estimation, and evaluation. Specification focussed on testing, generating, and selecting an appropriate model. Estimation stage used selected model to predict future visitation. Evaluation stage investigated various models over time to see if there were any need to modify the models. The results of their research showed that the MAPE values had indicated that ARIMA and SES appeared in the first position twice, while Naive1, TSCSREG, time series analysis with explanatory variable, Holt's, and SMA models appeared in the first position once. Naive 2 had appeared in last position five times, whilst SES appeared in last position twice, and Holt's and TSCSREG had appeared in last position once.

**Methodology**

This research was conducted based on the log production in Malaysia and Indonesia. The data used in this research were taken from Department of Statistics Malaysia and the World Data Bank. The variables chosen for this research were the amount of log production, carbon dioxide emission, labor force, forest area, and gross domestic product (GDP). Amount of log production would be the dependent variable while the independent variables were carbon dioxide emission, forest area, labor force, and GDP. There were only 21 yearly data that had been collected. with 16 observations were used to obtain the production model, while the other 5 observations were used for prediction.

Multiple regression is a technique that can be used to predict the relationship between the dependent variable with two or more independent variables relating to it. According to Zainodin *et al.* (2011), the general form of multiple regression can be shown as follows:

$$Y = \Omega_0 + \Omega_1 W_1 + \Omega_2 W_2 + \dots + \Omega_k W_k + random\ scatter \dots\dots\dots(1)$$

In equation (1), Y is the dependent variable,  $\Omega_0$  is the constant term,  $\Omega_j$  is the j-th coefficient of independent variable  $W_j$ , and  $W_j$  is the j-th independent variables (included the single independent variables, interaction variables, generated dummy variables and transformed variables) where  $j=1, 2, \dots, k$ . In this research, there was one dependent variable with four independent variables. Therefore, the model equation would be:

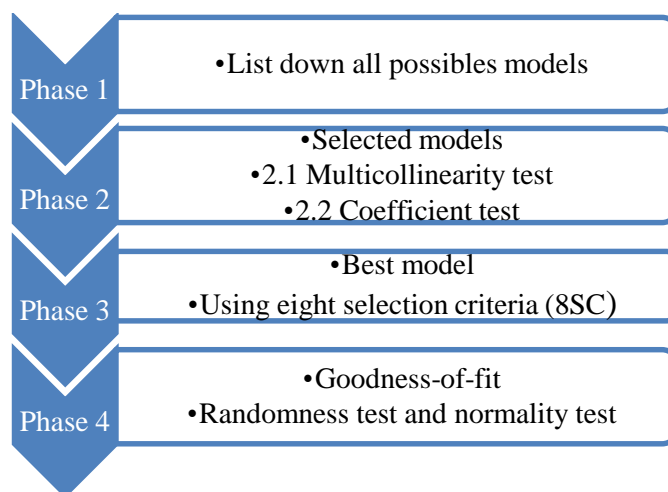
$$Y = \Omega_0 + \Omega_1 W_1 + \Omega_2 W_2 + \Omega_3 W_3 + \Omega_4 W_4 + random\ scatter \dots\dots\dots(2)$$

,where Y is the log production,  $W_1$  is the carbon dioxide emission,  $W_2$  is the forest area,  $W_3$  is the labor force, and  $W_4$  is the GDP.

**Modelling Procedures**

Zainodin *et al.* (2011) had demonstrated the four phases in the model building procedures of multiple regression which started from listing down all the possible models in Phase 1 to carrying out the goodness-of-fit on the residual of the best model in Phase 4, as depicted in Figure 1.

Figure 1. Model-Building Procedures

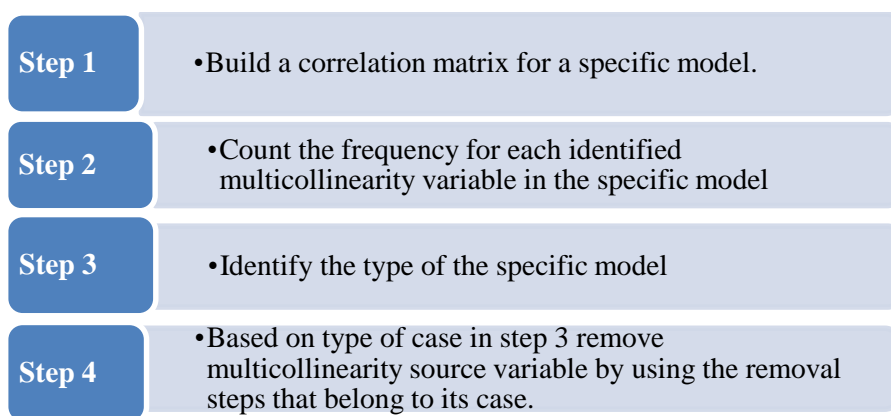


According to Noraini *et al.* (2012), all the possible models consisting of the different combinations of variables and their interaction were initially listed out for consideration. The number of possible models can be calculated and is given by the

