Journal of Economic & Financial Studies

Likelihood of financial distress in Canadian oil and gas market: An optimized hybrid forecasting approach

Mohammad Mahbobi[a]*, Rashmit Singh G. Sukhmani[a]

[a] School of Business and Economics, Thompson Rivers University, Kamloops, Canada.
*Corresponding author's email address: mmahbobi@tru.ca

A R T I C L E   I N F O

A B S T R A C T

Forecasting models are built on either multivariate parametric or nonparametric methodologies. We attempt to optimize the accuracy of the forecasts combining these approaches to make a robust hybrid forecasting model in predicting the likelihood of financial distress for companies in the Canadian oil and gas market. The proposed approach combined the forecasts out of a multivariate logit model based on the conventional Altman's Z-score with a nonparametric Artificial Neural Network (ANN) technique. The sample firms are publicly traded and listed on the Toronto Stock Exchange (TSX) and span over a period from first quarter of 1999 to the last quarter of 2014. The results of a proposed three-stage estimation process for the period of 2015-2020 indicated that besides the fact that Canadian energy sector will go through ups and downs regarding the likelihood of financial distress, this industry would face a hard time by late 2020. Results show that the forecasting accuracy out of the proposed three-stage forecasting technique is significantly superior to the outcomes of any individual forecasting techniques, i.e. ANN and logit models.

1.   Introduction

Corporate failures have been one of the main reasons for the economic downturns of various countries. The prediction of the corporation's financial distress is important and intensely studied topic (Tan, 1997). Corporate failure has a significant impact on the economy, in particular for the stakeholder dependent or affected by the actions of the corporation directly or indirectly. It is important to the creditors and investors to forecast the financial distress of the companies in order to make business decisions (Hauser, et al., 2011). Many parametric and non-parametric models are generated and published in the literature for the prediction of the financial distress. The statistical models used for the prediction are multivariate discriminant analysis, logistic regression, and neural networks. Research shows that "Artificial Neural Network (ANN)", "logit and probit" models have more accuracy to predict the financial distress (Andres, et al., 2005). Due to inadequacy of the logistic regression model, many non-parametric models have been proposed to train themselves based on the data fed into the model and provided acceptable results (Andres, et al., 2005).

2.   Background

Many models had been developed and tested to study the financial distress of the companies. One of the most traditional techniques by the econometrics is the discriminant analysis that is multivariate technique. In this

technique, Z-score is assigned based on the combination of financial ratios of the companies (Altman, 1968). A cut-off value is decided based on the sample data and businesses falling below this cut-off value will be categorized as a firm facing financial distress. The best thing about this technique is that it reduces the multidimensionality of the input variables for each company into one single Z-Score. The principal critique of this model is that it is not robust in nature and has bad accuracy in predicting the financial distress (Tan, 1997). The other type of models that are published in the literature are the logit, probit and Artificial Neural Network models (Tan, 1997). The Logit models have the similar accuracy as discriminant analysis techniques, whereas probit and ANN model has a better accuracy (Tan, 1997). The model should be such that it should grab an essence of the thinking ability of the decision-making ability of the human mind. The primary constraint on these kinds of studies is the complexity of the model. In recent studies, ANN models are proposed as it can take into account the relationship between the financial ratios variable and predict the financial distress of the firm (Edmister, et al., 1995). In our study, we will be using artificial neural network model and Logit model for predicting the financial distress of the Canadian companies.

2.1   Altman's Z-score model

Edward Altman derived a formula in 1967 for multi-discriminant analysis to predict the financial distress for the manufacturing companies(Anjum, 2012). Altman studied all the financial ratios to develop a formula to predict financial distress. Many well-known investing firms used this model to have an approximate idea of the state of the companies. There have been many revised version of Altman's Z-Score model to predict accurately the financial distress of the companies (Altman, 1968).The significantly used version of Altman Z-score model is (Deakin, 1972): $Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 1X_5$. Where, $X_1 = \frac{Working\ Capital}{Total\ Assets}$, $X_2 = \frac{Retained\ Earnings}{Total\ Assets}$, $X_3 = \frac{Earning\ Before\ Interest\ and\ Taxes\ (EBIT)}{Total\ Assets}$, $X_4 = \frac{Market\ Value\ of\ Equity}{Total\ Liabilities}$, and $X_5 = \frac{Total\ Sales}{Total\ Assets}$.

Table 1: Altman's financial distress criteria

Criteria	Zone	Description
$Z < 1.8$	Distress Zone	Highly likely to get bankrupt based on the financial data.
$1.8 < Z < 2.7$	Grey Zone	Likely to get bankrupt in next 2 years based on the financial data.
$Z > 2.7$	Safe Zone	Safe based on the financial data.

Error! Reference source not found.shows the cut-off value for the Z-Score for financial distress (Deakin, 1972). Many studies have been done by the researchers to study the service industry, manufacturing industry, publicly listed companies and banks (Anjum, 2012). The studies of 40 years show that Altman Z-Score model can be applied to the modern economies to have a constructive predictability of the financial distress (Altman, 1968).

2.2   Logistic regression model

The Linear Probability Model (LPM) has the output in terms of binary digits. The dependent variable y_i^* can written as:

$$y_i^* = x_i\beta_i + \epsilon_i \tag{2}$$

where x_i is the independent variable, β_i is the estimated parameters of the equation and ϵ_i is the error term associated with the estimation (Whittington, Briscoe, Mu, & Barron, 1990). The next step is to break the dependent variable into a binary format to represent a binary response model (BRM) by:

$$y_i = \begin{cases} 1\ if\ y_i^* > \pi \\ 0\ if\ y_i^* < \pi \end{cases} \tag{3}$$

where π is the threshold value decided based on the theoretical outcome (Zmijewski, 1984) of y_i^*. The outcome of the model is in the binary form that does not follow the normalized distribution of the error. The ordinary linear regression estimation techniques cannot be used to determine the β parameters of the equation. To overcome estimation problem, maximum likelihood estimation (MLE) technique is used. The relation between y_i^* and y_i in terms of the probability is (Zmijewski, 1984):

$$\Pr(y_i = 1) = \Pr(x_i\beta_i + \epsilon_i > \pi) \tag{4}$$
$$\Pr(y_i = 1) = \Pr(\epsilon_i - \pi > -x_i\beta_i) \tag{5}$$
$$\Pr(y_i = 1) = 1 - \Pr(\epsilon_i - \pi < -x_i\beta_i) \tag{6}$$
$$\Pr(y_i = 1) = 1 - F(-x_i\beta_i) \tag{7}$$

where F is the cumulative distribution function of $\epsilon_i - \pi$. The cumulative distribution function (CDF) of the $\Pr(y_i = 1)$ in the form of normal distribution function expressed as (Long, 1997):

$$\Pr(y_i = 1) \sim N(\mu_i, \sigma^2) = \prod_0^{\mu_i = x_n \beta_n} \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{\frac{-(y_i - \mu_i)^2}{2\sigma^2}} \qquad (8)$$

with $\mu_i = x_i \beta_i$ and σ^2 being the variance of the distribution. Now taking natural logs on the both side to determine the Log-Likelihood function, we get (Long, 1997):

$$\ln(Pr(y_i = 1)) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\sum_{i=1}^{N} \frac{(y_i - \mu_i)^2}{2\sigma^2} \qquad (9)$$

However, to find out ML estimator, we need to differentiate with respect to β_i & σ^2 and equalized it to zero to find out the value of β_i to maximize the equation (King, 1998).

