IJST EXPLORE International Journal of Science and Technology
CENTRE OF PROFESSIONAL RESEARCH PUBLICATIONS
ISSN 2049-7318
International Journal of Science and Technology Volume 5 No. 3, March, 2016

Comparative Study on Predicting Crude Palm Oil Prices Using Regression and Neural Network Models

Azme Khamis and Nursu'aidah Abd. Wahab
Department of Mathematics and Statistics
Faculty of Science, Technology and Human Development
Universiti Tun Hussein Onn Malaysia

ABSTRACT

Palm oil has known as the important source of vegetables oils in the global market. Malaysia is the one of the major producer and exporters of palm oil. An accurate forecasting on crude palm oil (CPO) prices is considered significant to the oil palm business. This study was conducted to identify suitable model between Multiple Linear Regression (MLR) model and Artificial Neural Network (ANN) model on predicting Malaysia crude palm oil (CPO) prices. The Malaysia crude palm oil was predicted by three other Malaysia primary commodity prices which are natural rubber (NR) prices, black pepper (BP) prices and cocoa beans (CB) prices. The analysis use weekly data on the prices from Jan 2004 until Dec 2013. The methods are compared to obtain the best model for predicting crude palm oil price. It was found that, the value of R^2 in ANN model is higher than MLR model by 20.61%. The value of mean squared error (MSE) in ANN model also lower compared to MLR model. Therefore, ANN model is preferred to be used as alternative model in estimating crude palm oil (CPO) prices compared to MLR model.

Keywords: *Multiple linear regression, neural network model, crude palm oil prices*

1. INTRODUCTION

Palm oil has known as the important source of vegetables oils in the global market. World Growth (2010) had stated that palm oil is the most important tropical vegetable oils in the global market for oils and fats markets, as measured by either production or international trade. Production of palm oil is more sustainable than other vegetables oils since it consumes considerable less energy in production due to long productive lifespan of 25 years, uses less land in terms of broad-acre cultivation method and generates more oil per hectares compares to other oils. Ming and Chandaramohan, (2002) showed that the productivity per unit area of palm oil is higher compared to any other crop. The study by Craven (2010) had revealed that Malaysia is the leading supplier of palm oil in the global market. Malaysia was the major producer and exporters of palm oil since 1960 until it has been surpassed by Indonesia due to limitation of land for more growth (Mahat, 2012). As one of the major crude palm oil producers, it is importance to know the future crude palm oil price which can lead to significant impact to Malaysia's economy.

The appropriate model should be considered in order to provide reliable forecast for CPO which can be a useful for policy makers. According to Sallehuddin *et al.* (2009) forecasting approach can be divided into two categories which are statistical and artificial intelligence (AI) based techniques. The artificial intelligence for forecasting approaches is important area for the time series forecasting. Artificial Neural Networks (ANN) model have been viewed as simplified models of neural processing which resemble the network process in brain. In more practical terms neural networks are non-linear statistical modeling or decision making tools. Moreover, ANN is a stable model for prediction (Duy *et al.*, 2009). The study by Rojas *et al*., (2008) and Sallehuddin *et al.* (2009) had indicated that ANN model is capable to predict nonlinear time series data. The evidence found by Maier and Dandy (1996) which recommends that the ANN model is better for the long term forecasting. Recent study shows a great extension in application of neural network approach in predicting prices (Sureshkumar and Elango, 2012). For example the study by Ernest (2002) had use ANN model to predict CBOT soybean prices. Mombeini (2015) propose ANN model perform better in predicting gold prices. There are also studies on predicting chicken prices in Iran using ANN model by Somayeh *et al.* (2012). And lastly the study by John Wei *et al*. (2012) had use ANN model to predict crude oil prices.

As our interest is toward CPO prices, so there is also some study which had apply ANN model to predict the CPO prices. Silalahi (2013) had applied the genetic algorithm neural network (GANN) to forecast the international price of crude palm oil (CPO) and soybean oil (SBO) and the result indicated that GANN is the great problem solving ability for both of data. Another study by Gunawan *et al.* (2013) had proposes the best neural network model to predict CPO prices are consists of joint network topology and regular normalization in momentum of 0.75, learning rate of 0.05 and minimum of 50000 iterations.