following formula:  $N = \sum_{j=1}^q j \binom{q}{j}$  where N is the number of all possible models, q is the number of single independent

variables (for j=1, 2, 3, ..., q) exclusive of dummy variables (Zainodin and Khuneswari, 2009). The possible models for (say) 4 single independent variables can thus be calculated as 32 models. After listing out all the possible models, the models would then go through the multicollinearity test, followed by the coefficient test of Phase 2. The multicollinearity test was used to remove multicollinearity source variables with absolute coefficients greater than 0.95. The removal of these non contributing variables could be found in Zainodin *et al.* (2011). Figure 2 below showed the multicollinearity test procedures that had been carried out.

Figure 2. Multicollinearity Test Procedures



The coefficient test was carried out on the reduced model to eliminate insignificant variables from each model using the backward elimination method. According to Zainodin and Khuneswari (2009), the coefficient test is used to test the coefficient of the corresponding variables where the insignificant variable would be eliminated. Detailed elimination procedures can be found in Noraini *et al.* 2011. After these tests had been carried out, the selected models were then obtained in Phase 3. The model labelling could also be found in Zainodin *et al.* 2011.

The best model in Phase 3 was chosen from the selected models based on the eight selection criteria (8SC). According to Zainodin *et al.* (2011), the eight selection criteria (8SC) were carried out on the selected models which had undergone filtration. The eight selection criteria (8SC) were shown in Table 1 below.

Table 1. Eight Selection Criteria

<p>AIC: (Akaike, 1974)</p> $\left(\frac{SSE}{n}\right) e^{(2m/n)}$	<p>RICE: (Rice, 1984)</p> $\left(\frac{SSE}{n}\right) \left[1 - \left(\frac{2m}{n}\right)\right]^{-1}$
--	--

FPE: (Akaike, 1969) $\left(\frac{SSE}{n}\right) \frac{n+m}{n-m}$	SCHWARZ: (Schwarz, 1978) $\left(\frac{SSE}{n}\right) n^{m/n}$
GCV: (Golub <i>et al.</i> , 1979) $\left(\frac{SSE}{n}\right) \left[1 - \left(\frac{m}{n}\right)\right]^{-2}$	SGMASQ: (Ramanathan, 2002) $\left(\frac{SSE}{n}\right) \left[1 - \left(\frac{m}{n}\right)\right]^{-1}$
HQ: (Hannan & Quinn, 1979) $\left(\frac{SSE}{n}\right) (\ln n)^{2m/n}$	SHIBATA: (Shibata, 1981) $\left(\frac{SSE}{n}\right) \frac{n+2m}{n}$

In Phase 4, the best model would go through the goodness-of-fit test. According to Maydeu and Garcia (2010), the goodness-of-fit would describe how well a statistical model fitted into a set of observations. Goodness-of-fit indices would summarize the discrepancy between the observed value and the values expected under a statistical model. There were two tests under the goodness-of-fit which were randomness test and the normality test. The randomness test was used to test the randomness of residuals (Zainodin *et al.*, 2011). Histogram and scatter plots of the residuals can be used to obtain the distribution of the residuals. The Kolmogorov-Smirnov test and Shapiro-Wilk test are the most common used to test the normality of the residuals. When the number of observations is large, Kolmogorov-Smirnov test is used while Shapiro-Wilk test is used when number of observations is small. In this research, Shapiro-Wilk test was used because the number of sample size was only 16. The null hypothesis would be accepted if the p-value was greater than 0.05. When the null hypothesis was accepted, it showed that the data was normally distributed. According to Ghasemi and Zahediasl (2012), scatter plot, histogram, and normal P-P plot could be used as supporting evidences for the normality test.

**Prediction Model Using MAPE**

Mean Absolute Percent Error (MAPE) measures the size of the error in percentage term. According to Ren (2009), MAPE was widely used as the accuracy measurement in forecasting with non-negative observations. Chen *et al.* (2008) stated that accuracy was the most important forecasting evaluation criterion. They stated that if the MAPE value was less than 10% is highly accurate forecasting. If the MAPE value was between 10% to 20%, it is a good forecasting model, while if the MAPE was more than 50 %, it would be an inaccurate forecasting model. MAPE for the best models from each country was calculated using the formula shown as:

$$MAPE = \left( \sum_{i=1}^k \frac{|A_t - F_t|}{A_t} \right) / k \times 100\% \quad \dots\dots\dots (3)$$

where, k is the number of observations used for forecasting,  $A_t$  is the actual value of Y and  $F_t$  is the estimated value of Y.