2.3   Artificial neural network (ANN)

Artificial neural networks are the developed form of models that are interconnections between the inputs with the output with the help of artificial neuron to mimic the decision-making process of the human brain (Tan, 1997). The artificial neural network models are free from the distributional assumption and avoid the collinearity problems in the data. The financial analyst uses these kinds of models because of its ability to determine the weights of the neuron on its own without the help of prior external assumption (Edmister, et al., 1995). ANN model used in the approach is a multi-layered feed-forward network. The inputs are fed at the bottom and propagated through a neural network to the output layer. The transfer function output 'y' is expressed by:

$$y_k = f(\alpha) = g\left\{\sum_{j=1}^{n} w_{jk} X_j - \theta_k\right\} \qquad (10)$$

where θ_k is the threshold activation value known as offset, w_{jk} are the weights assigned to the neurons, g is the sigmoid function and X_j is the value of the input nodes (Edmister, et al., 1995).

2.3.1   Artificial neural network training methodologies

ANN network model has to be trained to determine the input output relationship. If the ANN model is trained properly, then it provides the financial analyst with the accurate future forecast of the economic events (Edmister, et al., 1995). Backpropagation algorithm is used to train the weights of the neural network to get the accurate output. It's the iterative learning algorithm that trains the neural weights according to the inputs and desired outputs. The first step of the iterative process is to randomly initialize the weights and threshold. The neural network algorithm is run with the help of initialized weights and threshold values to generate outputs using the transfer function. The error is calculated with the help of the target value and output values. The error term δ is:

$$\delta = \left(T_{q,k} - Out_{q,k}\right)Out_{q,k}\left(1 - Out_{q,k}\right) \qquad (11)$$

where, $T_{q,k}$ is the desired output, $Out_{q,k}$ is the output of the neural network and n is the number of elements in the trail set. Once the error term is obtained, the weights are reinitialized to minimize the error term. The change weights w_i^+ of the connection between the input nodes and hidden layers is given by:

$$w_i^+ = w_i + \left(\delta \times Out_{q,k}\right) \qquad (12)$$

The next step is to change the weights of the hidden layer connection to the output node. The weights are changed by back propagating error form output layer to hidden layer by calculating the error. The error term δ_{hidden} is expressed as:

$$\delta_{hidden} = Out_{q,k}\left(1 - Out_{q,k}\right)\left(\delta \times w_{hidden}\right) \qquad (13)$$

The weights of the hidden connection are changed by the same equation used in changing the weights of input node connections. The whole process is repeated till the error is minimized to obtain the desired output. In 1990, Jeffery Elman developed a neural network known as Elman Network. It's the feed forward system with a difference of the context units connected to the hidden layers (Elman, 1990). Elman's idea was to develop time delay network that could recognize and predict learned sequential time-varying patterns. In Elman's neural architecture, the input units and context units activate the hidden layer that feed forward to activate the output layer (Ben, et al., 1996). The generated outputs are compared with the desired output to calculate the error. The

error is calculated which is back propagated to adjust the weights same as in standard feedforward backpropagation neural network (Elman, 1990). In the Elman network, in the initial stage the recurrent unit weights are not subjected to the adjustment. At the next time step, the context units are assigned the values of the immediate previous time state that are exactly same as hidden layers' node weights (Svetlana, et al., 2012). In this way, the context units' values provide temporary memory to the network of the previous training weights of the hidden layers. In this way, Elman made the neural network more dynamic to get adjusted to local changes and to avoid any convergence problem that are associated with the backpropagation algorithm (Svetlana, et al., 2012). Cascade is feedforward neural networks that are similar to the back propagation neural networks. The difference between the two is that there are more connectivity's between the nodes in the network (Allaf, et al., 2011). In the cascade neural network, inputs layers are connected to all the layers in the system with the adjustable weights (Allaf, et al., 2011). The architecture of cascade network makes the neural network more stable and adaptable to the problem. It is used in the short-term forecasting and have shown good performance with minimizing mean square errors (Shrivastava, et al., 2013). Cascade feed forward network have the potential to learn any input-output relationship due to the input layer connection to other layers in the network with the adjustable weights virtually (Shrivastava, et al., 2013).

3.   Data description

In the last nineteen years, the predictability of financial distress of many models evolved in the literature. This likelihood of the companies depends on the chosen financial ratios. Many models were proposed which used different sets of financial ratios for predicting the financial distress includes multivariate discriminant analysis (MDA), Logit Model, and recursive partitioning algorithm (RPA) (Edmister, et al., 1995). The financial ratios used in the recursive partitioning algorithm are so far, the best indicators of distress in all the prior studies (Frydan, et al., 1985). Built on (Tan, 1997), the financial indicators for the financial distress are chosen based on the following criteria:

a.   Previous studies showing the significant effect of the financial ratios on the distress analysis.
b.   Availability of the financial ratio data.
c.   Authenticity of the data.

The scope of this research was on developing a forecasting model to predict the likelihood of future financial distress of the Canadian energy industry. Canadian Energy industry in this paper consists of thirty-one of those energy service companies that have the core competencies in drilling. We used 33 Technology sector and 31 Energy sector firm's financial information for this analysis. The quarterly data from 1999 to 2015 can be collected from ADVFN website (ADVFN, n.d.). The required data calculated based on the Altman Z-score financial distress criteria. For the training of the ANN model, randomly categorized quarterly information of the financial ratios and the Altman Z-score are used as an input and output respectively. The aggregated values of the selected financial ratios as an input and the desired binary output based on the Altman distress criteria is used to generate the final output of probabilities that can be used for the sake of Logit model. The financial ratios used in our analysis are consistent with those consistent with the approach used in (Dorsey, et al., 1995). A list of the included ratios along with their descriptions have been provided in Appendix Table 2.

4.   Estimation process

The estimation process of this study carried out in three stages. In stage one, the Altman's Z-Score was calculated for each company with the help of financial ratios for every quarter. Then, a random sampling of quarterly data of each industry used to train the Artificial Neural Network (ANN) model. Nineteen financial ratios used as an input for the ANN model of those five financial variables corresponded to the Z-scores. The other 15 variables taken into consideration in order to determine their influence on the output and improve the accuracy of the model. Consistent with (Edmister, et al., 1995), random sampling helped in avoiding biases among the procedure of selecting the data for the purpose of training and testing the model. After training the model of the artificial neural network, the selected financial ratios of companies for each quarter were fed to obtain the associated Altman's Z-Score for all the selected companies of both industries. After aggregating and standardizing the Z-Scores for each industry in each quarter, they converted into probabilities of financial distress. The quarterly Z-Scores assigned the value 1 or 0 depending on whether the companies were doing well in that quarter. This approach was set to be conservative in nature. In the literature, Altman financial criteria are categorized in three cut-off values. But in our analysis, we converted the Altman Z-Score into two cut-off values of *safe zone* of greater than or equal to a threshold of 1.81 and a *distress zone* of less than 1.81. Same aggregation process was conducted to obtain quarterly financial ratios for each industry. These obtained financial ratios and binary outputs were used as inputs for the logit model to obtain the probabilities of the financial distress for each sector.