Malaysia's Agriculture primary commodities are consisting of palm oil, natural rubber, cocoa beans and pepper. Although there are different between commodities, Malaysia's share in the international market of world production and export of

these raw materials tend to provide important role for the country (Arshad and Ghafar, 1998). There is some study that had interest to identify the trend of commodity price involving agricultural markets. Baffes and Gohou (2001), shows that there is a strong co-movement between polyester and cotton prices. Another study by Chen *et al.* (2010) had found that there are significant relationship between the crude oil price and grain prices. Hairi *et al.* (2009) had proposed that there are linked within commodity prices of oil for corn, cotton, and soybeans, but not for wheat.

This study was intended to predicting crude palm oil price and how it was affected by other Malaysia primary commodity prices. MLR and ANN model will be used to predict the crude palm oil (CPO) prices. There were 524 weekly data of crude palm oil (CPO) prices that was taken out from Jan 2004 until Dec 2013. The primary commodity price data that will be used consists of natural rubber (NR) prices, black pepper (BP) prices and cocoa beans (CB) prices. The relationship between crude palm oil (CPO) prices and other Malaysia primary commodity prices are also investigated. Comparatives study between MLR model and ANN model will be carried out to determine which model is best model to forecast crude palm oil (CPO) prices.

2. RESEARCH METHODOLOGY

2.1. Multiple Linear Regressions

Regression is a fundamental operation in statistics and includes techniques for modeling and analyzing several variables at a time. Regression analysis is used for explaining the relationship between a dependent variable and a number of independent variables. The independent variables are also known as predictor or explanatory variables. In most regression analyses, the variables are assumed to be continuous. In simple regression, there is only one independent variable. However, most real world applications involve more than one variable which influence the outcome variable. The model for Multiple Linear Regression can be represented as (Michael et al., 2008):

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1} \qquad (1)$$

where β_0 is called intercept and $\beta_1, \beta_2, \beta_{p-1}$ are called slopes or regression coefficients. The difference between the predicted and the actual value of Y is called the error (ε) or can be written as $\varepsilon = Y_i - E\{Y\}$. Then, regression equation can be express as:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \qquad (2)$$

where Y_i is the actual value and ε_i is the error for the *i*th observation while $X_{i,1}, X_{i,2}, X_{i,p-1}$ are known constants.. The main assumptions for the errors ε_i is that $E(\varepsilon_i) = 0$ and $var\ (\varepsilon_i) = \sigma^2$. Also the ε_i are randomly distributed. The predicted value is also denoted by $\hat{Y}$. The various errors are given as:

$SSE = \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2$; $SST = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$

and $SSR = \sum_{i=1}^{n} (\bar{Y} - \hat{Y}_i)^2$

Where *SSE* is the sum of squares of error, *SSR* is the sum of squares of regression, and *SST* is the sum of squares total.

R^2 is the square of the correlation between the response values and the predicted response values. It is also called the square of the multiple correlation coefficients and the coefficient of multiple determinations. The coefficient of determination is the overall measure of the usefulness of a regression. It is given as, $R^2 = 1 - \frac{SSE}{SST}$

The value of R^2 can range between 0 and 1, a higher value indicates a better model. In terms of the sample, the estimate of the population total variance (*SST*) is denoted by Mean Sum of Squares Total (*MST*) and is calculated as; $MST = {SST}/{(n-1)}$

Where *n* is the sample size. Similarly, the estimated residual or error is called Mean Squared Error (*MSE*) and is calculated as, $MSE = {SSE}/{(n-p-1)}$

where *n* is the sample size, and *p* is the number of exploratory variables. A better estimate of the coefficient of determination is made by the Adjusted-R squared statistic:

$$R^2_{Adj} = 1 - \frac{{SSE}/{(n-p-1)}}{{SST}/{(n-1)}} = \frac{MSE}{MST}$$

The *F-test* in one way Analysis of Variance (ANOVA) is also used as a statistic to find the goodness of fit of the model. It is calculated as:

$$Ftest = \frac{\text{expalined variance}}{\text{unexplained variance}} = \frac{{SSR}/{p}}{{SSE}/{(n-p-1)}}$$

2.2. Artificial Neural Network Model

Artificial neural networks are nonlinear mapping systems with a structure loosely based on principles observed in biological nervous systems. Artificial neural network (ANN) offers many advantages over conventional statistical methods (Shachmurove, 2002).