**Results**

In this research, 21 data sets representing both countries (Malaysia and Indonesia) had been collected with 16 data observations were used for model building, while 5 were used for prediction using MAPE. The normality test was carried out and results indicated that to fulfil the assumptions of multiple regression, ladders-of-power transformation had to be performed on the non-normal data sets for normality and linearity. Table 2 below showed the normality tests (a) before and (b) after transformation on the variables. It could be seen that before transformation, the non-normal variables  $L_{og}P$  and GDP had significant p-values which were less than 0.05 (highlighted in green). Transformations carried out on the non-normal data of  $L_{og}P$  and GDP were inverse and square-root, so as to obtain the after transformation variables of Y and  $X_4$  respectively with p-values greater than 0.05. The corresponding transformed variables for modelling were indicated in Table 2 (b) below. It can be seen that all the variables had p-values more than 0.05, hence, were now have normal distribution as highlighted in yellow.

Table 2. Normality Tests Before and After Transformation (Malaysia).

(a) Before Transformation	Shapiro-Wilk			(b)After Transformation	Shapiro-Wilk		
	Stats	df	P-value		Stats	df	P-value
$L_{og}P$	0.810	16	0.004	$1/L_{og}P=Y$	0.810	16	0.070
$CO_2E$	0.898	16	0.075	$CO_2E=X_1$	0.898	16	0.075
$FA_{rea}$	0.932	16	0.261	$FA_{rea}=X_2$	0.932	16	0.261
$L_{abor} F$	0.958	16	0.625	$L_{abor} F=X_3$	0.958	16	0.625
GDP	0.863	16	0.021	$\sqrt{GDP}=X_4$	0.863	16	0.061

Similar procedures were carried out on the data sets for Indonesia. Table 3 below showed the normality tests (a) before and (b) after transformation on the variables. It could be seen that before transformation, the non-normal variable GDP had significant p-value (0.034) which was less than 0.05 (highlighted in blue). Transformation carried out on the non-normal data of GDP was inverse square-root, so as to obtain the after transformation variable of  $X_4$ . The corresponding transformed variables for

modelling were indicated in Table 3(b) below. It can be seen that all the variables had p-values more than 0.05, hence, the variables were normal as highlighted in pink.

Table 3. Normality Tests Before and After Transformation (Indonesia).

(a) Before Transformation	Shapiro-Wilk			(b) After Transformation	Shapiro-Wilk		
	Stats	df	P-value		Stats	df	P-value
LogP	0.916	16	0.145	LogP=Y	0.896	16	0.145
CO <sub>2</sub> E	0.962	16	0.689	CO <sub>2</sub> E=X <sub>1</sub>	0.898	16	0.689
FA <sub>area</sub>	0.907	16	0.103	FA <sub>area</sub> =X <sub>2</sub>	0.932	16	0.103
L <sub>abor</sub> F	0.951	16	0.512	L <sub>abor</sub> F=X <sub>3</sub>	0.958	16	0.512
GDP	0.876	16	0.034	1/√GDP=X <sub>4</sub>	0.973	16	0.883

### Multiple Regression Model Development

Model development consisted of four phases comprised of Phase 1 from listing down all the possible models to Phase 4 in carrying out the goodness-of-fit on the residual of the best model (Zainodin *et al.*, 2011). According to Noraini *et al.* (2012), all the possible models consisted of different combinations of variables and their interactions with N = the number of all possible models and q = the number of single independent variables (for j=1, 2, 3, ..., q) which does not include dummy variables. The number of possible models had been calculated as:

$$N = \sum_{j=1}^q j({}^qC_j) = 1({}^4C_1) + 2({}^4C_2) + 3({}^4C_3) + 4({}^4C_4) = 1(4) + 2(6) + 3(4) + 4(1) = 32 \text{ models}$$

All the model building procedures were carried out to get 6 selected models for Malaysia and 10 selected models for Indonesia. The best model for both countries was then chosen from these selected models based on the eight selection criteria (8SC). Table 4 showed all the selected model equations for both countries.

