The International Journal Of Engineering And Science (IJES)
|| Volume || 3 || Issue || 11 || Pages || 19-26 || 2014 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Forecasting Wheat Price Using Backpropagation And NARX Neural Network

Azme Khamis[1] and Siti Nabilah Syuhada Binti Abdullah[2]
Department of Mathematics and Statistics
Faculty of Science, Technology and Human Developtment,
Universiti Tun Hussein Onn Malaysia

--ABSTRACT--
This study aims to investigate suitable model and forecast future wheat price using backpropagation neural network (BPNN) and nonlinear autoregressive models with exogenous inputs (NARX) networks. The price of wheat was estimated using prices of 3 types of grains widely used in agriculture which are oats, barley and soybeans. The analysis was conducted using historical data on the prices from 1978 until 2012. Artificial neural networks are non-parametric statistical estimators, and thus have potential for use in estimation modeling. The variables selection criteria procedures are also developed to select a significance explanatory variable. The methods are compared to obtain the best model for predicting wheat price. Based on the results, the NARX model with 8 nodes in hidden layer and 4 tapped delay lines can be used as an alternative model predict wheat price.

KEYWORDS: *artificial neural networks, price estimation, BPNN, NARX*

Date of Submission: 31 October 2014 — Date of Accepted: 10 November 2014

I. INTRODUCTION

Artificial neural network (ANN) model have an ability to recognize time series patterns and nonlinear characteristics, which gives better accuracy over the others methods, it become most popular methods in making prediction (Vaziri, 1997; Sharda, 1994; Jones, 2004; Toriman *et al*., 2009). A comparative study were made using ANN and conventional Auto-Regression (AR) model networks in forecasting the river flow for two well known River in USA and they found that ANN performed better than AR model. It has been reported that the results using Radial Basis Function (RBF) neural network is better than Back Propagation Neural Network (BPNN) in modeling a meteorological problems such as weather forecasting. In addition, there was analysis shows that pre-processing data analysis also can influenced the performance of prediction model (Zhang, 2002; Suguna and Thanuskodi, 2011). From previous work, the method of BPNN is better compared to SARIMA in obtaining water level prediction at Dungun River, Terengganu (Arbain and Wibowo, 2012). This study was carried out on predicting wheat price and how it was affected by other grains prices such as soybean, oats and barley. Two types of ANN which are Backpropagation Neural Network (BPNN) and nonlinear autoregressive models with exogenous inputs (NARX) network will be used to obtain prediction of wheat price. The relationship between wheat price with other grain product such as soybean, oats and barley are also investigated. Comparatives study will be carried out to determine which model is best suited to model and forecast wheat price.

II. RESEARCH METHODOLOGY

2.1 Backpropagation Neural Network (BPNN) : Backpropagation is the most widely used learning algorithm and is a popular technique because it is easy to implement. It requires data for conditioning the network before using it for predicting the output. Training a network by backpropagation involves three stages: the feed forward of the input training pattern, the backpropagation of the associated error, and the adjustment of the weights (Laurene, 1994).In BPNN network, there are four important considerations are comprised in network designs which are the network architecture determination, hidden neuron number determination, activation function optimization and training algorithm optimization (Hagan *et al*., 1996). As shown in Figure 1, the network consists of three layers. The first layer, which is the input layer, is triggered using the activation function whereas the second layer is hidden layer and third layer is the output layer. A network of these two transfer function can be trained to approximate any function.

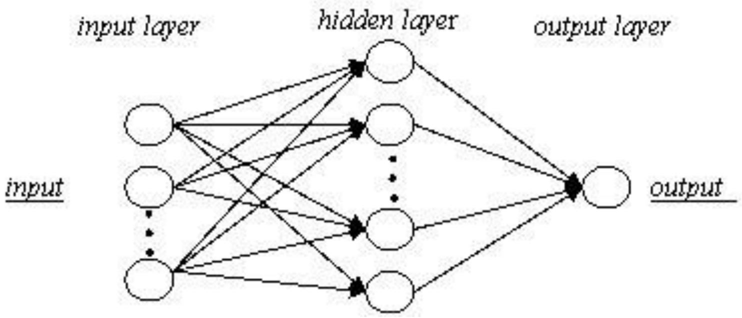

Figure 1. General BPNN with input, hidden and output layer

The network is trained using Levenberg-Marquardt algorithm. In the case of supervised learning, the network is presented with both the input data and the target data called the training set. Externally provided correct patterns are compared with the ANN output during training and feedback is used to adjust the weights until all training patterns are correctly categorized by the network.