In stage two, in order to improve the prediction accuracy of the obtained result from ANN and Logit model, a Forecast Artificial Neural Network (FANN) model was estimated by combining the results out of the Logit and ANN models found in stage one. The neural network technique used for the FANN model was Elman Back Propagation approach to increase the performance and avoid convergence problem. For training purpose, random weighted probabilities were generated using the actual ANN and Logit outputs. Once the training of ANN model completed, inputs of each industry's ANN and Logit model were fed to obtain the optimized quarterly probabilities of the financial distress for each industry. Following the proposed combining processes proposed by (Brandt, et al., 1981) and (Li, et al., 2004), and in order to further improve accuracy of the predicted probabilities, an optimizing algorithm was used to calibrate the intercept and slope of a linear regression model by minimizing the mean square error of the combined outputs. Finally, in stage three, cascade forward backpropagation neural network, which mostly employed in time series analysis as a way to improve the performance of the time series model, was used. Inputs of the ANN model was the time "t" for the year 1999 to 2014, and desired outputs were the corresponding optimized quarterly results of the industry. These datasets were then used to train each industry's ANN outputs to obtain the ultimate forecasts for 2015 to 2020. The diagram shown in **Error! Reference source not found.** outlines the process of the proposed three stages.

5.   Empirical results

The calculated Altman Z-Score for each representative company in each quarter used to create an aggregated Altman Z-Score for the entire industry for the period of 1999 to 2014. Values of the aggregated Altman Z-Score were then standardized to obtain the probability of financial distress in each quarter for the entire industry. In the following sections, the simulated/estimated outcomes out of each model described consistent with the proposed three-stage process will be provided.

Figure 1: Steps of the three stages of the proposed hybrid forecasting process.
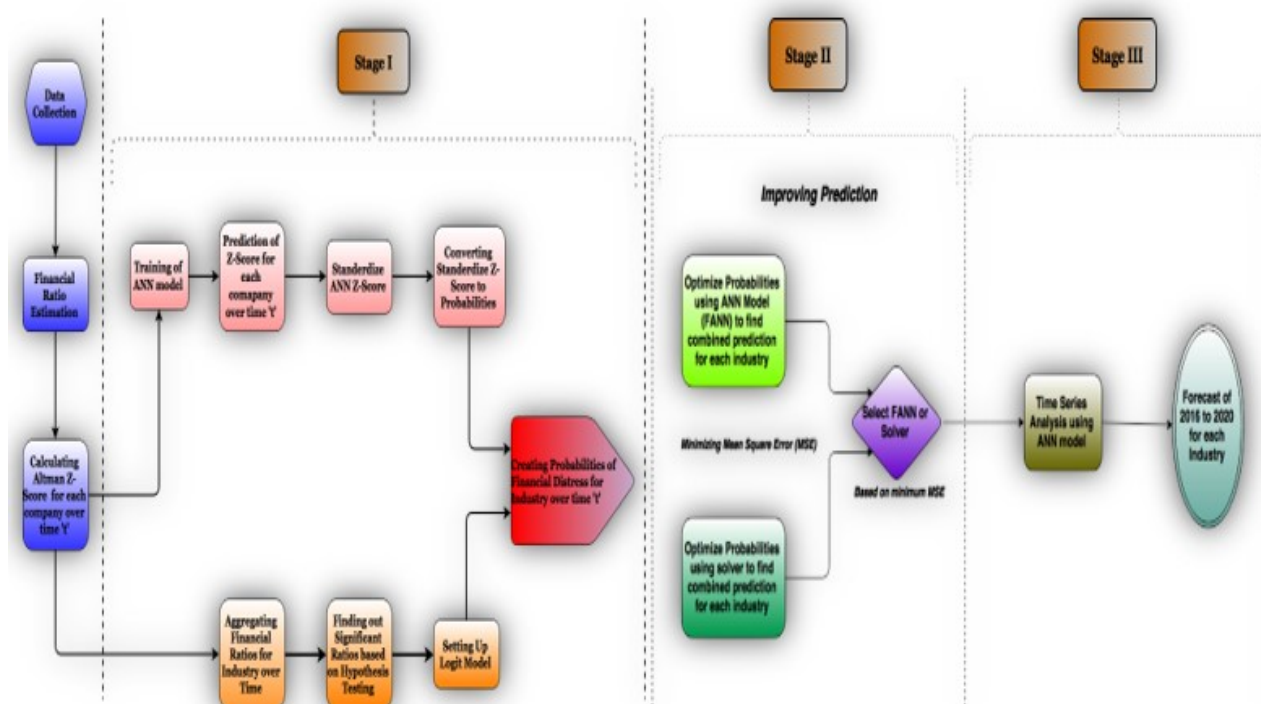
5.1   Stage I: Performances of ANN and Logit models

Elman feedforward backpropagation network used in Stage I consists of 1 input layer, 5 hidden layers and 1 output layer. The input layer consists of 20 nodes whereas output layer consists of 1 node. **Error! Reference source not found.**shows the training performance for the ANN model.

The performance curve shows that there are some overfittings as test curve increased before validation curve could increase. It also shows the performance of the Elman feedforward backpropagation ANN model for Energy Services Sector with training error (MSE) of 0.000005. The performance result does not show any problem in the training. The test curve and the validation curves are very similar to each other. There was some overfitting as test curve increased by some values before validation curve could increase. **Error! Reference source not found.**

below shows the graphical representation of the actual values of the Altman Z-Score and ANN model output for the Energy sector.