Table 4. Comparisons of Selected Model Equations for Malaysia and Indonesia

MALAYSIA	INDONESIA	
M1.0.0: Y=β <sub>0</sub> +β <sub>1</sub> X <sub>1</sub>	M2.0.0: Y=β <sub>0</sub> +β <sub>2</sub> X <sub>2</sub>	M25.5.0: Y=β <sub>0</sub> + β <sub>24</sub> X <sub>24</sub>
M2.0.0: Y=β <sub>0</sub> +β <sub>2</sub> X <sub>2</sub>	M4.0.0: Y=β <sub>0</sub> +β <sub>4</sub> X <sub>4</sub>	M28.4.0: Y=β <sub>0</sub> +β <sub>2</sub> X <sub>2</sub> +β <sub>24</sub> X <sub>24</sub> + β <sub>124</sub> X <sub>124</sub>
M3.0.0: Y=β <sub>0</sub> +β <sub>3</sub> X <sub>3</sub>	M9.0.0: Y=β <sub>0</sub> +β <sub>2</sub> X <sub>2</sub> +β <sub>4</sub> X <sub>4</sub>	M29.4.1: Y=β <sub>0</sub> + β <sub>4</sub> X <sub>4</sub> + β <sub>14</sub> X <sub>14</sub> + β <sub>134</sub> X <sub>134</sub>
M4.0.0: Y=β <sub>0</sub> +β <sub>4</sub> X <sub>4</sub>	M18.1.0: Y=β <sub>0</sub> +β <sub>4</sub> X <sub>4</sub> + β <sub>14</sub> X <sub>14</sub>	M31.10.2: Y=β <sub>0</sub> +β <sub>124</sub> X <sub>124</sub> + β <sub>134</sub> X <sub>134</sub>
M6.0.0: Y=β <sub>0</sub> +β <sub>1</sub> X <sub>1</sub> +β <sub>3</sub> X <sub>3</sub>	M20.1.0: Y=β <sub>0</sub> +β <sub>2</sub> X <sub>2</sub> + β <sub>24</sub> X <sub>24</sub>	
M19.2.0: Y=β <sub>0</sub> +β <sub>23</sub> X <sub>23</sub>	M23.3.1: Y=β <sub>0</sub> +β <sub>14</sub> X <sub>14</sub> + β <sub>24</sub> X <sub>24</sub>	

From Table 5, model M6.0.0 was the best model among the six selected models for Malaysia

Table 5. Eight Selection Criteria (8SC) for MALAYSIA Best Model

Model	AIC	FPE	GCV	HQ	RICE	SCHWARZ	SGMASQ	SHIBATA
M1.0.0	4.4299 X10 <sup>-11</sup>	4.4357 X10 <sup>-11</sup>	4.5061 X10 <sup>-11</sup>	4.4519 X10 <sup>-11</sup>	4.600 X10 <sup>-11</sup>	4.8790 X10 <sup>-11</sup>	3.9429 X10 <sup>-11</sup>	4.3125 X10 <sup>-11</sup>
M2.0.0	3.0736 X10 <sup>-11</sup>	3.0777 X10 <sup>-11</sup>	3.1265 X10 <sup>-11</sup>	3.0889 X10 <sup>-11</sup>	3.1917 X10 <sup>-11</sup>	3.3853 X10 <sup>-11</sup>	2.7357 X10 <sup>-11</sup>	2.9922 X10 <sup>-11</sup>
M3.0.0	2.4477 X10 <sup>-11</sup>	2.4509 X10 <sup>-11</sup>	2.4898 X10 <sup>-11</sup>	2.4598 X10 <sup>-11</sup>	2.5417 X10 <sup>-11</sup>	2.6958 X10 <sup>-11</sup>	2.1786 X10 <sup>-11</sup>	2.3828 X10 <sup>-11</sup>
M4.0.0	4.8151 X10 <sup>-11</sup>	4.8214 X10 <sup>-11</sup>	4.8980 X10 <sup>-11</sup>	4.8390 X10 <sup>-11</sup>	5.0000 X10 <sup>-11</sup>	5.3033 X10 <sup>-11</sup>	4.2857 X10 <sup>-11</sup>	4.6875 X10 <sup>-11</sup>
M6.0.0	1.8006 X10 <sup>-11</sup>	1.8087 X10 <sup>-11</sup>	1.8746 X10 <sup>-11</sup>	1.8140 X10 <sup>-11</sup>	1.9800 X10 <sup>-11</sup>	2.0812 X10 <sup>-11</sup>	1.5231 X10 <sup>-11</sup>	1.7016 X10 <sup>-11</sup>
M19.2.0	2.2792 X10 <sup>-11</sup>	2.2821 X10 <sup>-11</sup>	2.3184 X10 <sup>-11</sup>	2.2904 X10 <sup>-11</sup>	2.3667 X10 <sup>-11</sup>	2.5102 X10 <sup>-11</sup>	2.0286 X10 <sup>-11</sup>	2.2188 X10 <sup>-11</sup>
MIN	1.8006 X10 <sup>-11</sup>	1.8087 X10 <sup>-11</sup>	1.8746 X10 <sup>-11</sup>	1.8140 X10 <sup>-11</sup>	1.9800 X10 <sup>-11</sup>	2.0812 X10 <sup>-11</sup>	1.5231 X10 <sup>-11</sup>	1.7016 X10 <sup>-11</sup>