Firstly, the data have to undergo preprocessing step. The neural network training can be made more efficient if certain preprocessing steps on the network inputs and targets are performed. The normalization of the input and target values mean to mapping them into the interval [-1, 1]. This simplifies the problem of the outliers for the network. The normalized inputs and targets that are returned will all fall in the interval [-1, 1].

The total number of input neurons for the network was set into three variables input data which they are black pepper (BP) prices, natural rubber (NR) prices and cocoa bean (CB) prices while crude palm oil (CPO) prices was set as target

data. The data was randomly divided up the 100% into 70% for training, 15% for validation and 15% for testing. Figure 1 show the example of three layer feed forward network architectures with one output neuron were used throughout this study. It has been proven in many studies, for example, Fu (1994) and Masters (1995) that using a network with only one hidden layer is sufficient to approximate any continuous nonlinear function, thus a network with one hidden layer was used in this study.

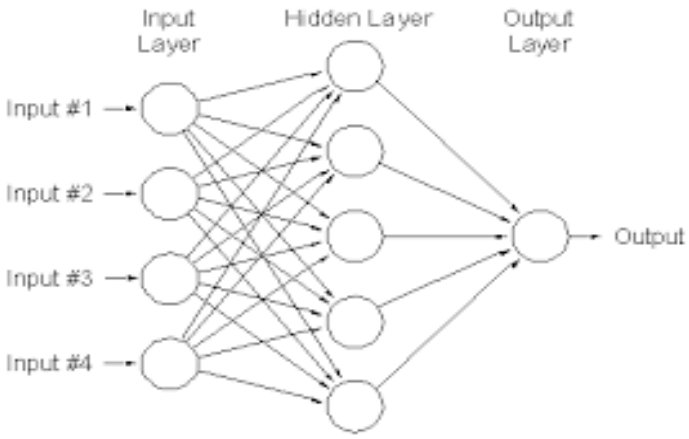

Figure 1: Three Layer Feed Forward Network architectures

Decisions about the optimum number of neurons in the hidden layer were obtained by trial and error, where the number of neurons in the hidden layer varied from 1 to 15. Each network was trained with different initial weights. The weights were initialized randomly using a built-in MATLAB function .The maximum number of epochs to train the ANN was set at 10,000. This was to ensure that the network was sufficiently trained. The network will be train by using Levenberg Marquat algorithm. According to Jianchao and Chern (2001), Levenberg Marquat algorithm is the fastest and most widely used for neural network training. Levenberg Marquat algorithm is a combination of two algorithms, the steepest descent with similar formulation to gradient descent and the Gauss-Newton. Gauss-Newton is a second order error minimization method that uses the information of the curvature of the error surface.

3. DATA ANALYSIS AND RESULTS

3.1. Data Preparation

The data for this study was obtained from a different source including MPOB, MRB, MCB, and MPB official website for Malaysia primary commodity price. The data consist of prices of crude palm oil (CPO), natural rubber (NR), black pepper (BP) and cocoa beans (CB) throughout Jan 2004 until Dec 2013. A total of 524 weekly data points was used for this study.