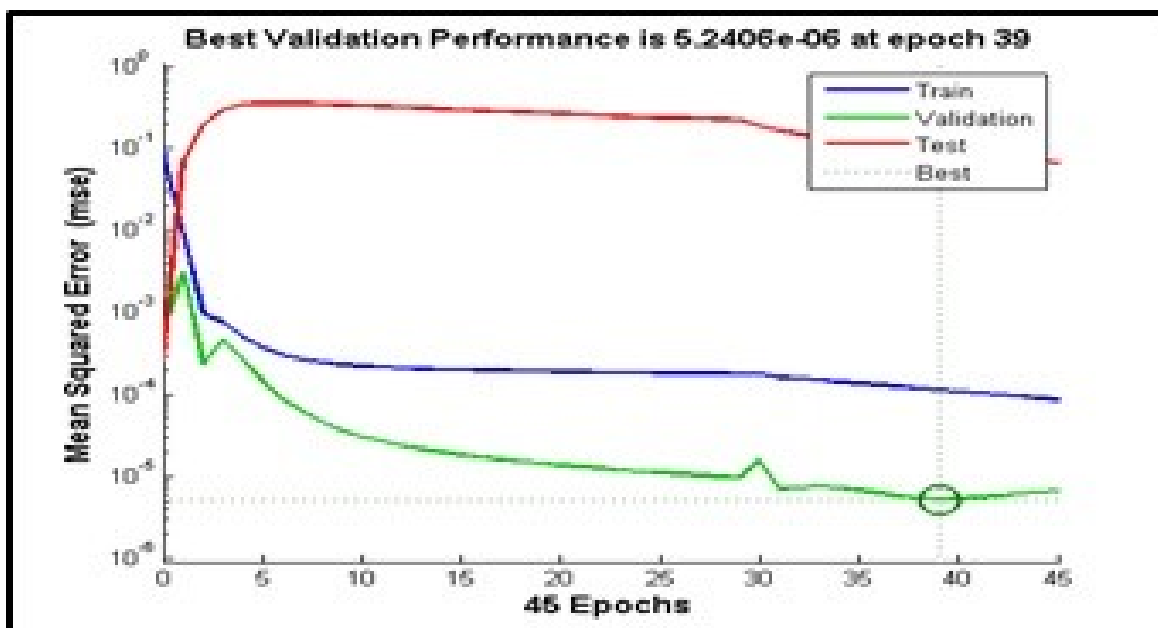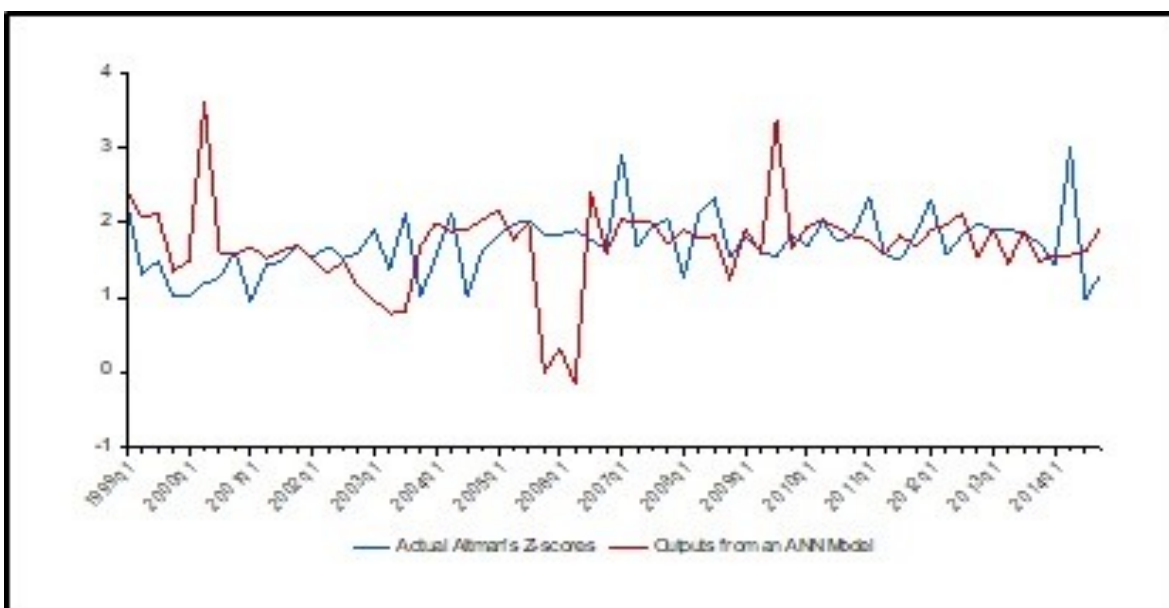Figure 2: ANN training performance for energy industry

Figure 3: Actual Altman Z-score and ANN output for energy industry

The aggregated values of the financial ratios and Altman Z-Score for each quarter for the industry are calculated and can be used as the inputs for the logit model. All Canadian industries had suffered from the Global Financial Crises (GFC). Therefore, to incorporate possible impacts of GFC, we included a dummy variable to capture this effect. Table 8 shows the descriptive statistics of the input variables of the Logit Model for Energy Industry. Table 9 shows the estimated output of the Logit model for Energy Services Industry. The estimated output shows that likelihood ratio test statistic is 25.758 (distributed chi-squared), with two degrees of freedom. The associated p-value is less than 0.001, indicating that the model with all seven predictors are overall significant. In the estimated output, the McFadden R-Squared value (0.293) shows that the independent variables used in the model can explain more than 29% of the characteristics of the dependent variable.

Table 3 shows the estimated outputs for the logit model in energy sector and Table 4 shows the accuracy of the prediction of energy industry by the estimated logit model in which the correct and incorrect classification based on the prediction rule along with the expected value calculation have been calculated. The classification is based

on the cut-off value of 0.5 for the classifying the predicted probability of y=0 and y=1. In Energy Services industry, 21 of the Y=0 and 30 of the Y=1 have correctly been estimated by the Logit model (63.94% of the Y=0 and 71.95% of the Y=1 observations are correctly classified, and overall accuracy of 68.45%). Table 5 shows the cut-off value for the industry for the probabilities of the financial distress. The cut-off values are generated based on Altman's financial distress criteria.

Table 2: Descriptive statistics of the variable used in the logit model for canadian energy sector

Variables	Mean	Med	Max	Min	Std. Dev.	Skew	Kurtosis
Y	0.563	1.0	1.0	0.0	0.5	-0.252	1.063
TDTS	2.028	2.051	3.200	1.255	0.380	0.245	3.577
RETA	-0.055	-0.037	0.379	-1.201	0.294	-2.093	8.696
QATS	2.227	1.580	26.087	0.976	3.115	7.193	55.554
QATA	0.275	0.276	0.347	0.196	0.028	0.134	3.358
NITA	0.007	0.011	0.049	-0.047	0.021	-0.586	3.459
MVETD	2.086	1.992	5.073	1.353	0.599	2.785	13.792
CATA	0.347	0.351	0.410	0.270	0.029	-0.167	2.675

Source: Authors' Calculations

Table 3: Estimation output of the Logit model for energy services industry.

Dependent Variable: Y

Variable	Coefficient	Std. Error	z-Statistic	Prob.
TDTS	3.110311	1.160401	2.680376	0.0074
RETA	2.579156	1.420000	1.816307	0.0693
QATS	-0.218982	0.080810	-2.709834	0.0067
QATA	54.25873	28.18970	1.924771	0.0543
NITA	9.475975	22.62185	0.418886	0.6753
MVETD	0.724876	0.546168	1.327203	0.1844
CATA	-76.49619	29.36309	-2.605182	0.0092
C	4.739115	4.235605	1.118875	0.2632
Mean dependent variance	0.562500	S.D. dependent variance		0.500000
S.E. of regression	0.420888	Akaike info criterion		1.218150
Sum squared residual	9.920221	Schwarz criterion		1.488011
Log likelihood	-30.98081	Hannan-Quinn criteria		1.324462
Restricted log likelihood	-43.86011	Avg. log likelihood		-0.484075
LR statistic (7 df)	25.75859	McFadden R-squared		0.293645
Probability (LR stat)	0.000556			
Obs with Dep=0	28	Total observations		64
Obs with Dep=1	36			

Source: Authors' calculations

Table 4: Accuracy of the prediction of energy services industry by Logit model.

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	21	6	27	0	0	0
P(Dep=1)>C	7	30	37	28	36	64
Total	28	36	64	28	36	64
Correct	21	30	51	0	36	36
% Correct	75.00	83.33	79.69	0.00	100.00	56.25
% Incorrect	25.00	16.67	20.31	100.00	0.00	43.75
Total Gain*	75.00	-16.67	23.44			
Percent Gain**	75.00	N/A	53.57			
	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	17.90	10.10	28.00	12.25	15.75	28.00
E(# of Dep=1)	10.10	25.90	36.00	15.75	20.25	36.00
Total	28.00	36.00	64.00	28.00	36.00	64.00
Correct	17.90	25.90	43.81	12.25	20.25	32.50
% Correct	63.94	71.95	68.45	43.75	56.25	50.78
% Incorrect	36.06	28.05	31.55	56.25	43.75	49.22
Total Gain*	20.19	15.70	17.67			
Percent Gain**	35.89	35.89	35.89			

*Change in "% Correct" from default (constant probability) specification

**Percent of incorrect (default) prediction corrected by equation

Source: Authors' calculations

5.2   Stage II: Improving the predictions

In search for improving the prediction of the outputs found in Stage I, a Forecast Artificial Neural Network (FANN) was the first method used to optimize and to combine the results of the logit and ANN models found in Stage I. The neural network applied in FANN model was Elman Back Propagation technique proposed by (Brandt, et al., 1981) and to increase the performance and avoid convergence problem. For training purpose, random weighted probabilities are generated using the actual ANN and logit results. These random probabilities as a desired output for the randomize inputs are fed into the model. Once the training of the ANN model completed, the input of out of the ANN and logit results fed to obtain the optimized single quarterly probabilities of the financial distress for Canadian energy industry. Alternatively, the solver optimizing algorithm was used to minimize the mean square error.

