academicJournals

Vol. 7(5), pp. 127-136, May, 2015
DOI: 10.5897/JEIF2014.0629
Article Number: 8A2F9D052928
ISSN 2141-6672

http://www.academicjournals.org/JEIF

Journal of Economics and International Finance

Full Length Research Paper

Optimizing the monthly crude oil price forecasting accuracy via bagging ensemble models

Hacer Yumurtaci Aydoğmuş[1]*, Aykut Ekinci, Halil İ[2]. Erdal[3] and Hamit Erdal[4]

[1]Akdeniz University, Turkey.
[2]Development Bank of Turkey, Turkey.
[3]Turkish Cooperation and Coordination Agency, Turkey.
[4]Turkish Military Academy, Turkey.

Received 25 November, 2014; Accepted 29 April, 2015

The study investigates the accuracy of bagging ensemble models (i.e., bagged artificial neural networks (BANN) and bagged regression trees (BRT)) in monthly crude oil price forecasting. Two ensemble models are obtained by coupling bagging and two simple machine learning models (i.e., artificial neural networks (ANN) and classification and regression trees (CART)) and results are compared with those of the single ANN and CART models. Analytical results suggest that ANN based models (ANN & BANN) are superior to tree-based models (RT & BRT) and the bagging ensemble method could optimize the forecast accuracy of the both single ANN and CART models in monthly crude oil price forecasting.

Key words: Artificial neural networks, bagging (bootstrap aggregating), classification and regression trees, ensemble models, forecasting.

INTRODUCTION

Oil is an important component of the economic activity and the adverse effect of the crude oil prices on the level of the output is widely recognized in numerous empirical studies (Hamilton, 1983; Hamilton and Herrera, 2004; Huntington, 2005; Barsky and Kilian, 2004; Kilian, 2008). Therefore, forecasting crude oil prices is a very important topic, although it is an extremely hard one due to its intrinsic difficulty and practical applications. The supply and demand forces which are influenced by factors like gross domestic product, stock market activities, foreign exchange rates, weather conditions and political events determine the crude oil prices (Bernabe et al., 2004; Yousefi and Wirjanto, 2004). These factors among others may cause the highly nonlinear and chaotic tendency of the crude oil prices (Yang et al., 2002).

In the past decades, traditional statistical and econometric techniques have been widely applied to crude oil price forecasting. Abramson and Finizza (1991)

*Corresponding author. E-mail: hacera@akdeniz.edu.tr.

utilized a probabilistic model for predicting oil prices. Gulen (1998) attempted to predict the West Texas Intermediate (WTI) price using co-integration analysis. Morana (2001) offered a semiparametric statistical method based on the GARCH properties of crude oil price. Similarly, the GARCH model was used by Morana (2001) to forecast short-term oil prices. Ye et al. (2002, 2005, 2006) presented a single equation model to forecast short-run WTI crude oil prices, using OECD petroleum inventory levels, relative inventories, and high and low-inventory variables. Lanza et al. (2005) used error correction models to predict oil prices. Xie et al. (2006) employed a linear ARIMA model to forecast crude oil prices, argued that oil prices exhibit nonlinear behavior which cannot be captured by linear techniques.

As the traditional and econometric models have some limitations, some non-linear and emerging artificial intelligent models like artificial neural networks (ANN), support vector machines (SVM) and genetic programming (GP) can provide powerful solutions to nonlinear crude oil prediction. Abramson and Finizza (1991) attempted to predict crude oil prices using neural network models. Tang and Hammoudeh (2002) used a non-linear regression model to forecast OPEC basket price. Mirmirani and Li (2004) applied the VAR and ANN techniques to make ex-post forecast of U.S. oil price movements. Their analysis suggests that the BPN-GA model noticeably outperforms the VAR model. Xie et al. (2006) proposed a support vector machine model to forecast WTI prices. To evaluate the forecasting ability of SVM, authors compared its performance with those of ARIMA and BPNN. The experiment results showed that SVM outperforms the other two methods. Shambora and Rossiter (2007) and Yu et al. (2007) also used the ANN model to predict crude oil price. Gori et al. (2007) forecasted oil prices and consumption in the short term under three scenarios: parabolic, linear and chaotic behavior. Silva et al. (2010) used a wavelet decomposition to forecast oil price trends. Azadeh et al. (2010) applied an adaptive intelligent algorithm for forecasting gasoline demand based of artificial neural network (ANN), conventional regression and design of experiment (DOE).

In the recent years, there has been a growing interest in ensemble methods for integrating multiple predictions. To our knowledge there have been very few applications of ensemble models within energy economics. For example, Zhanga et al. (2008) used ensemble empirical mode decomposition (EEMD) for crude oil price analysis. Yu et al. (2008) proposed using an empirical mode decomposition (EMD) based neural network ensemble learning paradigm for crude oil forecasting. Authors found that across different forecasting models, for the two main crude oil prices – WTI crude oil spot price and Brent crude oil spot price – in terms of different criteria, the EMD-based neural network ensemble learning model performs the best. The ensemble methods provide an enhancement of the forecasting accuracy of their individual constituent members such as artificial neural networks and classification and regression trees. The most popular and widely used method is bagging. Thus, we employ bagging in constructing ensemble models in the present study.

The organization of this paper is as follows. Section two is devoted to bagging, classification and regression trees and artificial neural networks. Section tree describes the data, performance statics, application details and empirical results. Finally, some discussions, conclusions and future study directions are given in section four.