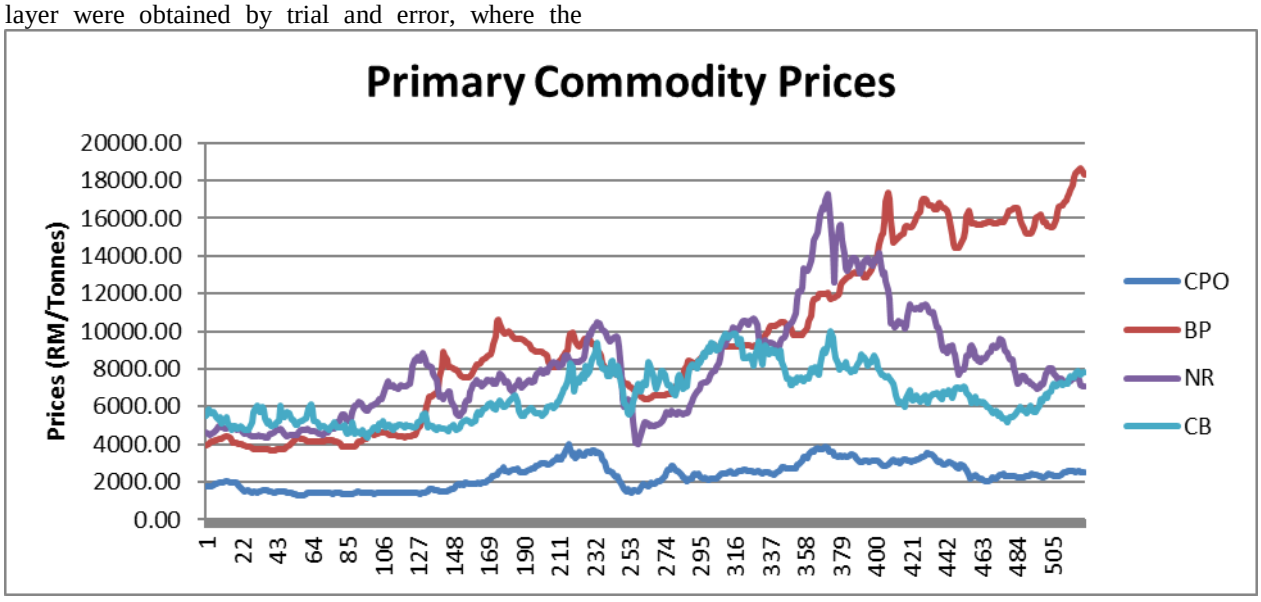

Figure 2: Price of Malaysia's Primary Commodity Prices from Jan 2004-Dec 2003

3.2. Regression Analysis

The relationship between dependent variable with independent variables was performed using Pearson correlation. It was found that the correlation coefficient of CPO prices is positive and strongly significance to BP prices (0.637), NR prices (0.783) and CB prices (0.642).

The value of *F* statistic is 411.328 and *p*-value is 0.000 indicate that the model is suitable and can be fitted to the data. The coefficient of determination R^2 = 0.704, it shows that 70.4% variance in CPO price can be explained by PB prices, NR prices and CB prices. The regression equation for the CPO prices can be written as;

$$CPO = 134.789 + 0.042 * PB + 0.120 * NR + 0.127 * CB$$

3.3. Artificial Neural Network Model

The number of neurons in the hidden layer is determined by trial and error method. The trials initialize at error with 2 nodes first. Then, the process is repeated until 15 nodes were used. A comparison of the MSE value for all number of nodes was carried out. The lowest MSE value will be selected as optimum number of nodes in hidden layer. The network architecture of the ANN model has been determined iteratively. Based on Table 3, the optimum number nodes in the hidden layer are 13.

Table 3: Optimum number of nodes in hidden layer

Number of nodes	MSE	R value for training process	R value for validation process	R value for testing process
2	0.0553	0.8907	0.8752	8870
3	0.0512	0.9000	8955	9032
4	0.0543	0.9051	8732	9016
5	0.0381	0.9280	9127	9300
6	0.0401	0.9200	9096	9125
7	0.0396	0.9250	9159	9287
8	0.0328	0.9356	9303	9273
9	0.0365	0.9401	9202	9265
10	0.0337	0.9317	9337	9377
11	0.0339	0.9383	9357	9314
12	0.0330	0.9391	9088	9313
13	**0.0298**	**0.9504**	**0.8934**	**0.9215**
14	0.0315	0.9467	8960	9170
15	0.0312	0.9453	8976	9160

The architecture in Figure 5 shows that there are 3 input layers, 1 hidden layer consists of 13 nodes and 1 output layer. Figure 6 shows retrained performance graph of neural network model that had been created during its training. The training stopped after 13 epochs because the validation error increased and performance MSE obtained is 0.048825. It is a useful diagnostic tool to plot the training, validation, and test errors to check the progress of training. The result shows a good network performance because the test set error and the validation set error have similar characteristics, and it doesn't appear that any significance over fitting has occurred.