Table 6 below showed that model M29.4.1 was the best model among the 10 selected models for Indonesia.

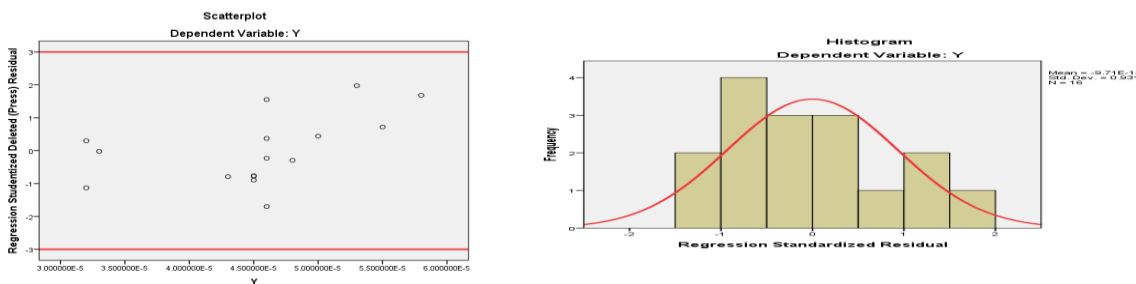
Table 6. Eight Selection Criteria (8SC) for INDONESIA Best Model

Model	AIC	FPE	GCV	HQ	RICE	SCHWARZ	SGMASQ	SHIBATA
M2.0.0	4.5260 $\times 10^{13}$	4.5320 $\times 10^{13}$	4.6039 $\times 10^{13}$	4.5485 $\times 10^{13}$	4.6998 $\times 10^{13}$	4.9849 $\times 10^{13}$	4.0284 $\times 10^{13}$	4.4061 $\times 10^{13}$
M4.0.0	7.8969 $\times 10^{13}$	7.9072 $\times 10^{13}$	8.0328 $\times 10^{13}$	7.9360 $\times 10^{13}$	8.2001 $\times 10^{13}$	8.6975 $\times 10^{13}$	7.0287 $\times 10^{13}$	7.6876 $\times 10^{13}$
M9.0.0	2.9343 $\times 10^{13}$	2.9475 $\times 10^{13}$	3.0549 $\times 10^{13}$	2.9561 $\times 10^{13}$	3.2267 $\times 10^{13}$	3.3917 $\times 10^{13}$	2.4821 $\times 10^{13}$	2.7729 $\times 10^{13}$
M18.1.0	2.9145 $\times 10^{13}$	2.9276 $\times 10^{13}$	3.0343 $\times 10^{13}$	2.9362 $\times 10^{13}$	3.2050 $\times 10^{13}$	3.3688 $\times 10^{13}$	2.4654 $\times 10^{13}$	2.7543 $\times 10^{13}$
M20.1.0	2.9224 $\times 10^{13}$	2.9355 $\times 10^{13}$	3.0425 $\times 10^{13}$	2.9441 $\times 10^{13}$	3.2136 $\times 10^{13}$	3.3779 $\times 10^{13}$	2.4720 $\times 10^{13}$	2.7617 $\times 10^{13}$
M23.3.1	2.5479 $\times 10^{13}$	2.5593 $\times 10^{13}$	2.6526 $\times 10^{13}$	2.5668 $\times 10^{13}$	2.8018 $\times 10^{13}$	2.9450 $\times 10^{13}$	2.1552 $\times 10^{13}$	2.4078 $\times 10^{13}$
M25.5.0	6.3675 $\times 10^{13}$	6.3759 $\times 10^{13}$	6.4771 $\times 10^{13}$	6.3990 $\times 10^{13}$	6.6120 $\times 10^{13}$	7.0131 $\times 10^{13}$	5.6674 $\times 10^{13}$	6.1988 $\times 10^{13}$
M28.4.0	2.3676 $\times 10^{13}$	2.3933 $\times 10^{13}$	2.5529 $\times 10^{13}$	2.3911 $\times 10^{13}$	2.8720 $\times 10^{13}$	2.8720 $\times 10^{13}$	1.9147 $\times 10^{13}$	2.1540 $\times 10^{13}$
M29.4.1	1.0489 $\times 10^{13}$	1.0536 $\times 10^{13}$	1.0920 $\times 10^{13}$	1.0567 $\times 10^{13}$	1.1534 $\times 10^{13}$	1.2124 $\times 10^{13}$	8.8723 $\times 10^{12}$	9.9120 $\times 10^{12}$
M31.10.2	1.5139 $\times 10^{13}$	1.5207 $\times 10^{13}$	1.5761 $\times 10^{13}$	1.5252 $\times 10^{13}$	1.6648 $\times 10^{13}$	1.7499 $\times 10^{13}$	1.2806 $\times 10^{13}$	1.4307 $\times 10^{13}$
MIN	1.0489 $\times 10^{13}$	1.0536 $\times 10^{13}$	1.0920 $\times 10^{13}$	1.0567 $\times 10^{13}$	1.1534 $\times 10^{13}$	1.2124 $\times 10^{13}$	8.8723 $\times 10^{12}$	9.9120 $\times 10^{12}$