METHODS AND DATA

Bagging

Bagging (short for bootstrap aggregating) was proposed by Breiman (1996). It works as follows (van-Wezel and Potharst 2007): A training set D consists of data $\{(X_i, Y_i), i = 1,2,\ldots,n\}$ where X_i is a realization of a multidimensional predictor variable and Y_i contains the label of the case i. For a regression problem, Y_i is a realization of a real valued variable. A replica dataset of size n is randomly drawn with replacement from the original dataset of the n patterns. A bootstrap sample D^* may contain some in D multiple times, whereas others are not included. When a bootstrapped sample is drawn, approximately 37% of the data is excluded from the sample and the remaining data is replicated to bring the data to full size. The excluded one third of the samples is known as the out of bag samples (OOB), while the replicated dataset is known as the in bag samples (Ismail and Mutanga, 2010). A more detailed version of bagging is described in Breiman (1996). Model structure of bagging ensemble developed in the present study is shown in Figure 1. Given a learning model h, bagging is defined for regression problems as follows (Pino-Mejias et al. 2008):

Definition 1. Bagging.

1 Input:
Training sample $= \{(X_i, Y_i), i = 1,2,..., n\}$;
Base learning model h;
2 Process:
I Construct a bootstrap sample $D^* = \{(X^*{}_i, Y^*{}_i), i = 1,2,\ldots,n\}$, according to the empirical distribution of the pairs $U_i, i = 1,2,\ldots,n$, in D.
II Fit h to D^*, obtaining the bootstrapped model h^*.
3 Output:
The bagged predictor is $h^*{}_b(x) = E\left[h^*(x) / D\right]$.

Classification and Regression Trees

Classification and regression trees (CART) was proposed by

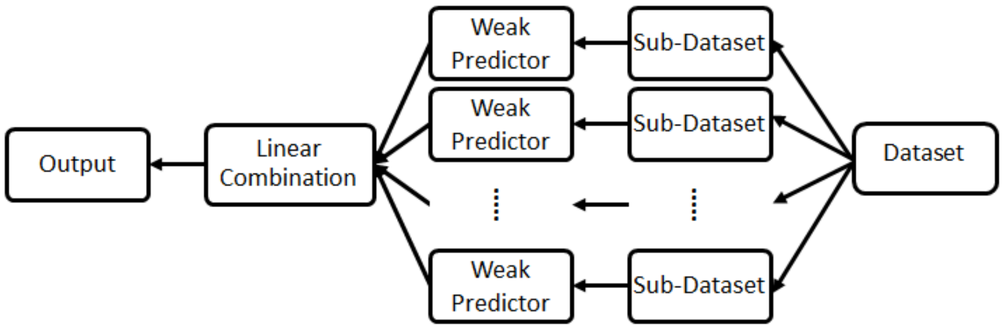

Figure 1. Bagging ensemble model structure.

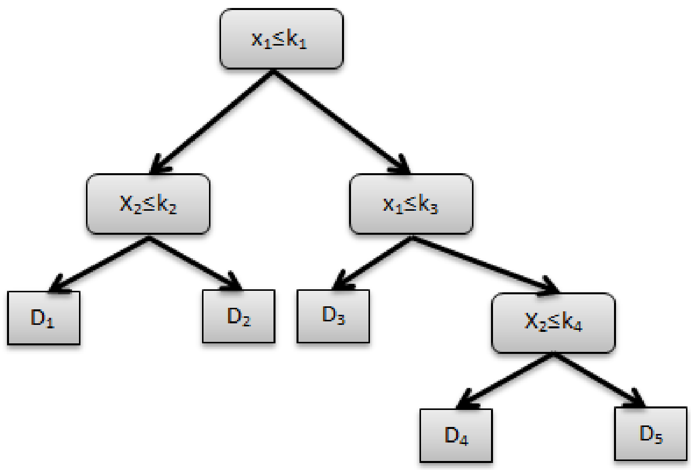

Figure 2. A CART structure.

Breiman et al. (1984) which is a nonlinear statistical technique (Cao et al., 2010).The CART method is based on binary recursive partitioning. A node, which is always partitioned into exactly two new nodes, is called a parent node. The new nodes are called child nodes. The method is recursive since the process can be repeated by treating each child node as a parent node (Grunwald et al., 2009). A terminal node is anode that has no child nodes. The main aim of CART is to estimate the response y by selecting some appropriate variables from a large dataset. It works as follows (Hancock et al., 2005): Each node within the tree has a partitioning rule. For regression problems, the partitioning rule is determined through minimization of the relative error statistic (RE):

$$RE(d) = \sum_{l=0}^{L} (y_1 - \overline{y_L})^2 + \sum_{r=0}^{R} (y_r - \overline{y_R})^2 \qquad (1)$$

Where y_l and y_r are the left and right branches with L and R observations of y in each, with respective means $\overline{y_L}$ and $\overline{y_R}$ .The decision rule d is a point in some predictor variable x that is used to determine the left and right branches. The partitioning rule that minimizes the RE is then used to construct a node in the tree. In the last decade, CART has gained popularity in machine community. However, CART is very sensitive to small changes in the training dataset. More specifically, minor changes in the values of the training dataset can lead to significant changes in the selection of variables (Hastie et al. 2008; Ismail and Mutanga, 2010).Thus, CART is identified as unstable predictor that is prone to overfitting (Breiman, 1996). A CART structure is depicted in Figure 2.