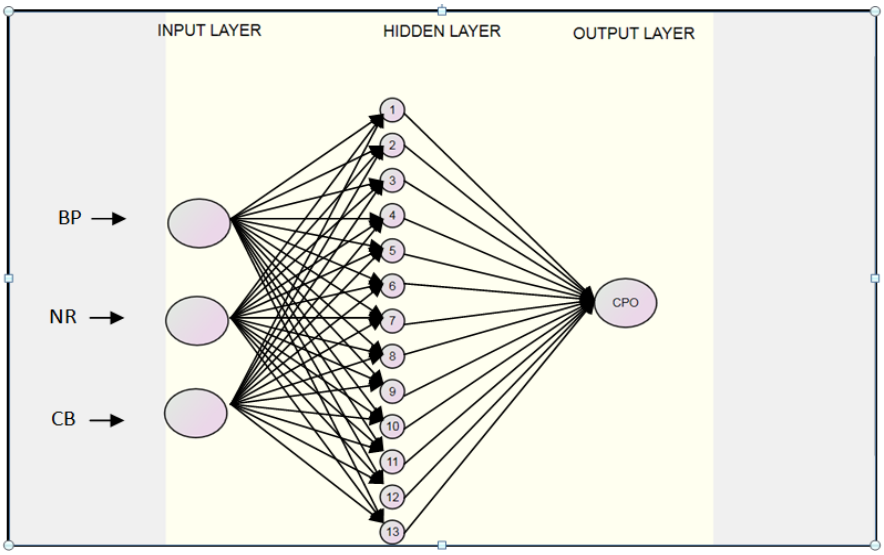

Figure 5: Architecture of Artificial Neural Network with Three Layers

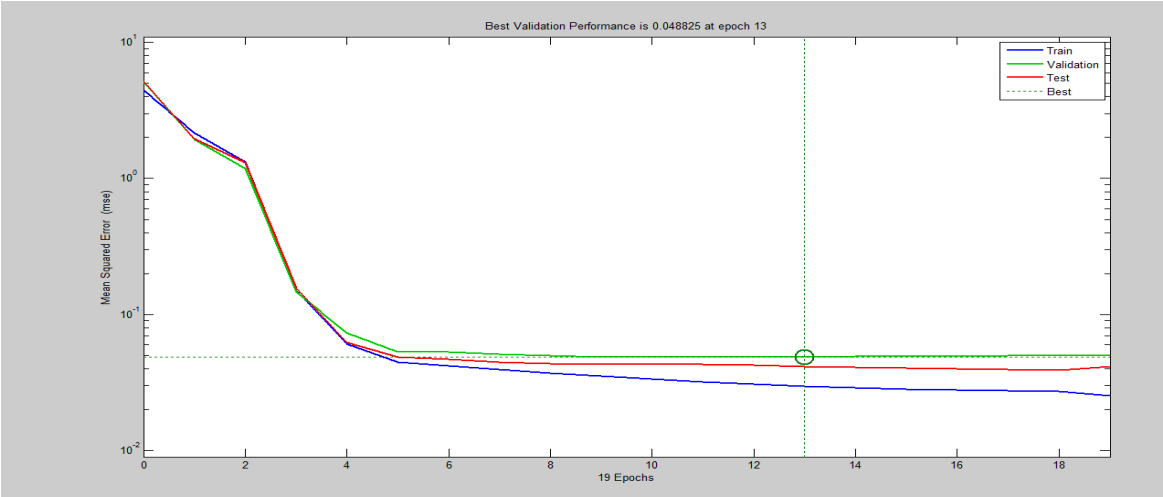

Figure 6: Performance Plot

Based on Figure 7, the values of *R* are above 0.9 during training of network. This shows that the output produced by the network is closely similar to the target and that the model is satisfactory.