Nonlinear Autoregressive Network with Exogenous Inputs (NARX) : Nonlinear autoregressive models with exogenous inputs (NARX) recurrent neural architectures (Chen, *et. al*., 1990; Narendra and Parthasarathy,1990) as opposed to other recurrent neural models, have limited feedback architectures that come only from the output neuron instead of from hidden neurons. The NARX models are commonly used in the system of identification area (Xie *et al*., 2009). All the specific dynamic networks discussed so far have either been focused networks, with the dynamics only at the input layer, or feed forward networks. The nonlinear autoregressive network with exogenous inputs (NARX) is a recurrent dynamic network, with feedback connections enclosing several layers of the network. The NARX model is based on the linear ARX model, which is commonly used in time-series modeling.

NARX is an important class of time series nonlinear systems that can be written as

$$y(t+1) = f\,[y(t),...,x(t-d_y+1);\ u(t-k),\ u(t-k-1),...,\ u(t-k-d_u+1)], \qquad (1)$$

where $u(n) \in \boldsymbol{R}$ and $y(n) \in \boldsymbol{R}$ denote, respectively, the input and output of the model at discrete time step n, while $d_u \geq 1$, $d_y \geq 1$ and $d_u \leq d_x$, are the input-memory and output-memory orders, respectively. The parameter k ($k \geq 0$) is a delay term, known as the process dead-time. Without lack of generality, we always assume $k = 0$ in this paper, thus obtaining the following NARX model:

$$y(t+1) = f[y(t),...,\ y(t-d_y+1);\ u(t),u(t-1),...,u(t-d_u+1)], \qquad (2)$$

which may be written in vector form as

$$y(t+1) = f\,[\boldsymbol{y}(t);\ \boldsymbol{u}(t)], \qquad (3)$$

where the vectors $\boldsymbol{y}(t)$ and $\boldsymbol{u}(t)$ denote the output and input regressors, respectively.

The nonlinear mapping $f(\cdot)$ is generally unknown and can be approximated, for example, by a standard multilayer perceptron (MLP) network. The resulting connectionist architecture is then called a NARX network, a powerful class of dynamical models which has been shown to be computationally equivalent to Turing machines (Siegelmann, *et. al.* 1997). Figure 2 shows the topology of a three-hidden-layer NARX network, which are input layer, hidden layer and output layer.

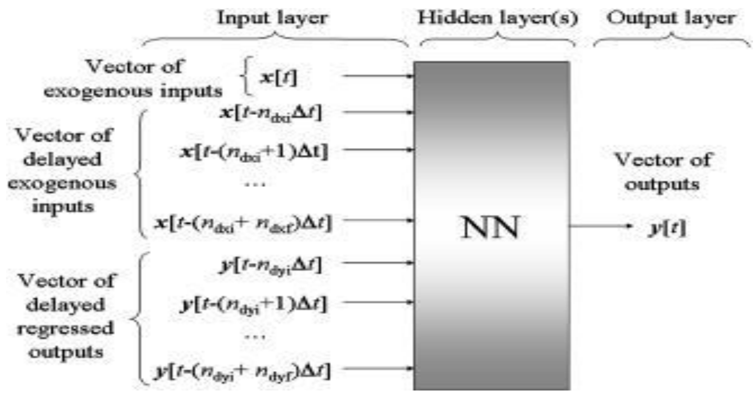

Figure 1. NARX network with three hidden layers

Takens (1981) has shown that, under very general conditions, the state of a deterministic dynamic system can be accurately reconstructed by a time window of finite length sliding over the observed time series as follows:

$$y(t+1) = f[\mathbf{y}(t); \mathbf{u}(t)], \quad (4)$$

where $x(t)$ is the sample value of the time series at time n, dE is the embedding dimension and τ is the embedding delay. Eq. (4) implements the delay embedding theorem (Kantz and Schreiber, 2006). According to this theorem, a collection of time-lagged values in a dE-dimensional vector space should provide sufficient information to reconstruct the states of an observable dynamical system.