The outputs then generated from both techniques and the mean square error was compared. The results that had the minimum mean square error was used to conduct a time series analysis in the next Stage. For the FANN model, the neural network consisted of 1 input layer, five hidden layers, and one output layer. The inputs of the model were the outputs of the ANN and logit models, and the desired outputs were a weighted average of both models. These weights were randomly generated to avoid any biases towards the final output. **Error! Reference source not found.** shows the training performance of the backpropagation model with the training error of 0.00657 and **Error! Reference source not found.** shows the regression plot for the training of this model.

Error! Reference source not found. shows the graphical representation of the output generated by the logit and FANN models for the Energy sector. It is evident that the solver model comes with the lowest MSE compared to other models. The MSE of the logit model is relatively high compared to other models. Therefore, in Stage III, the outputs from the solver model will be used to forecast the likelihood of financial distress for the period of 2015 to 2020.

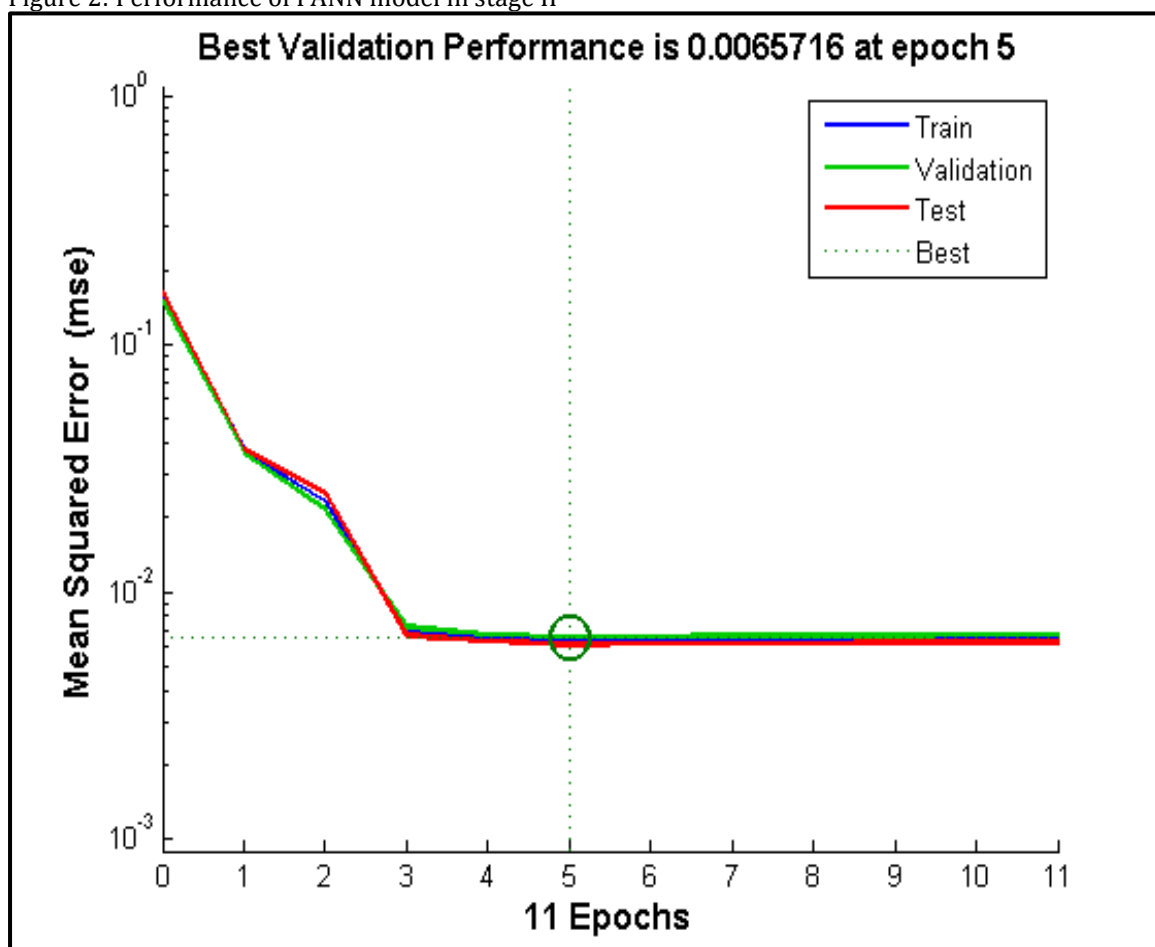Figure 2: Performance of FANN model in stage II