The best models would then run through the goodness-of-fit tests which were the run test and normality test. The run test is used to check whether the observations occurred in a random order. The hypothesis nol ( $H_0$ ) is that the standardized residuals were randomly distributed, and the hypothesis alternative ( $H_1$ ) is that the standardized residuals were not randomly distributed. For model M6.0.0 (Malaysia), since the probability Asymp.Sig (2-tailed) = 0.438 >  $\alpha$  = 0.05, therefore  $H_0$  could not be rejected, meaning that the standardized residuals were randomly distributed.

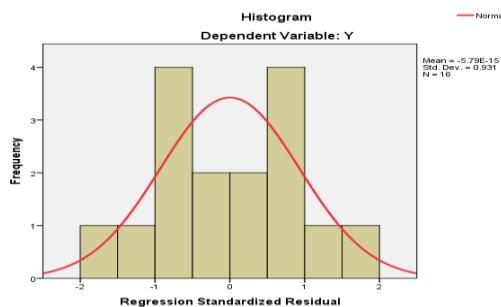
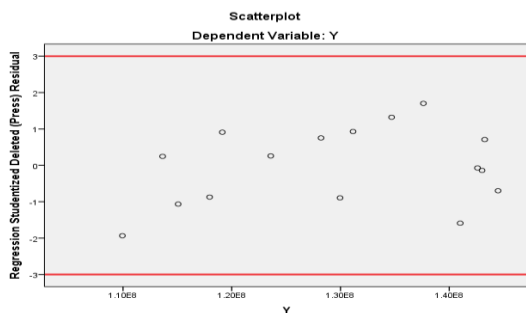
The normality can be observed visually using histogram and by the significant statistical test using Shapiro-Wilk. The statistical test of the standardized residuals gave a value of 0.951 at 0.507 significance level. Since its p-value is greater than 0.05, hence it could be said the standardized residuals had a normal distribution. Figure 3 showed that the scatter plot and histogram of the standardized residuals of model M6.0.0 (Malaysia) were randomly and normally distributed as supporting evidences.

Figure 3. Scatter plot and Histogram of best model M6.0.0 (Malaysia)



Similarly, the run test and normality test were carried out for model M29.4.1 (Indonesia). Since the probability Asymp.Sig (2-tailed) = 0.796 >  $\alpha$  = 0.05, therefore  $H_0$  could not be rejected, meaning that the standardized residuals were randomly distributed. The normality can be observed visually using histogram and by the significant statistical test using Shapiro-Wilk. The statistical test of the standardized residuals gave a value of 0.959 at 0.637 significance level. Since its p-value is greater than 0.05, hence it could be said the standardized residuals had a normal distribution. Figure 4 showed that the scatter plot and histogram of the standardized residuals of model M6.0.0 (Indonesia) were also randomly and normally distributed as supporting evidences.