In order to use the full computational abilities of the NARX network for nonlinear time series prediction, we propose novel definitions for its input and output regressors. The input signal regressor, denoted by $\mathbf{u}(t)$, is defined by the delay embedding coordinates of Eq. (4):

$$\mathbf{u}(t) = \mathbf{x}1(t) = \{x(t), x(t-\tau),\ldots, x(t-(d_E-1)\tau)\}, \quad (5)$$

where we set $du = dE$. The input signal regressor $\mathbf{u}(t)$ is composed of dE actual values of the observed time series, separated from each other of τ time steps. The output signal regressor $\mathbf{y}(t)$ can be written accordingly as

$$\mathbf{y}_*(t) = \{x(t),\ldots, x(t-d_y-1)\}, \quad (6)$$

Then, the NARX networks implement following predictive mappings:

$$\hat{x}(t+1) = \hat{g}\,[\mathbf{y}_*(t), \mathbf{u}(t)] = \hat{g}\,[\mathbf{y}_{sp}(t), \mathbf{x}_1(t)], \quad (7)$$

where the nonlinear function $\hat{g}(.)$ is readily implemented through an multilayer perceptron trained with plain backpropagation algorithm.

In addition, during training, the inputs to the feed forward network are just the real/true ones – not estimated ones, and the training process will be more accurate. The network training function updates the weight and bias values according to Levenberg-Marquardt optimization. In general, in function approximation problems, for networks that contain up to a few hundred weights, the Levenberg-Marquardt algorithm will have the fastest convergence. This advantage is especially noticeable if very accurate training is required. However, as the number of weights in the network increases, the advantage of this algorithm decreases.

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 equation used to transform the data into a more centralized set of normalized data, with zero being the central point is as follows:

$$x_{i_{-1\ to\ 1}} = \frac{x_i - \left(\frac{x_{Max} + x_{Min}}{2}\right)}{\left(\frac{x_{Max} - x_{Min}}{2}\right)}$$

The selection of appropriate predictors is one of the most important steps in enhancement of accuracy in predicting wheat price. The predictors are chosen based on correlation analysis. Correlation is a statistical technique that can show whether and how strongly pairs of variables are related (Creative Research Systems, 2013).

The evaluation of performance is essential with the purpose of finding the best neural network architecture, which gives the most reliable and accurate predictions. The network performance function can be measured using mean square error (MSE) for training and testing (Hagan *et al*., 1996; Muhammad, 2009) and can be defined as, MSE = $\frac{SSE}{df}$, where SSE is sum of squares error and *df* is degree of freedom. The neural network model with the smallest MSE value is considered to be the best neural network model.

III. DATA ANALYSIS

This study describes comparison of the modeling method of BPNN and NARX. The models are constructed using MATLAB neural network toolbox. In the network modeling, out of the 409 data points, 70% was used for training, 15% for validation and another 15% for testing.The data for this research was obtained from a data repository blog, “Understanding Dairy Markets” maintained by Prof. Brian W. Gould of the Dept. of Agric. and Applied Economics (Gould, 2013). The data consist of prices of wheat, barley, oats and soybeans throughout May 1978 until May 2012. A total of 409 data points was used for this study. In the following parts, the wheat price is regarded as *y(t)* and oats, barley and soybeans prices are the $x_i(t)$.