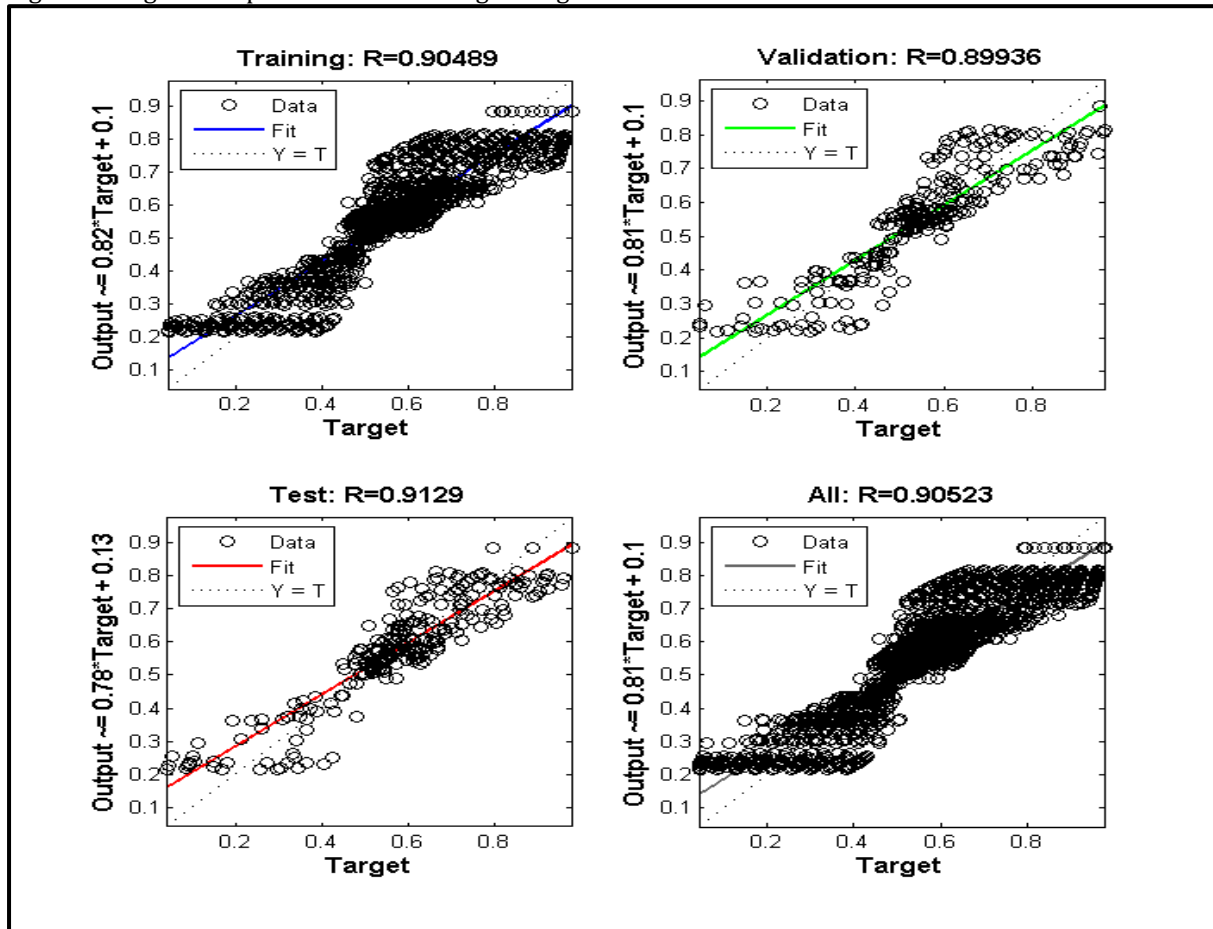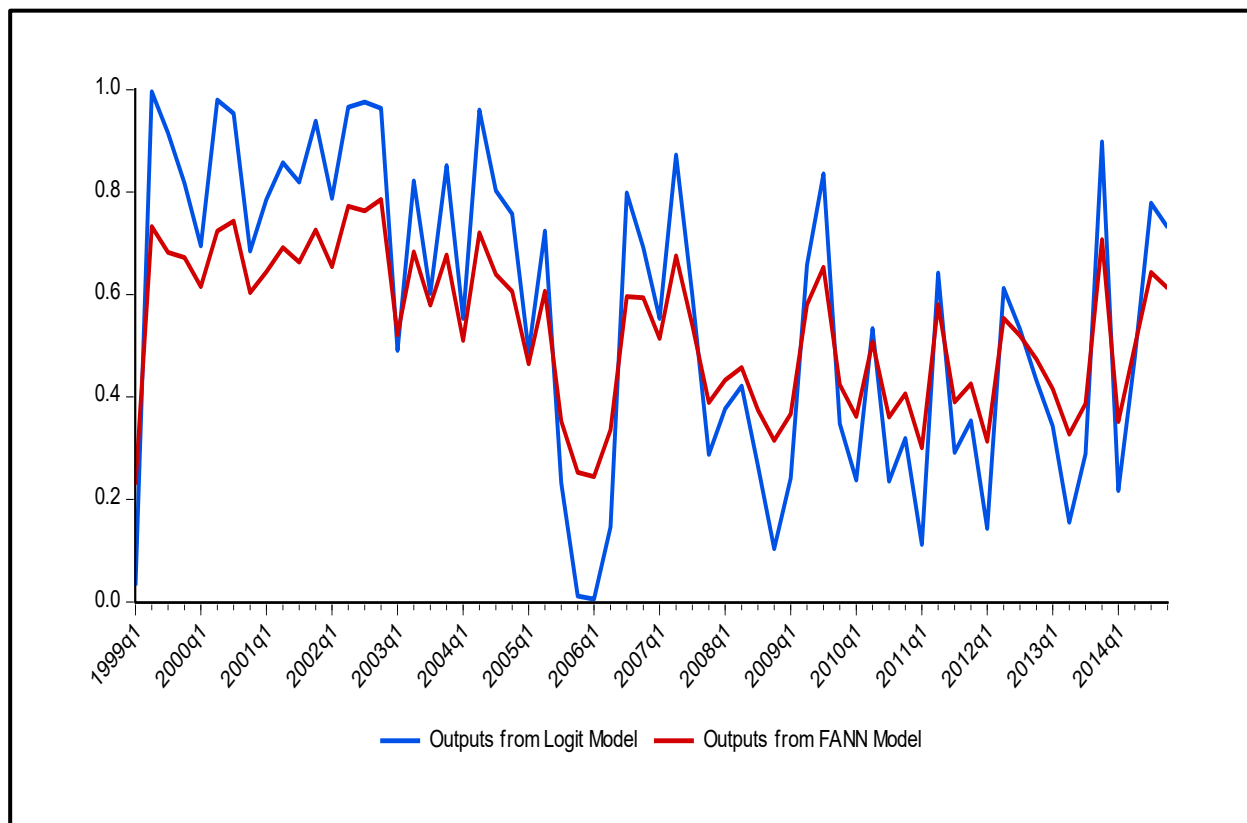Figure 3: Regression plot of FANN training in stage II

Figure 4: Comparison of actual and predicted outputs for the energy industry.

5.3   Stage III: Final Forecasts

Time series analysis was performed using the ANN model. The input for the ANN model is time whereas the corresponding desired outputs are the results of the solver model found in Stage II for the purpose of training an ANN model. **Error! Reference source not found.** shows the performance of the model at the time of training for Canadian Energy industry. **Error! Reference source not found.**(below) represents outputs of the solver model and the forecast values obtained by time series model for Canadian Energy industry.
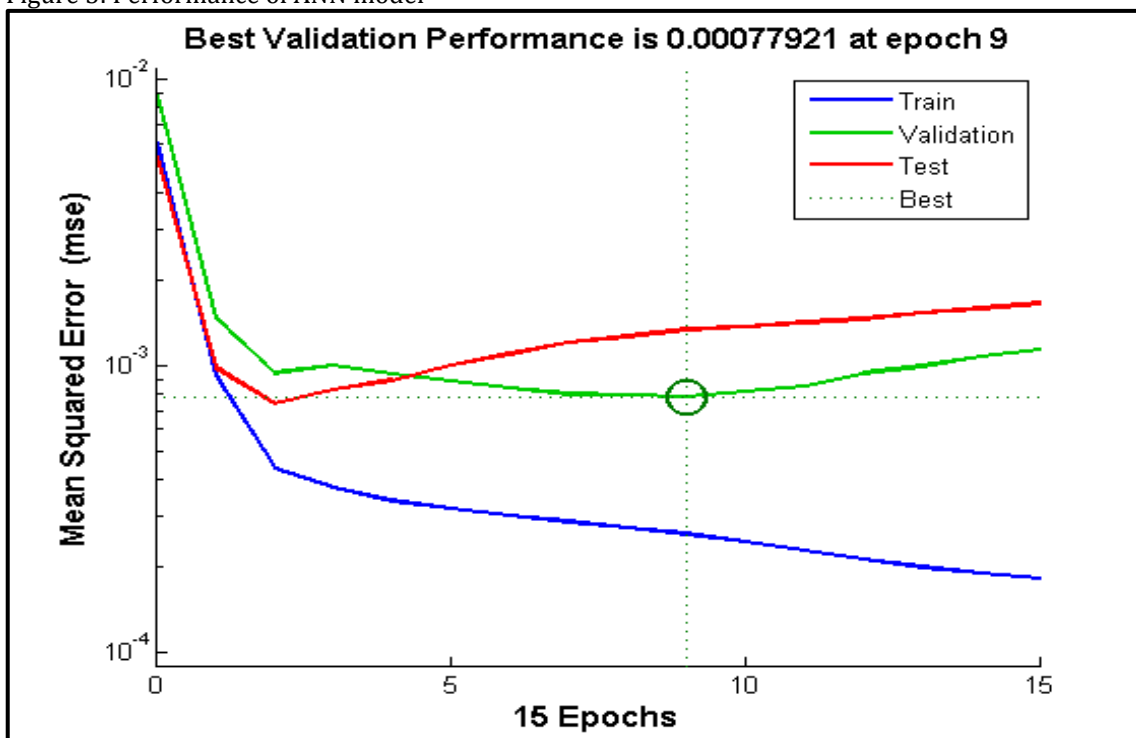
Figure 5: Performance of ANN model

Table 5 represents the Mean Square Error (MSE) of all estimated models for the Canadian energy Industry. The forecast improvement is evident that the final proposed optimized model is more accurate than all of the other examined models in the research. In fact, the final model increased the accuracy of the forecasts out of the ANN model by almost 19%.