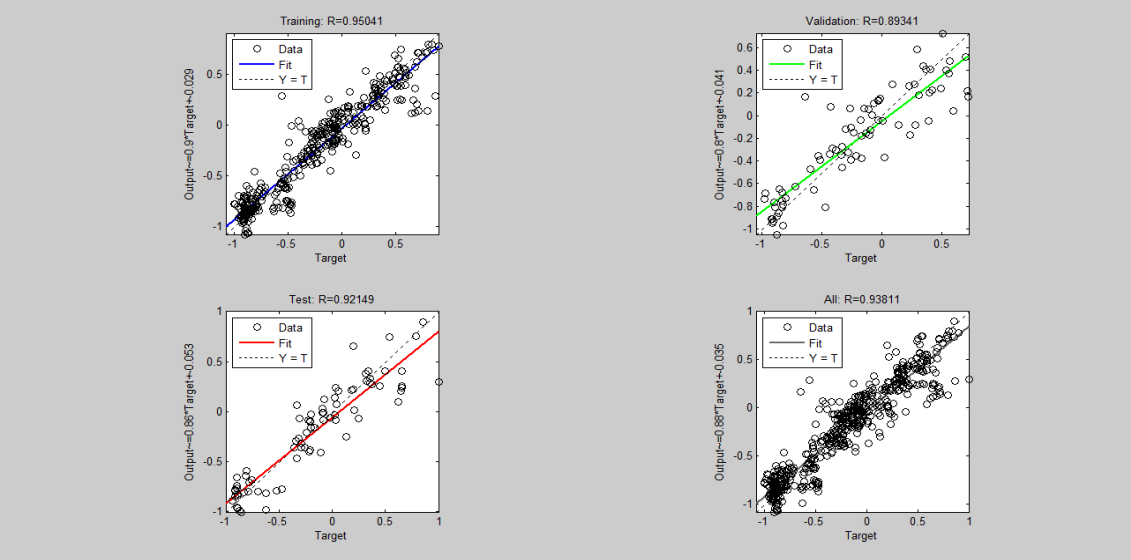

Figure 7: Regression Plots

4. MODEL COMPARISON

Based on the results as provided on Table 4, the value of R^2 for ANN model shows that 91.01 % variance in CPO prices can be explained by PB prices, NR prices and CB prices. The result also show that the R^2 value for ANN model is higher compared to MLR model and the value of MSE in ANN model is lower compared to MLR model. Therefore, the ANN model was selected as best model to predict CPO prices.

Table 4: Model Selection Criteria

	MLR	ANN
MSE	148019.89	65162.71
R^2	0.7040	0.9101
R	0.8390	0.9504

5. CONCLUSION AND RECOMMENDATION

The model's accuracy in predicting crude palm oil (CPO) price was measured by a number of criteria. The value of R^2 and MSE were compared to select preferred model. By using ANN model, the R^2 value was increase about 20.61% higher than MLR model. It can be concluding artificial neural network (ANN) model is preferred to predict CPO prices compared to MLR model. Besides, the chosen model can be used as an alternative way to estimate the future price of crude palm oil (CPO) prices.

6. REFERENCES

Arshad, F.M. & Ghaffar, R.A. (1988). Malaysia's Primary Commodities: Constant-Market-Share Analysis. *Malaysian Agricultural Policy: Issues and Directions*, Universiti Putra Malaysia, Serdang, Selangor: pp. 197-216.

Baffes, J. & Gohou, G. (2005). The Co-movement between Cotton and Polyester Prices. *World Bank Policy Research Working Paper* 3534. World Bank.

Chen, S.T., Kuo. H.I., & Chen, C.C. (2010). Modeling the relationship between the oil price and global food prices, *Applied Energy*, p. 2517-2525.

Craven, C. (2010). The Honduran Palm Oil Industry: Employing Lessons from Malaysia in the search for economically and environmentally sustainable energy solution", *Energy Policy*.

Duy, T. V. T. Sato Y. & Inoguchi, Y. (2009). Improving Accuracy of Host Load Predictions on Computational Grids by Artificial Neural Networks, *IEEE International Sym- posium on Parallel & Distributed Processing* (IPDPS 2009), Rome, 23-29 May 2009, pp. 23-29. doi:10.1109/IPDPS.2009.5160878

Ernest A. Foster, (2002) Commodity Futures Price Prediction, an Artificial Intelligence Approach, Department of Science, University of Georgia, Athens, 2002.