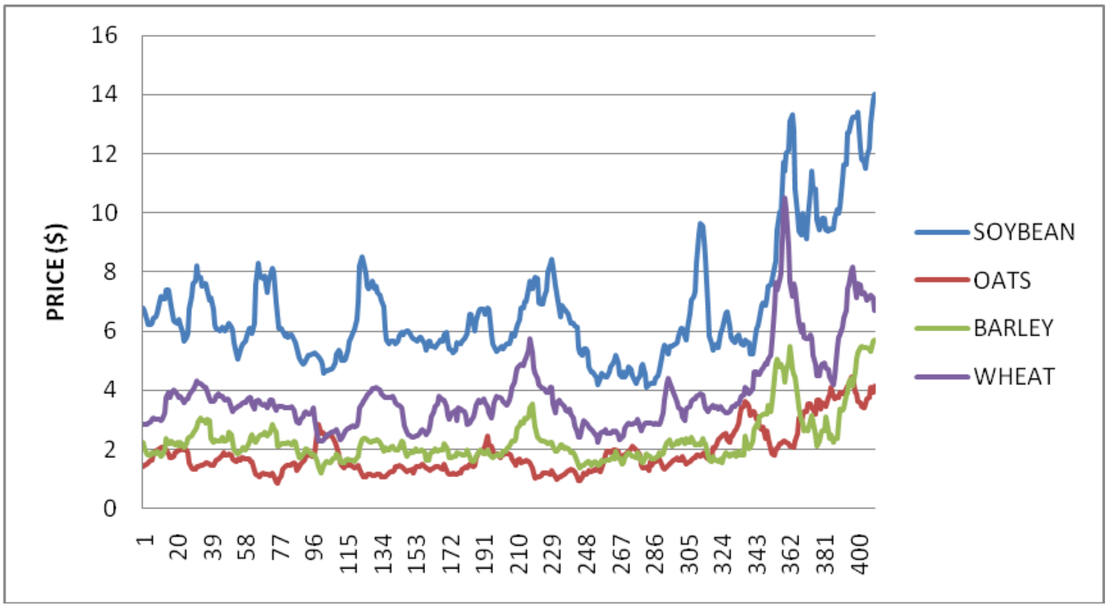

Figure 3: Price of Grain from May 1978-May 2012

Correlation Analysis : Correlation measure the strength of the linear relationship between independent and dependent variable. In this study, correlation coefficient (*r*) is measured by Pearson correlation. The value always lies in the range -1 to 1. A negative value of r indicates an inverse relationship which means that as one gets larger, the other gets smaller; whereas, a positive value of r indicates a direct relationship which means that as one variable gets larger the other gets larger. Furthermore, a zero value of r indicates that there is no relationship between the variables. The closer r is to +1 and -1, the stronger the relationship between two variables. The correlation analysis has shown that wheat price have a positive and significant to soybean (0.853), oats (0.554) and barley (0.913) prices respectively.

Back Propagation Neural Network (BPNN) Method : BPNN is performed in this experiment to build the model for wheat price. In the study, two types of sigmoid activation functions are selected which are logarithmic sigmoid function (*logsig*) and hyperbolic tangent sigmoid function (*tansig*). According to Zhang and Wu, both sigmoid functions are often used in hidden layer due to the ability of authoritative nonlinear approach (Zhang and Wu, 2009). The training algorithm used here is *trainlm* function that modified bias and weight values based on Lavenberg-Marquardt optimization. This part presents the results of the experiment using BPNN.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. Based on Table 2, the lowest MSE value is 0.006 with 11 nodes in hidden layer. This result is also supported by the *R* value which is the highest (0.972) at 11 nodes. The *R* value measures correlation between Outputs and Targets. An *R* value close to 1 means a close relationship while vice versa indicates a random relationship. Hence, 11 nodes are selected as optimum number of nodes in hidden layer.

Table 2: BPNN Model Selection

No. of nodes	MSE	*R*	No. of nodes	MSE	*R*
2	0.110	0.948	9	0.008	0.964
3	0.124	0.942	10	0.008	0.962
4	0.008	0.96	**11**	**0.006**	**0.972**
5	0.009	0.956	12	0.011	0.954
6	0.009	0.956	13	0.011	0.952
7	0.101	0.953	14	0.009	0.959
8	0.009	0.957	15	0.007	0.967

The architecture in Figure 4 shows that there are 3 input layers, 1 hidden layer consists of 11 nodes and 1 output layer. The inputs are barley, oats and soybean prices and the wheat price as an output.