Figure 4. Scatter plot and Histogram of model M29.4.1 (Indonesia)



**Prediction Model Using Mape**

MAPE for the best models from each country will be calculated by using the formula in (3). The estimated Y value was calculated first by substituting the 5 observations that had been reserved earlier into the best model, M6.0.0 (Malaysia) and model M29.4.1 (Indonesia).

$$M\ 6.0.0: Y^{-1} = -2.947x10^{-5} - 1.773x10^{-10} X_1 + 9.972x10^{-12} X_3 \dots\dots\dots(4)$$

$$M\ 29.4.1: Y = 1.900x10^8 + 2.620x10^8 X_{14} - 2.412x10^{-12} X_{134} \dots\dots\dots (5)$$

Table 7. MAPE For M6.0.0 (Malaysia)			Table 8. MAPE For M29.4.1 (Indonesia)		
A <sub>t</sub>	F <sub>t</sub>	( A <sub>t</sub> - F <sub>t</sub>  )/A <sub>t</sub>	A <sub>t</sub>	F <sub>t</sub>	( A <sub>t</sub> - F <sub>t</sub>  )/A <sub>t</sub>
40100	32405.4571	0.1919	163843496	155507833.4	0.0509
39859	32194.7136	0.1923	162227584	158379610.8	0.0237
43510	31163.327	0.2838	158234920	159523601.1	0.0081
37260	31814.7111	0.1461	154243096	156129794.1	0.0122
35671	29959.8538	0.1601	147701464	151331751.5	0.0246

$$MAPE_{Malaysia} = (0.1919+0.1923+0.2838+0.1461+0.1601)/5 \times 100\% = \mathbf{19.48\%}$$

$$MAPE_{Indonesia} = (0.0509 + 0.0237 + 0.0081 + 0.0122 + 0.0246)/5 \times 100\% = \mathbf{2.39\%}$$

**Discussions**

**Model Comparison Between Malaysia and Indonesia and Its Implications**

The amount of log production was the dependent variable, Y. The four independent variables were carbon dioxide emission (kt), X<sub>1</sub>; forest area (sq. km), X<sub>2</sub>; labor force, X<sub>3</sub>, and last will be the GDP (current US\$), X<sub>4</sub>. In this research, the models for both countries had the same dependent variable, and independents variable. Besides that, all the sample size of the countries was 21 with 16 observations used for modeling and 5 observations were used for forecasting. All models for both countries had undergone the modelling procedures as described in the methodology section. The best model for Malaysia was found to be:

$$Log\ Production^{-1} = -2.947x10^{-5} - 1.773x10^{-10} Carbon\ Dioxide\ Emission + 9.972x10^{-12} Labour\ Force$$

On the other hand, the best model for Indonesia was found to be:

$$Log\ Production = 1.900x10^8 + 2.620x10^8 Carbon\ Dioxide\ Emission\ with\ GDP - 2.412x10^{-12} Carbon\ Dioxide\ Emission\ with\ Labour\ Force\ and\ GDP$$

It could be seen from the prediction model equations for both countries that the significant contributing factors would involve environmental and economic factors which were the main variables on carbon dioxide emission and labour force. In addition, for Indonesia, the first and second order interaction variables were included, namely, variables on carbon dioxide emission with GDP and carbon dioxide emission with labour force and GDP respectively.

The MAPE values for Malaysia and Indonesia were 19.48% and 2.39% respectively. Comparing the MAPE values between both countries showed that the best model for Indonesia was the more accurate model that could be used for forecasting. This was because the MAPE value was less than 10% (Chen *et al.*, 2008). The MAPE value for the best model M6.0.0 of Malaysia

was 19.48%. It could also be used for forecasting purposes because the MAPE was less than 25%. Even though the MAPE value was greater than 25%, the model obtained could still be considered as acceptable. From model M6.0.0, when one unit of the variable  $X_1$  had increased, the value of  $Y$  would decrease by  $1.773 \times 10^{-10}$  tonnes. Similarly goes for variable  $X_3$ . When variable  $X_3$  had increased by one unit, the value of  $Y$  would increase by  $9.972 \times 10^{-12}$  tonnes. On the other hand, for Indonesia, from model M29.4.1, when one unit of  $X_{14}$  had increased, the value of  $Y$  would also increase by  $2.62 \times 10^8$  tonnes. On other hand, when the variable  $X_{134}$  had increased by one unit, the value of  $Y$  would decrease by 2.412 tonnes. What could be seen from the models of both countries, the interaction variables from model for Indonesia had accounted for the model difference and better estimation in log production.

## Conclusion

In this research, log production is able to be forecasted using multiple regression analysis for Malaysia and Indonesia, and the prediction model for Indonesia is acceptable to be excellent with a small MAPE value. The interaction variables had contributed significantly to the prediction model when compared to the main variables. The modelling procedures comprise of the various statistical tests and steps hence would enable researchers to produce a robust model for forecasting in their future work.

## Recommendation

The data used in this research were online data. It was better if original data were obtained directly from the related department. Second, it was recommended that the sample size be large for parameter estimation.