Artificial Neural Networks

This study uses a multilayer perceptron (MLP) which is a conventional back-propagation artificial neural network. Back-propagation process is applied in two phases. The first phase is the forward phase; it involves feeding an input data to the input layer and propagating the signal as far as the output of the network to obtain the prediction. Next, the second phase is the backward phase; the error is employed to adjust the weights of the connections from the hidden to the output neurons. The error is also back propagated and used to adjust the weights of the connections from the input to the hidden neurons (Oliveira et al., 2010).The output signal for the l^{th} neuron in the nth layer is given by,

$$y_l^n(t) = \varphi[\sum_{j=1}^{p} w_{lj}^n(t) y_j^{n-1}(t) + \Psi_l^n] \quad (2)$$

where $\varphi(\cdot)$ is the activation function, w_{lj}^n is the connection weight, t is the time index and $\Psi_l^n = w_{l_0}^n(t)$ is the weighted. For an n-layer network, the synaptic weight $w_{ji}^n(t)$ is given by

$$w_{ji}^n(t+1) = w_{ji}^n(t) + \Delta w_{ji}^n(t) \quad (3)$$

subject to $l \leq n \leq N$ and it can be revised as given by

$$\Delta w_{ji}^n(t) = \eta \lambda_j^n(t) y_i^{n-1}(t) \quad (4)$$

subject to $0 < \eta < 1$

where η is the learning rate, and $\lambda_j^n(t) \equiv -\partial E_t \big/ \partial u_j^n$ is the local error gradient. To improve the back-propagation algorithm, a momentum term α is added

$$\Delta w_{ji}^n(t) = \eta \lambda_j^n(t) y_i^{n-1}(t) + \alpha \Delta w_{ji}^n(t-1) \quad (5)$$

subject to $0 < \alpha < 1$

For the output layer, the local error gradient is given by

$$\lambda_j^N(t) = [d_j(t) - y_j^N(t)]\varphi[u_j^N(t)] \equiv e_j(t)\varphi[u_j^N(t)] \quad (6)$$

where $d_j(t)$ is the goal output signal, and $\varphi(\cdot)$ is the activation function.

2.4. Dataset and experimental settings

The data used in this analysis consist of the monthly West Texas Intermediate (WTI) spot price from January 1982 to November 2011 gathered from the Federal Reserve Bank of St. Louis Federal Reserve Economic Data (FRED). There are various data sets for oil price in the literature, but WTI data is most common due to having long period and providing data continuously from FRED. Bagging ensemble model was applied in forecasting prices in the monthly WTI. Prices are forecasted using time as inputs. In this study, the results are obtained by using a 10-fold cross-validation for each model. The 10-fold cross-validation procedure is applied as follows: First, the WTI dataset is randomized and then data are partitioned into three parts as training set (8 distinct folds), cross-validation set (1 fold) and testing set (1 fold). The training set is employed for the model training and the testing set is used to evaluate the accuracy of models. The cross-validation set is used to apply an early stopping process to avoid overfitting of the training data. Data mining toolkit WEKA (Waikato Environment for Knowledge Analysis) version 3.7.4 is used for experiment. WEKA is an open source toolkit, and it consists of a collection of machine learning algorithms for solving data mining problems (Witten and Frank, 2005).

In this study, the model-specific parameter values we use are as follows: the parameters for MLP are: the number of hidden layers is 5 and 10; the learning rate is 0.3, 0.4 and 0.5; the momentum factor was 0.3, 0.4, and 0.5; and the training time is 300, 400 and 500. The experiments indicate that the best MLP parameters are as follows: the number of hidden layers is 5; the number of the learning rate is 0.3; the momentum factor is 0.4; and the training time is 500. The parameters for the CART are the following: number of folds; the minimum total weight; and number of seeds. In this case, the values for these parameters were 2, 2 and 1 for CART respectively. The bagging parameters are the size of each bag (as a percentage); the number of iterations; and the number of seeds. The best configuration parameters for the bagging are 100, 40, and 1 respectively. The base models (i.e., CART, ANN) parameters are identical to the case in which are they are separately applied. In this study, we offer a better forecasting method for oil price, so we run the program for the each parameter values specified above and select giving the best value. We examined the effects of all the model parameters from the highest values to the least that can be applied in a proper way through the method algorithms. The parameter values that give the highest first three ones are selected for further examination and analyzed for the best values through which we can obtain the least prediction error. Prediction results for each parameter values are compared by using the root mean squared error, the mean absolute error, relative absolute error and root relative squared error accuracy measures.

APPLICATION AND EMPIRICAL RESULTS

The predictive models proposed in this study (i.e., ANN, RT, BRT and BANN) are evaluated by using the four accuracy measures (i.e., the root mean squared error RMSE, the mean absolute error MAE, relative absolute error RAE and root relative squared error RRSE) and also six numerical descriptors (maximum, minimum, mean, variance, maximum under-prediction MUP and maximum under-prediction MOP) are computed to investigate the statistical relation between original data and predicted data.