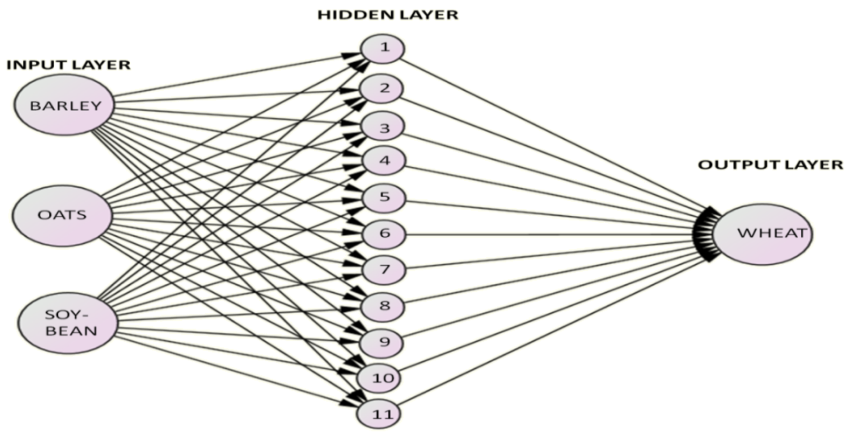

Figure 4: BPNN Model Architecture

Figure 5 shows retrained performance (MSE) graph of neural network model, created during its training. The training stopped after 12 epochs because the validation error increased. 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. After initial training of neural network model, it is retrained for 12 epochs and performance MSE is obtained 0.0173 in training

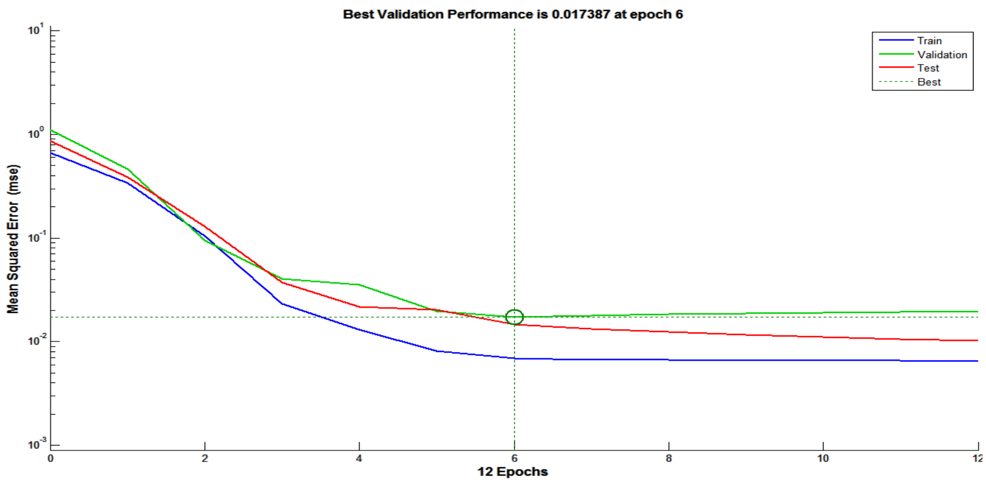

Figure 5: Performance Plot

Based on Figure 6, the value for *R* is 0.9549 This shows that the output produced by the netwrok is closely similar to the target and that the model is satisfactory.

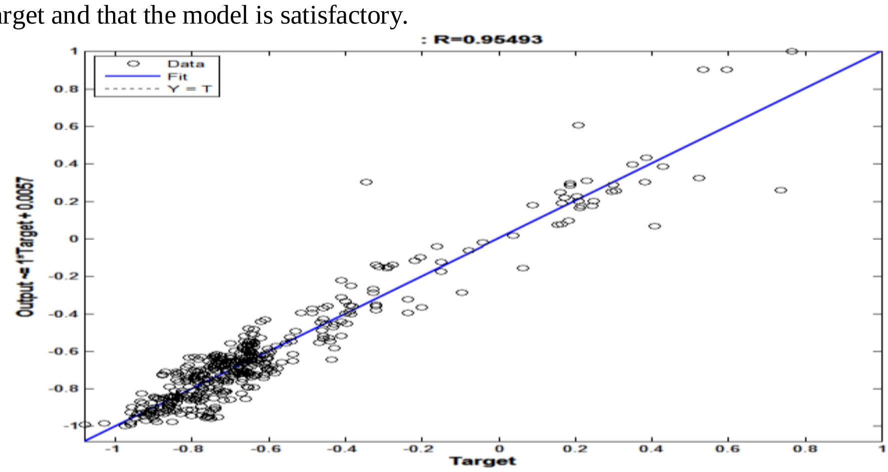

Figure 6: Correlation plot predicted versus actual observation

NARX method : In this study, nonlinear autoregressive with exogenous (external) input or NARX will be used in order to make future prediction values of a time series, *y (t)*, from past values of that time series and past values of other time series, x_i *(t)*. In these experiments, we also conducted NARX with variety numbers of hidden layers, numbers of tapped delay lines (*d*) and one output neuron with two layer feed forward networks (hyperbolic tangent transfer function in the hidden layer and linear transfer function in the output layer) were used in these experiments. The standard Lavenberg-marquardt backpropagation algorithm is used to train the network with learning rate equal to 0.001. Training automatically stops when generalization stops improving, as indicated by an increase in the Mean Square Error (MSE) of the validation samples.