## Acknowledgement

The authors would like to extend their gratitude to Universiti Malaysia Sabah for the partial financial support in this research.

## References

- Chen, R.J.C., Peter, B., & Frederick, W.C. (2008). Comparing forecasting models in tourism. *Journal of Hospitality & Tourism Research*, 32(3): 1-20.
- Department of Statistics Malaysia. Available at: <http://www.statistics.gov.my/>
- Forestry Department of Peninsular Malaysia, Sabah and Sarawak. (2009).
- Ghasemi, A. & Zahediasl, S. (2012 Spring). Normality test for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology Metabolism*, 10(2): 486-489.
- Indonesia Forestry Department. Available at: <http://www.indonesia.go.id/en/ministries/ministers/state-minister-for-the-environment/1642-profile/266-kementerian-lingkungan-hidup.html>
- Maydeu-Olivares, A. & García-Forero, C. (2010). Goodness of fit testing. In Peterson, P., Baker, E., & McGaw, B. (Eds). *International Encyclopedia of Education* (3rd ed.) (pp. 190-196). Oxford: Elsevier
- Noraini, A., Amran, A. & Zainodin, H.J. (2011). An Improved Volumetric Estimation Using Polynomial Regression. *Journal of Science & Technology (UTM)*, 3(2):29-42.
- Noraini, A., Zainodin, H.J. & Amran, A. (2012). Sustainable urban forest using multiple regression models. *Research Journal of Forestry*, 6(1): 1-15.
- Sopipan, N., Kanjanavajee, N. & Sattayatham, P. (2012). Forecasting SET50 index with multiple regression based on principal component analysis. *Journal of Applied Finance & Banking*, 2(3): 271-294.
- Thang, H.C. (2009). Malaysia forestry outlook study. *Asia-Pacific Forestry Sector Outlook Study II*.
- The World Bank Data Bank. Available at: <http://databank.worldbank.org>
- Zainodin H.J. and Khuneswari G. (2011). Zainodin, H.J. & Khuneswari, G. 2009. A case study on determination of house selling price model using multiple regression. *Malaysian Journal of Mathematical Sciences*, 3(1): 27-44.
- Zainoddin, H.J., Noraini, A., & Yap, S.J. (2011). An alternative multicollinearity approach in solving multiple regression problem. *Trends in Applied Science Research*, 6(11): 1241-1255.

## APPENDIX

### All Possible Models

$$M1 : Y = \beta_0 + \beta_1 X_1 + u$$

$$M2 : Y = \beta_0 + \beta_2 X_2 + u$$

$$M3 : Y = \beta_0 + \beta_3 X_3 + u$$

$$M4 : Y = \beta_0 + \beta_4 X_4 + u$$

$$M5 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$



- M6 :  $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + u$
- M7 :  $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_4 + u$
- M8 :  $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + u$
- M9 :  $Y = \beta_0 + \beta_2 X_2 + \beta_4 X_4 + u$
- M10 :  $Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 + u$
- M11 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$
- M12 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + u$
- M13 :  $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4 + u$
- M14 :  $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$
- M15 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$
- M16 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_{12} + u$
- M17 :  $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_{13} X_{13} + u$
- M18 :  $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_4 + \beta_{14} X_{14} + u$
- M19 :  $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_{23} X_{23} + u$
- M20 :  $Y = \beta_0 + \beta_2 X_2 + \beta_4 X_4 + \beta_{24} X_{24} + u$
- M21 :  $Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 + \beta_{34} X_{34} + u$
- M22 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{23} X_{23} + u$
- M23 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + \beta_{12} X_{12} + \beta_{14} X_{14} + \beta_{24} X_{24} + u$
- M24 :  $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{34} X_{34} + u$
- M25 :  $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{34} X_{34} + u$
- M26 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{34} X_{34} + u$
- M27 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{23} X_{23} + \beta_{123} X_{123} + u$
- M28 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + \beta_{12} X_{12} + \beta_{14} X_{14} + \beta_{24} X_{24} + \beta_{124} X_{124} + u$
- M29 :  $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{34} X_{34} + \beta_{134} X_{134} + u$
- M30 :  $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{34} X_{34} + \beta_{234} X_{234} + u$
- M31 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{34} X_{34} + \beta_{123} X_{123} + \beta_{124} X_{124} + \beta_{134} X_{134} + \beta_{234} X_{234} + u$
- M32 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{34} X_{34} + \beta_{123} X_{123} + \beta_{124} X_{124} + \beta_{134} X_{134} + \beta_{234} X_{234} + \beta_{1234} X_{1234} + u$