Mean absolute error:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - a_i| \quad (7)$$

Root mean squared error:

$$RMSE = \sqrt{\sum_{i=1}^{n}(p_i - a_i)^2 / n} \quad (8)$$

Relative absolute error:

$$RAE = \sum_{i=1}^{n}|p_i - a_i| \Big/ \sum_{i=1}^{n}(\overline{a_i} - a_i)^2 \quad (9)$$

Root relative squared error:

$$RRSE = \sqrt{\sum_{i=1}^{n}(p_i - a_i)^2 \Big/ \sum_{i=1}^{n}(\overline{a_i} - a_i)^2} \quad (10)$$

where a =actual target $\bar{a}$ = average and p =predicted

Table 1. The comparison of performance statics for ANN and BANN models.

Model inputs	ANN				BANN			
	MAE	RMSE	RAE (%)	RRSE (%)	MAE	RMSE	RAE (%)	RRSE (%)
WTI $_{t-1}$	3,306	4,084	17,502	16,382	2,040	3,133	10,801	12,567
WTI $_{t-1}$, WTI $_{t-2}$	2,238	2,952	11,849	11,840	1,868	2,526	9,893	10,131
WTI $_{t-1}$, WTI $_{t-2}$, WTI $_{t-3}$	2,317	3,081	12,269	12,359	1,712	2,271	9,063	9,109

Table 2. Numerical descriptors for ANN models and actual data.

Model	Model inputs	min	max	mean	varience	MOP	MUP
Actual		12,850	125,390	34,371	1181,370		
ANN	WTI $_{t-1}$	15,839	110,017	36,641	1342,529	6,816	-15,373
	WTI $_{t-1}$, WTI $_{t-2}$	15,057	112,815	35,595	1266,988	5,281	-12,575
	WTI $_{t-1}$, WTI $_{t-2}$, WTI $_{t-3}$	15,023	113,344	35,700	1274,472	6,167	-12,046
BANN	WTI $_{t-1}$	12,849	109,208	33,735	1138,021	3,707	-16,182
	WTI $_{t-1}$, WTI $_{t-2}$	13,193	113,394	33,861	1146,567	3,426	-11,996
	WTI $_{t-1}$, WTI $_{t-2}$, WTI $_{t-3}$	13,439	116,530	34,267	1174,234	4,823	-8,860

target. Three input combinations based on preceding monthly crude oil prices are developed to forecast current monthly crude oil price. The input combinations evaluated in the study are; (1) WTIt-1, (2) WTIt-1, WTIt-2 and (3) WTIt-1, WTIt-2, WTIt-3. In all cases, the output is the WTIt for the current month. We purposely do not give the training performance statistics, because good testing accuracy gives no guarantee for a low test error. The performance statistics of ANN and BANN models in the test period are given in Table 1. The table indicates that the BANN model whose inputs are the prices of three previous months (input combination 3) has the best accuracy. It can be seen from Table 1 that the BANN model performs better than the single ANN model from the various performance criteria viewpoints. The table shows that the relative MAE, RMSE, RAE and RRSE differences between the BANN (input combination 3) and ANN (input combination 2) models are 23.514%, 23.065%, 2.786% and 2.731% in the test period, respectively. Table 2 summarizes the numerical descriptors (max, min, mean, variance, maximum over prediction and maximum under prediction) for the ANN and BANN models. The numerical descriptors estimated for the ANN and BANN models indicate that the BANN model yields more similar estimates and distributions when compared with the actual WTI data.

Table 3 indicates that the BRT model whose inputs are the prices of two previous months (input combination 2) has the smallest MAE, RMSE, RAE and RRSE in testing period. And it is found that the RT model has the best accuracy for the input combination 3. Compared with the RT models, the BRT models yield better accuracy in monthly crude oil price forecasting. The relative MAE, RMSE, RAE and RRSE differences between the BRT (input combination 2) and RT (input combination 3) models are 14.581, 16.761, 2.403 and 3.918% in the test period, respectively. The numerical descriptors shown in Table 4 for the RT and BRT models show that the BRT model provides more similar estimates and distributions than RT. The BANN, ANN, BRT and RT residuals in test period are shown in Figure 3 for all input combinations respectively. It can be seen from the residuals that BANN approximates the actual values better than the others. The underestimations are obviously seen for the tree-based models.

The direct relationship between the MAE, RMSE, RAE and RRSE is very clear according to Tables 1 and 3. The best model for minimizing MAE (1.712) and RMSE (2.271) is BANN, the 2nd model is ANN (MAE=2.238, RMSE =2.952), the 3th model is BRT (MAE=2.659, RMSE =4.850) and finally the worst model is RT (MAE=3.112, RMSE =5.827). Tables 1 and 3 indicate BANN (RAE=9.063, RRSE=9.109) and ANN models (RAE=11.849, RRSE=11.840) are superior to the BRT (RAE=14.076, RRSE=19.457) and RT (RAE=16.479, RRSE=23.375) models for determining, RAE and RRSE statics. In general, note from Tables 1 to 3 that ensemble learning always provides a good improvement and

Table 3. The comparison of performance statics for RT and BRT models.

Model inputs	RT				BRT			
	MAE	RMSE	RAE (%)	RRSE (%)	MAE	RMSE	RAE (%)	RRSE (%)
WTI $_{t-1}$	3,295	5,855	17,449	23,486	2,755	4,989	14,588	20,014
WTI $_{t-1}$, WTI $_{t-2}$	3,259	5,924	17,254	23,763	2,659	4,850	14,076	19,457
WTI $_{t-1}$, WTI $_{t-2}$, WTI $_{t-3}$	3,112	5,827	16,479	23,375	2,826	5,078	14,964	20,368

Table 4. Numerical descriptors for RT models and actual data.