The model specification was conducted by experimenting with number of nodes and numbers of tapped delay lines. 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. On the other hand, the number of tapped delay lines was initialized at 2 steps up until 4 steps. A comparison of the MSE value for all networks was carried out. The lowest MSE value will be selected as optimum network. Based on Table 3, the lowest MSE value is 0.0047 with 8 nodes in hidden layer and 4 tapped delay lines. This result is also supported by a high *R* value (0.9728) at 8 nodes that shows a close relationship between Output and Target. Hence, 8 hidden nodes with 4 tapped delay lines are selected as optimum network.

Table 3: NARX model Selection

	No. of delay				No. of delay		
No. of nodes	2	3	4	No. of nodes	2	3	4
2	0.012	0.0117	0.0114	9	0.0083	0.0069	0.0048
3	0.0101	0.012	0.0085	10	0.0089	0.0079	0.0098
4	0.0115	0.0088	0.0091	11	0.0093	0.0077	0.0085
5	0.0086	0.0095	0.0085	12	0.011	0.0085	0.0059
6	0.1147	0.0073	0.008	13	0.0073	0.0051	0.007
7	0.0095	0.0069	0.0103	14	0.0061	0.0079	0.0059
8	0.0088	0.0073	**0.0047**	15	0.0106	0.0059	0.0087

The architecture in Figure 7 shows that there are 4 input layers with 4 tapped delay line, 1 hidden layer consisting of 8 nodes and 1 output layer. The inputs are barley, oats and soybean prices and feedback of wheat price and the output is wheat price.

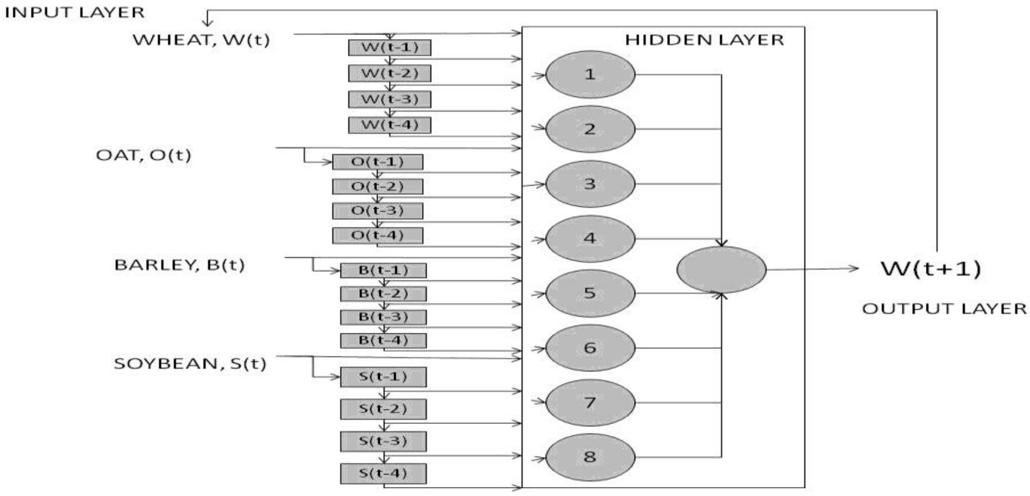

Figure 7: NARX Model Architecture

Figure 8 shows retrained performance (MSE) graph of neural network model, created during its training. The training stopped after 20 epochs because the validation error increased. 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. After initial training of neural network model, it is retrained for 20 epochs and performance MSE is obtained 0.0758 in training