Table 5: Mean square error (MSE) of different models for Canadian energy industry

Estimated model	MSE
ANN Model	0.0032
Logit Model	0.0766
FANN Model	0.0185
Combined Solver (Optimized) Model	0.0026

Source: Authors' calculations

6.   Business Implications

In this section, the predicted likelihoods of financial distress by the proposed three-stage model will be used to distinguish and to describe significant time periods of the Canadian energy industry in the past 15 years. In addition, an outlook of this industry by year 2020 will be pictured. The relation of the estimated likelihoods can be analyzed on the basis of the number wells drilled from 1999 to 2014. The drilling activities throughout the decade were affected the financial situation of the firms. The activities of the drilling are directly dependent on the cash flows and the capital expenditure made in the energy industry (The Fiscal Pulse of Canada's Oil and Gas Industry: First Quarter 2015, 2015). In 2001, the drilling activities fell due the economic recession caused by the dot-com bust experienced in Canada (Mearns, 2013).

Figure 6: Outputs out of the Optimized Model and its Predictions
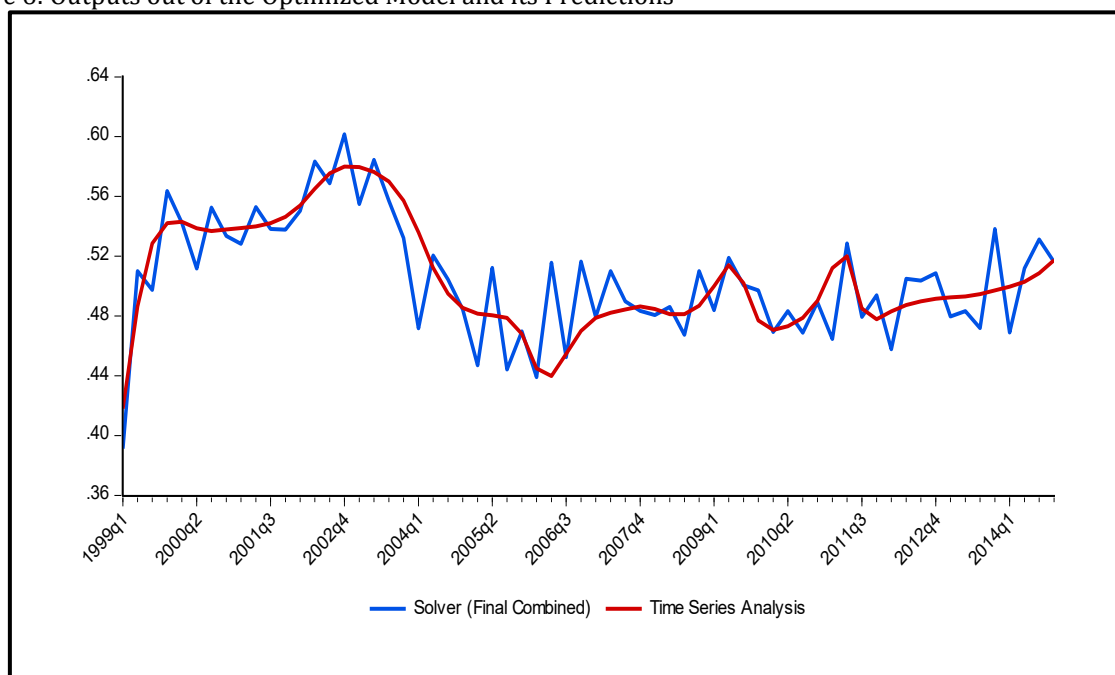
Three distinguished periods can be recognized for Canadian energy industry over the last 15 years. Figure 7 is a graphical representation of the final predicted values of the likelihood of financial distress for Canadian Energy industry. The outlook beyond 2015 on the graph indicates the forecast values of the likelihood of financial distress as a result of the developed model in the research to the last quarter of 2020. As illustrated, the first range is from 1999 to 2002, the second period covers 2003 to 2008, and the third period is from 2009 to 2014. The cut-off value associated with the financial distress is calculated based on the adjusted Altman's Z-score discussed earlier, and is found to be 0.48856. For all periods, in which the probability stays above this cut-off point, it is expected that Canadian energy market to face severe financial distress. The selected companies in our analysis are service providers to all oil and gas producers in Canada. Majority of these firms has the core competencies in providing drilling services to this industry. Another relevant core competency is to provide services such as infrastructure and other necessary services like pipelines, etc. to the oil rigs. The dip in the drilling activities is evident from the

increase in the financial distress probability from 2001 to 2002 as supported by the predicted likelihoods of the first period shown in Figure 7: Predicted and forecasts of probabilities of financial distress in Canadian energy sector.