Model	Model inputs	min	max	mean	varience	MOP	MUP
Actual		12,850	125,390	34,371	1181,370		
	WTI $_{t-1}$	13,358	93,636	32,535	1058,512	6,691	-31,754
RT	WTI $_{t-1}$, WTI $_{t-2}$	14,143	93,636	32,544	1059,127	6,691	-31,754
	WTI $_{t-1}$, WTI $_{t-2}$, WTI $_{t-3}$	14,143	93,636	32,758	1073,063	6,691	-31,754
	WTI $_{t-1}$	13,456	99,190	33,017	1090,136	7,517	-26,200
BRT	WTI $_{t-1}$, WTI $_{t-2}$	13,619	99,130	33,102	1095,750	6,873	-26,260
	WTI $_{t-1}$, WTI $_{t-2}$, WTI $_{t-3}$	13,046	96,860	33,032	1091,116	7,830	-28,530

ensemble models (i.e., BANN, BRT) seems to be more adequate than the single ANN and RT models for forecasting monthly crude oil prices (Table 4).

The actual and predicted WTI distributions of the input combinations 1, 2 and 3 for testing period are depicted with boxplots presented in Figures 4, 5 and 6. The box height corresponds to the interquartile range, the whiskers depict the 5th and 95th percentiles and the horizontal line is the median. Dots indicate values outside the range and the horizontal line within each boxes indicate the median values. The performance of BANN model was better than the ANN, RT and BRT models when compared to the distribution of the actual WTI data. Moreover the distribution of WTI data predicted by the BANN model is similar to the distribution of actual data and the BANN model did the best job at the capturing the actual data for test phases.

DISCUSSION AND CONCLUSION

Ensemble learning is the supervised learning from the information generated by the base predictors. The main goal is to build an ensemble model that provides base predictor functionality and to increase the accuracy by combining the individual models (Chou et al., 2011). Integrating multiple instances of the same model type can reduce the variance and enhance prediction accuracy (Wang et al., 2009). In the present study, we have investigated the potential use of bagging ensemble models for monthly crude oil price forecasting. The ensemble models (i.e., bagged artificial neural networks BANN, bagged regression trees BRT) are obtained by coupling bagging and two single unstable machine learning model (i.e., ANN, CART). We have also employed the base models ANN and CART as benchmark models and used tree input combination to test proposed predictive models. In general, the bagging method can be very effective procedure when applied to unstable learning algorithms, such as classification and regression trees and artificial neural networks (Mejias et al. 2010). Moreover, bagging ensembles can inherit almost all advantages of their base models while overcoming their primary problem, which is inaccuracy. Breiman (1996) pointed out that the bagged model variance is smaller than or equal to the variance of a simple model (i.e. CART, ANN), leading to increasing prediction accuracy (Louzada et al. 2011).

The obtained results from the study indicate that (i) bagging always provides a considerable enhancement. Bagged models (i.e., BANN, BRT) reduce the mean absolute errors, root mean squared errors, relative absolute errors and root relative squared errors with respect to the single ANN and CART models by 23.514-14.581%, 23.065-16.761%, 2.786-2.403% and 3.918-2.731%, respectively; (ii) ANN-based predictive models (i.e., BANN, ANN) are found better than tree-based predictive models (i.e., BRT, RT). (iii) BANN model is a