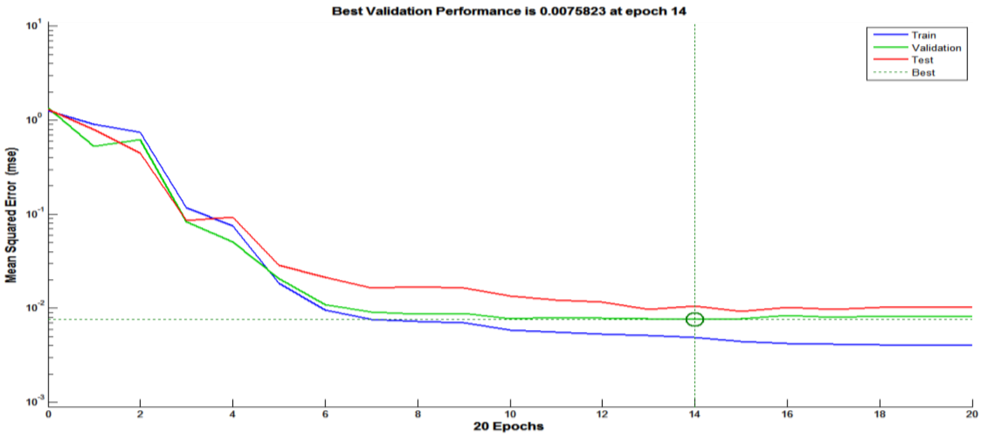

Figure 8: Performance Plot

Based on Figure 9, the all values of *R* are above 0.9 during training of network. This shows that the output produced by the networ k is closely similar to the target and that the model is satisfactory.

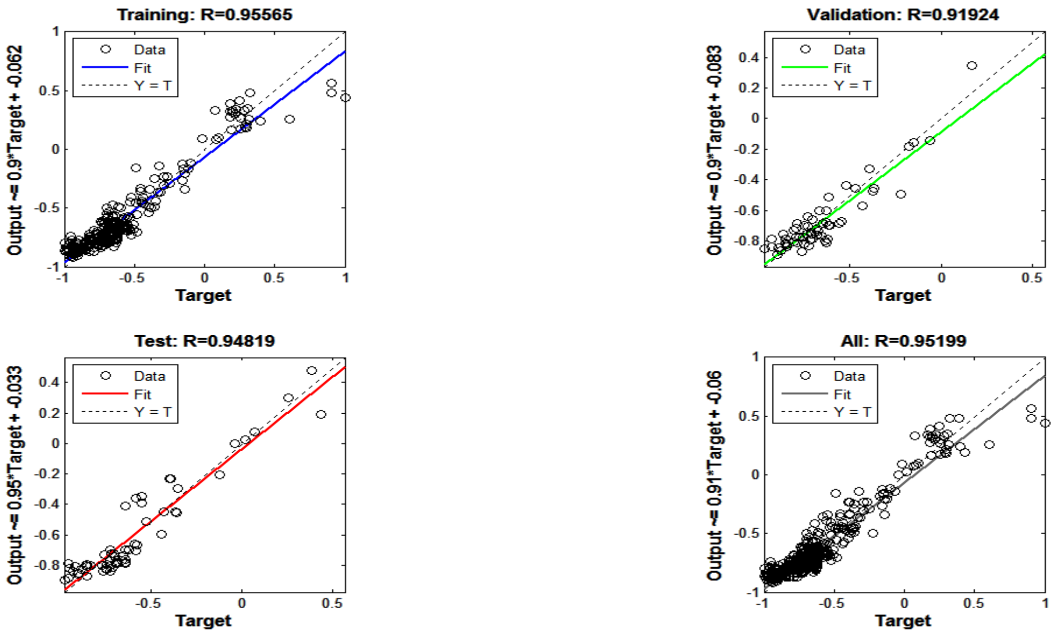

Figure 9: Regression Plots

Model Selection : For this analysis, the best model was selected in order to predict wheat price based on performance of BPNN and NARX methods. The performance is indicated based on accuracy measurement using Mean Square Error (MSE) of both methods. The value of MSE and *R* for NARX model is 0.0047 and 0.9728, while BPNN model is 0.0059 and 0.9724 respectively. The value of MSE is lower compared to BPNN model. The MSE is the sample estimate of the variance of the regression residuals. On the other hand, the value of *R* in NARX model is higher compared to BPNN model. The *R* value measures correlation between Outputs and Targets. An R value close to 1 means a close relationship while vice versa indicates a random relationship. The model with the lowest MSE value and highest *R* value is considered to be the best neural network model. Therefore, NARX model can be used as an alternative model to predict wheat price.