Fu, L. (1994). *Neural Networks in Computer Intelligence*. New York McGraw-Hill, Inc.

Gunawan, R. Leylia, M. & Harlili (2013). Commodity Price Prediction Using Neural Network, *2013 International Conference on Computer, Control, Informatics and Its Applications*,243–248.

Harri, A., Nalley, L. Hudson, D. (2009). The Relationship Between Oil, Exchange Rates, and Commodity Prices. *Journal of Agricultural and Applied Economics* 41(2): 501-510.

Jianchao, Y., & Chern, C. T. (2001). Comparison of Newton-Gauss with Levenberg-Marquardt Algorithm for Space Resection. Paper presented at the *22nd Asian Conference on Remote Sensing.*

Mahat, S. B. A. M. (2012). The Palm Oil Industry From The Perspective of Sustainable Development : A Case Study of Malaysian Palm Oil Industry, (September).

Maier H. R. & Dandy, G. C. (1996). Neural Network Models for Forecasting Univariate Time Series, *Water Resource Research*, Vol. 32, No. 4, 1996, pp. 1013-1022.

Masters, T. (1995). *Neural, novel & hybrid algorithms for time series prediction*. New York John Wiley and Sons.

Michael H. K., Christopher J. N., & John N. (2008). *Applied Linear Regression Models*. Fourth Edition. New York: McGraw-Hill/Irwin

Ming, K. K. & Chandramohan, D. (2002). Malaysian Palm Oil Industry at Crossroads and its Future Direction. *World*, *2*(2), 2–7.

Mombeini, H., & Yazdani-Chamzini, A. (2015). Modeling Gold Price via Artificial Neural Network. *Journal of Economics, Business and Management*, 3(7), 699–703. http://doi.org/10.7763/JOEBM.2015.V3.269

Rojas, I. Valenzuela, O. Rojas, F. Guillen, A. Herrera, L. J. Pomares, H. Marquez L. & Pasadas, M. (2008). Soft-Computing Techniques and ARMA Model for Time Series Prediction, *Neurocomputing*, Vol. 71, No. 4-6, 2008, pp. 519-537.

Sallehuddin, R. Shamsuddin, S. M. Hashim S. Z. & Abraham, A. (2009). Forecasting Time Series Data Using Hybrid Grey Relational Artificial Neural Network and Autore-gressive Integrated Moving Average Model, *Neural Network World*, Vol. 6, No. 7, 2009, pp. 573-605.

Shan Hu John Wei, Chung Hu Yi, & Wen Lin Ricky Ray, (2012). Applying Neural Networks to Prices Prediction of Crude Oil Futures, *Mathematical Problems in Engineering*, vol. 2012

Shachmurove, Y.,(2002). Applying Artificial Neural Networks to Business, Economics and Finance," Departments of Economics, The City College of the City University of

New York, 2002, available online from http://www.econ.upenn.edu/Centers/CARESS/CARESSpdf/02-08.pdf

Silalahi, D. D (2013). Application of Neural Network Model with genetic Algorithm to Predict the International Prices of Crude Palm Oil (CPO) and Soybean Oil (SBO), *12th National Convention on Statistics* (NCS) EDSA Shangri-La Hotel, Mandaluyong City, October 1-2, 2013Page 1 of 12, 1–12.

Sureshkumar K. K. & Elango, N. M. (2012). Performance Analysis of Stock Price Prediction using Artificial Neural Network, *Global Journal of Computer Science and Technology*, vol. 12, no. 1, pp. 19-25, January 2012.

Somayeh Ebrahimi, Shahrokh Shajari, & Mohammad Hassan Tarazkar, (2012). Prediction of Agricultural Commodity Price Using Artificial Neural Networks: Case of Chicken Price in Fars province, Iran, *Journal of Basic and Applied Scientific Research*, vol. 2, no. 11, pp. 11537-11541, 2012.

World Growth (2010). Palm Oil and Food Security: The Impediment of Land Supply. *World Growth Organization* December 2010.