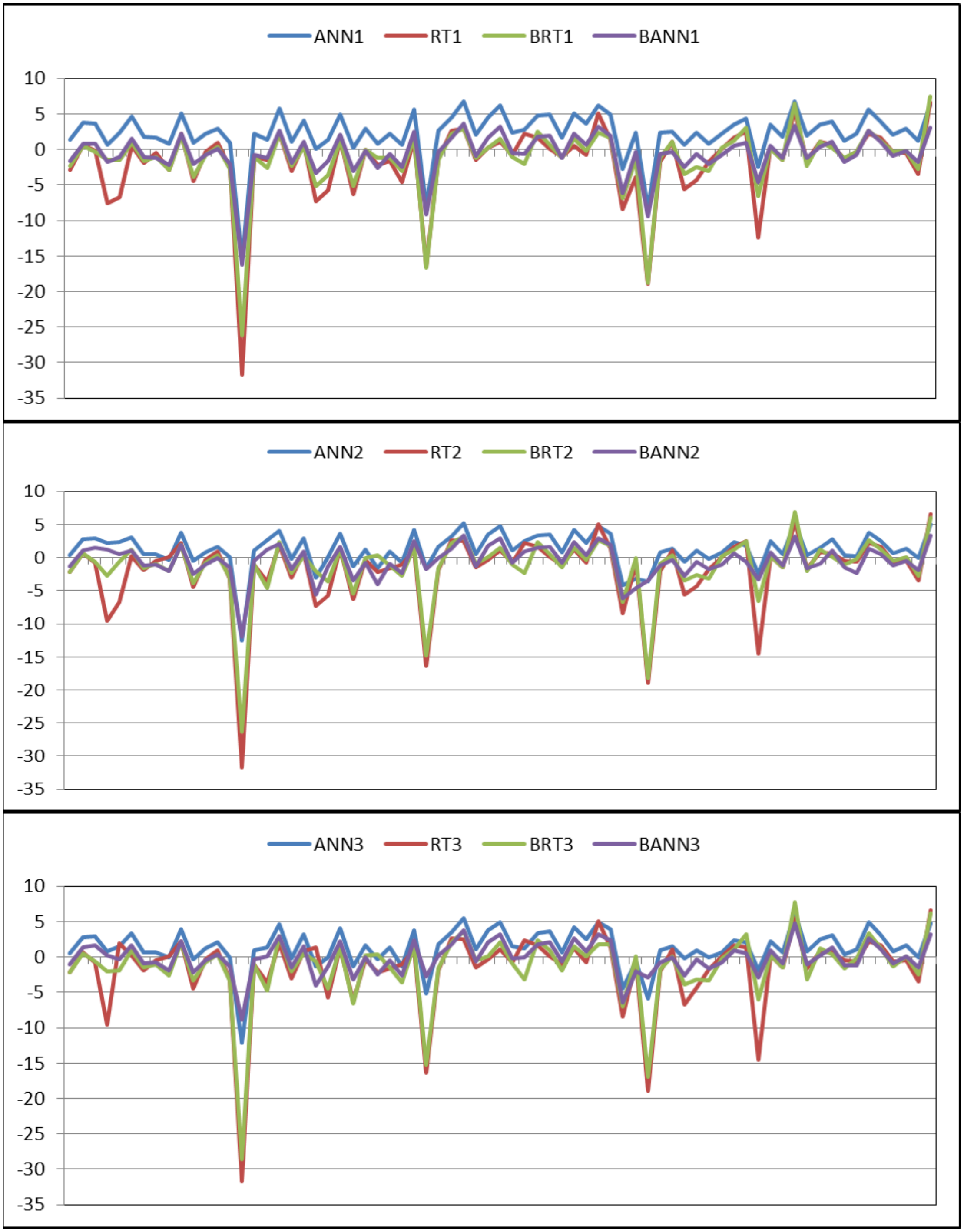

Figure 3. Residuals for the ANN, BANN, RT and BRT models.

promising approach for monthly crude oil price forecasting and finally (iv) the numerical descriptors (maximum, minimum, mean, variance, maximum under-prediction and maximum under-prediction) estimated for the proposed predictive models indicate that the BANN model yields statically similar estimates and distributions when compared with the actual WTI data. In this study, bagging method is used in building ensemble models. The other ensemble models (e.g., boosting, random forest) could be used for construction of ensemble

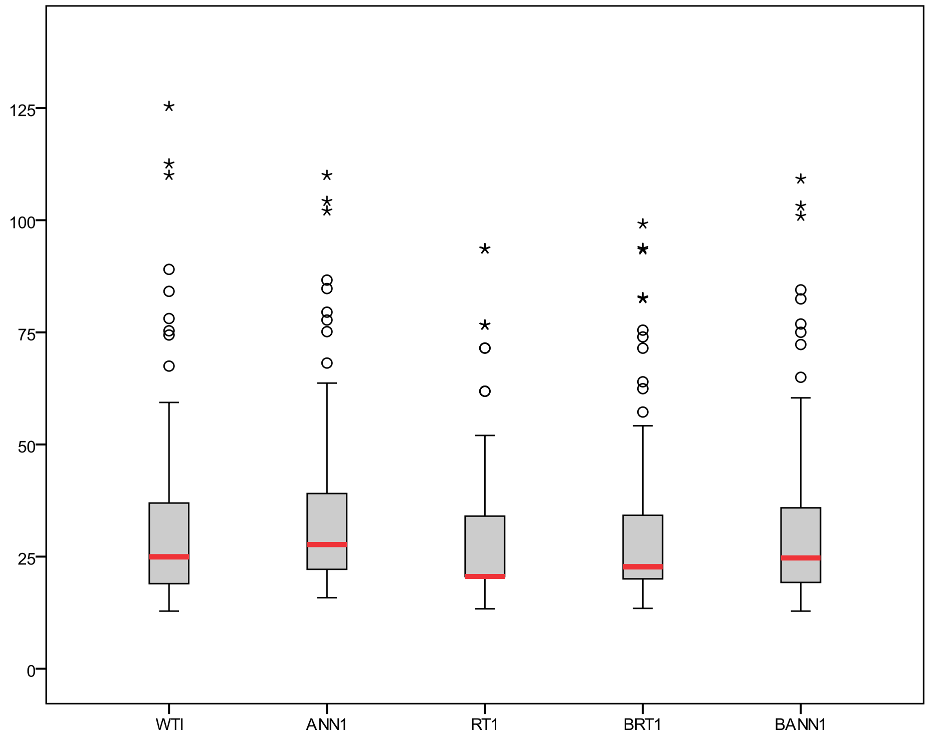

Figure 4. Box plots of actual and predicted WTI distributions for input combination 1.