IV. CONCLUSION

The approaches from BPNN and NARX models were compared to find the best model to predict Price of Wheat. After optimization of the NARX model using several tapped delay lines (*d*), it can clearly be seen that NARX model is better than BPNN. Therefore, the NARX model is the suitable model to predict the wheat price with tapped delay lines (*d* = 4), and 8 nodes in the hidden layer using soybean, oats and barley prices as it's predictors. The model's accuracy in predicting wheat price was measured by *R* square and MSE value. The NARX model can be considered as an alternative model for forecasting purposes of wheat price in the future.

REFERENCES

[1] Arbain, S. & Wibowo, A. (2012). Time series methods for water level forecasting of Dungun river in Terengganu Malaysia. *Int. J. Eng. Sci. Technol*., 4: 1803-1811.

[2] Chen, S, Billings, S. A. & Grant, P. M. (1990). Non-linear system identification using neural networks, *Int. J. Contr*., vol. 51, no. 6, pp. 1191-1214.

[3] *Creative Research Systems*. (2013). Retrieved December 22, 2013, from Survey System: http://www.surveysystem.com/correlation.htm

[4] Gould, P. B. (2013). Retrieved from Understanding Dairy Markets:

[5] Hagan, M.T., Demuth H.B. & Beale, M.H. (1996). *Neural Network Design*. 1st Edn., PWS Publishing Company, Boston, ISBN-10: 0534943322, pp: 736.

[6] Jones, A.J. (2004). New tools in non-linear modeling and prediction, *Computational Management Science*1: 109-149.

[7] Kantz, H & Schreiber, T. (2006). *Nonlinear Time Series Analysis*, second ed., Cambridge University Press, Cambridge.

[8] Laurene, F. (1994). *Fundamentals of Neural Networks : Architectures, Algorithm, and Applications.* United States: Florida Institute of Technology.

[9] Muhammad, Z.M., (2009). *Solid Model reconstruction using neural network and the mathematical representation*. Master Thesis, UTM Skudai, Malaysia.

[10] Narendra, K. S. and Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks, *IEEE Trans. Neural Networks*, vol. 1, pp. 4-27.

[11] Sharda, R., (1994). Neural networks for the MS/OR analyst: An application bibliography. *Interfaces*, 24: 116-130.

[12] Siegelmann, H. T. Horne, B. G. & Giles, C. L. (1997). Computational capabilities of recurrent narx neural networks, *IEEE Trans. Syst., Man Cybern*., pt. B, vol. 27, p. 208.

[13] Sivanandam, S. N., & Deepa, S. N. (2006). *Introduction to neural networks using Matlab 6.0*. Tata McGraw-Hill Education

[14] Suguna, N. & Thanuskodi, K.G. (2011). Predicting missing attribute values using K-means clustering. *J. Comput. Sci.*, 7: 216-224. DOI: 10.3844/jcssp.2011.216.224

[15] Takens, F (1981). Detecting strange attractors in turbulence, in: Rand, D.A. & Young L.S. (Eds.), Dynamical Systems and Turbulence, *Lecture Notes in Mathematics*, vol. 898, Springer, Berlin.

[16] Toriman, M.E., Juahir, H.. Mokhtar, M Barzani G.M. & Mastura S.A.S. (2009). Predicting for discharge characteristics in Langat River, Malaysia using neural network application model. *Res. J. Earth Sci*., 1: 15-21.

[17] Vaziri, M., (1997). Predicting caspian sea surface water level by ANN and ARIMA models. *J. Waterway, Port, Coastal Ocean Eng.*, 123: 158-162.

[18] Xie, H., Tang, H. & Liao, Y.H. 2009. Time series prediction based on NARX neural networks: An advanced approach. *Proceeding of the International Conference on Machine Learning and Cybernetics*, Jul. 12-15, IEEE Xplore Press, Baoding, pp: 1275-1279. DOI: 10.1109/ICMLC.2009.5212326

[19] Zhang, G.P., (2002). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50: 159-175. DOI: 10.1016/S0925-2312(01)00702-0

[20] Zhang, Y. & Wu, L. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Syst. Appli.*, 36: 8849-8854. DOI:10.1016/j.eswa.2008.11.028