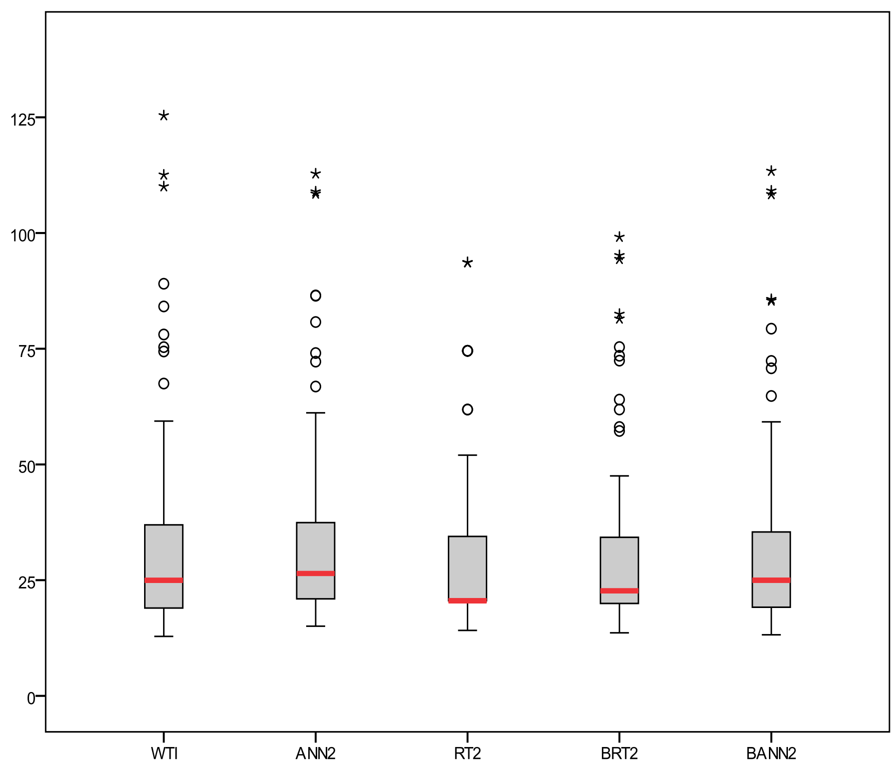

Figure 5. Box plots of actual and predicted WTI distributions for input combination 2.

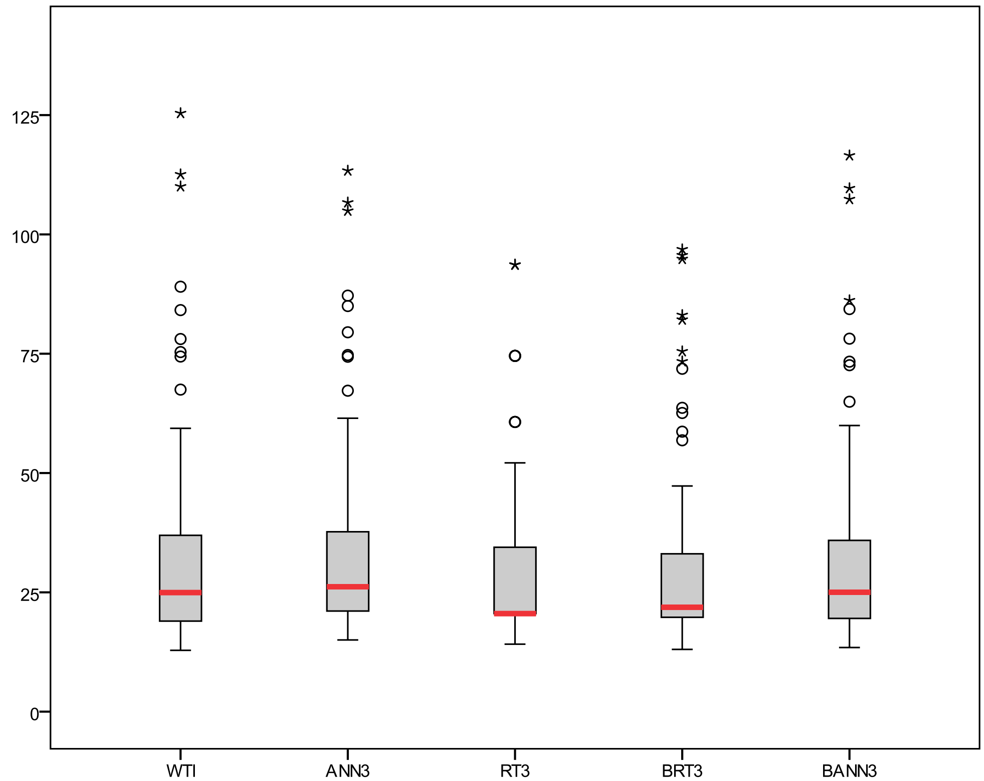

Figure 6. Box plots of actual and predicted WTI distributions for input combination 3.

models. We propose to investigate the usage of other ensemble models for future work.

Conflict of Interests

The authors have not declared any conflict of interests.

REFERENCES

Abramson B, Finizza A (1991). Using belief networks to forecast oil prices. Int. J. Forecast. 7(3):299–315.

Azadeh A, Arab R, Behfard S (2010). An adaptive intelligent algorithm for forecasting long term gasoline demand estimation: The cases of USA, Canada, Japan, Kuwait and Iran. Expert Syst. Appl. 37:7427-7437.

Barsky RB, Kilian L (2004). Oil and the macroeconomy since the 1970s. J. Econ. Persp. 18:115-134.

Bernabe A, Martina E, Alvarez-Ramirez J, Ibarra-Valdez C (2004). A multi-model approach for describing crude oil price dynamics. Physica A: Stat. Mechanics Appl. 338(3-4):567–584.

Breiman L (1996). Bagging predictors. Machine Learn. 24:123–140.

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984). Classification and regression trees, Belmont, California: Wadsworth, Int. Group. pp.357-358.

Chou JS, Chiu CK, Farfoura M, Al-Taharwa I (2011). Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques. J. Comput. Civ. Eng. 25:242-253.

Cao DS, Xu QS, Liang YZ, Chen X, Li HD (2010). Automatic feature subset selection for decision tree-based ensemble methods in the prediction of bioactivity. Chemometrics Intelligent Laboratory Systems 103:129–136.

Grunwald S, Daroub SH, Lang TA, Diaz OA (2009). Tree-based modeling of complex interactions of phosphorus loadings and environmental factors. Sci. Total Environ. 407:3772–3783.

Gulen SG (1998). Efficiency in crude oil future markets. J. Energ. Financ. Dev. 3:13–21.

Gori F, Ludovisi D, Cerritelli PF (2007). Forecast of oil price and consumption in the short term under three scenarios: parabolic, linear and chaotic behavior. Energy 33(4):1291–1296.

Hamilton JD (1983). Oil and the macro economy since world war II. J. Polit. Econ. 91:228-248.

Hamilton JD, Herrera AM (2004). Oil shocks and aggregate economic behavior: the role of monetary policy. J. Money, Credit Bank. 36:265-86.

Hancock T, Put R, Coomans D, Vanderheyden Y, Everingham Y (2005). A performance comparison of modern statistical techniques for molecular descriptor selection and retention prediction in chromatographic QSRR studies. Chemometrics Intelligent Laboratory Syst. 76(2):185-196.

Hastie T, Tibshirani R, Friedman J (2001). The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer-Verlag, New York, p.500.

Huntington HG (2005). The economic consequences of higher crude oil prices. Energy Modeling Forum. Special Report 9, Stanford University.

Ismail R, Mutanga O (2010). A comparison of regression tree ensembles: Predicting Sirexnoctilio induced water stress in Pinuspatula forests of KwaZulu-Natal, South Africa. Int. J. Appl. Earth Observation Geoinformation 12S:S45–S51.

Kilian L (2008). The economic effects of energy price shocks. J. Econ. Literature 46:871-909.

Lanza A, Manera M, Giovannini M (2005). Modeling and forecasting cointegrated relationships among heavy oil and product prices. Energ. Econ. 27: 831–848.

Louzada F, Anacleto-Junior O, Candolo C, Mazucheli J (2011). Poly-bagging predictors for classification modelling for credit scoring. Expert Syst. Appl. 38:12717–12720.

Mirmirani S, Li HC (2004). A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil. Adv. Econ. 19:203-223.

Morana C (2001). A semiparametric approach to short-term oil price forecasting. Energ. Econ. 23:325-338.

Oliveira ALI, Braga PL, Lima RMF, Cornélio ML (2010). GA-based method for feature selection and parameters optimization for machine learning regression applied to software effort estimation. Inf. Software Technol. 52:1155–1166.

Pino-Mejias R, Jimenez-Gamero MD, Cubiles-de-la-Vega MD, Pascual-Acosta A (2008). Reduced bootstrap aggregating of learning algorithms. Pattern Recognition Letters 29:265–271.

Silva GdS, Legey FL, Silva AdS (2010). Forecasting oil price trends using wavelets and hidden Markov models. Energ. Econ. 32:1507-1519.

Shambora WE, Rossiter R (2007). Are there exploitable inefficiencies in the futures market for oil? Energ. Econ. 29:18–27.

Tang L, Hammoudeh S (2002). An empirical exploration of the world oil price under the target zone model. Energ. Econ. 24:577–596.

Van-Wezel M, Potharst R (2007). Improved customer choice predictions using ensemble methods. Eur. J. Oper. Res. 181:436-452.

Witten IH, Frank E (2005). Data mining: Practical machine learning tools and techniques. Morgan Kaufman Publishers, Boston.

Xie W, Yu L, Xu SY, Wang SY (2006). A new method for crude oil price forecasting based on support vector machines. Lecture Notes Comput. Sci. 3994:441-451.

Yu L, Lai KK, Wang SY, He KJ (2007). Oil price forecasting with an EMD-based multiscale neural network learning paradigm. Lecture Notes Comput. Sci. 4489:925–932.

Yousefi A, Wirjanto TS (2004). The empirical role of the exchange rate on the crude-oil price information. Energ. Econ.. 26:783-799.

Yang X, Khumera H, Zhang W (2009). Back propagation wavelet neural network based prediction of drill wear from thrust force and cutting torque signals. Comput. Inf. Sci. 3(2):75–86.

Ye M, Zyren J, Shore J (2002). Forecasting crude oil spot price using OECD petroleum inventory levels. Int. Adv. Econ. Res. 8:324-334.

Ye M, Zyren J, Shore J (2005). A monthly crude oil spot price forecasting model using relative inventories. Int. J. Forecast. 21:491-501.

Ye M, Zyren J, Shore J (2006). Forecasting short-run crude oil price using high and low-inventory variables. Energ. Policy 34:2736–2743.

Yu L, Wang S, Lai KK (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energ. Econ. 30(5):2623–2635.

Zhanga X, Laic KK, Wang SY (2008). A new approach for crude oil price analysis based on Empirical Mode Decomposition. Energ. Econ. 30(3):905–918.