Figure 7: Predicted and forecasts of probabilities of financial distress in Canadian energy sector
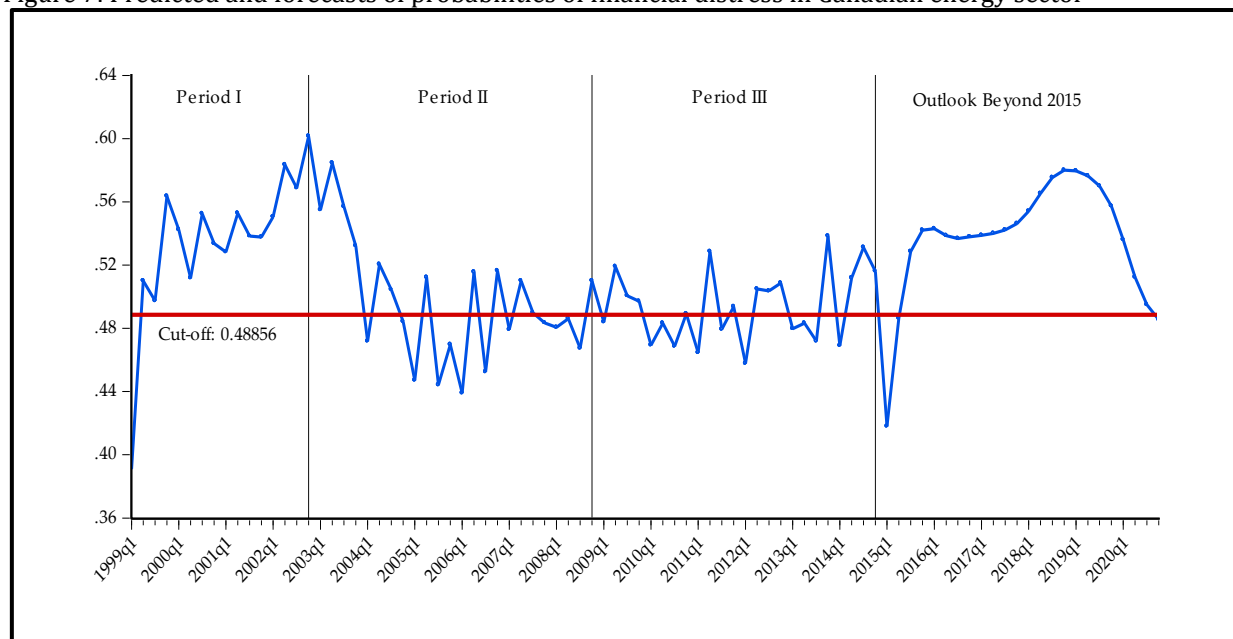
The distress probability reached to 58% due to falling oil prices and dot-com bust experienced. The reduced activities led to the reduction of the profits for the firms that has the core competencies in drilling services for the oil rig. There was a constant increase in oil and gas drilling from 2005 to 2006 (Petroleum Resources Branch, 2011). The rise in drilling activities is evident from the predicted distress probabilities shown in the second period of Figure 7: Predicted and forecasts of probabilities of financial distress in Canadian energy sector, as there was an enormous dip from 49% to 41%. Also, there was a high trend in the capital spending in the oil rig and energy industry which was increasing from 2005 to 2007 (Oil and Gas Fiscal Regimes Western Canadian Provinces and Territories, 2011). The rise of the capital investment was almost up to 50% that resulted in profitability of the drilling industry of Canada for oil and gas. The world was affected by the Global Economic Crises that started in the USA in late 2007. This crisis led to decrease in the energy use that eventually affected the drilling activities for the oil and gas (Industrial Energy Use in Canada: Emerging Trends, 2010). The ripple effect of this crisis has led to the lowest drilling activities in the history of Canadian energy sector. The reason for the lower drilling activities was due to the reduction in the capital investment as investors became more cautious about this market after the economic crisis (Oil and Gas Fiscal Regimes Western Canadian Provinces and Territories, 2011).

The third period shown in Figure 7: Predicted and forecasts of probabilities of financial distress in Canadian energy sector indicates a rapid increase in the probability of financial distress. One of the main reason for this increase from 2008 to 2009 was the lack of financial stability of the firms involved in the drilling activities (Oil and Gas Fiscal Regimes Western Canadian Provinces and Territories, 2011). The drilling firms were not able to complete the ongoing projects of drilling due to cut down of the capital investment (Oil and Gas Fiscal Regimes Western Canadian Provinces and Territories, 2011). The decline in the drilling activities led to a reduction in the profit and resulted in the increase in the likelihood of financial distress shooting to almost 53% in 2009. Then, there is a dip in the financial distress in 2010 but increased again in 2011 to the previous level of 53%. This time, the reason for the rise in the distress probabilities was due to the lower prices of the Canadian oil in the world market (Tertzakian, 2015). The rig count increased in 2011 but due to lower price of Canadian oil per barrel, all the supporting activities of the oil and natural gas producer reduced (Tertzakian, 2015). The trend of the lower oil prices has dragged to 2014 which supported by an increase in the predicted likelihood of the distress probabilities over time. Finally, consistent with the expectation of the Canadian Association of Oilwell Drilling Contractors (CAODC), a decline of the drilling activities by 10% in 2015 was expected (Hussain, 2014). According the outlook beyond 2015 shown in Figure 7, the likelihood of financial distress is going to increase sharply up until 2019. This trend also shows that the distress probabilities will peak to 58% by 2019, significantly above the cut-off level, after which the oil and natural gas producers will expect a decline in the likelihood of financial distress in this industry in late 2020.

7.   Conclusions

Nonparametric technique along with parametric and conventional statistical procedures are conglomerated to improve the forecast accuracy of the likelihood of the financial distress in Canadian oil and gas energy industry over the period of 1999-2014. In order to examine the current and future states of this industry, a three-stage robust hybrid forecasting model was developed to predict the likelihood of financial distress by end of 2020. Typically, financial distress is categorized in safe, distress, and danger zones based on the Altman's financial distress criteria. To guarantee a more conservative outlook, the first two categories were combined and the likelihood of financial distress for the selected companies divided into two categories; safe and dangerous zones. Nineteen of the most relevant financial ratios were used for each of the thirty selected Canadian oil and gas companies. All selected companies are publicly traded and listed on the Toronto Stock Exchange (TSX) from the first quarter of 1999 to the last quarter of 2014. From each of the representative company's income, balance sheet, and cash flow statements, nineteen crucial ratios were tracked back to the first quarter of 1999. We used the Altman Z-Score financial ratios along with other fifteen financial ratios to consider the complete financial stability of the selected firms. These inputs were used in both nonparametric and parametric models to categorize the financial health of the companies.

The optimized model could explain most of the economic variations that occurred from 1999 to 2014 in Canadian energy sector. The results indicate that the industry was relatively stable in last decade due to an impact of the changing economic conditions. It also showed a sign of improvement by the decrease in the value of the probability of financial distress in the mid of 2005 to 2007, but another major economic crisis again struck it back in early 2008. The results suggested that the energy supporting companies are going to have a hard time due to changing capital structure of investment and uncertainty of the oil prices by 2019. Finally, the results confirmed that the performance of the proposed three-stage hybrid forecasting model were superior to each of the individual nonparametric and parametric forecasting models with respect to the accuracy of the forecasts.

References

2013 Edition of the Branham300 – Canada's Top Tech Companies. (2013). (Branham Group Inc.) Retrieved from Branham: https://www.branham300.com/index.php?year=2013

ADVFN. (n.d.). Retrieved June 10, 2016, from http://ca.advfn.com/

Allaf, A., Omaina, N., & AbdAlKader, S. A., (2011). "Nonlinear Autoregressive neural network for estimation soil temperature: a comparison of different optimization neural network algorithms". *UBICC J*, 42-51.

Altman, E. I. (1968). Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 589-609.

Andres, J. d., Landajo, M., & Lorca , P. (2005). Forecasting business profitability by using classification techniques: A comparative analysis based on a Spanish case. *European Journal of Operational Research*, 518-542. Retrieved January 20, 2015, from http://www.sciencedirect.com/science/article/pii/S0377221704002267

Anjum, S. (2012). Business bankruptcy prediction models: A significant study of the Altman's X-Score model. *Asian Journal of Management Research 3.1*, 212-219. Retrieved from http://www.ipublishing.co.in/ajmrvol1no1/volthree/EIJMRS3018.pdf

Brandt, J., & Bessler, D. (1981). Composite Forecasting: An Application with U.S. Hog Prices. *American, 63*(1), 135-140.

Canada Economy - overview. (2015, June 30). Retrieved from O index mundi: http://www.indexmundi.com/canada/economy_overview.html

Deakin, E. B. (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*.

Dorsey, R. E., Edmister, R. O., & Johnson, J. D. (1995). *Bankruptcy Prediction Using Artificial Neural Systems.* The Research Foundation of The Institute of Chartered Financial Analysts. Retrieved from http://books.google.ca/books/about/Bankruptcy_prediction_using_artificial_n.html?id=3ndaAAAAYAAJ&redir_esc=y

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 179-211.

Frydan, H., Altman, E. I., & Kao, D. L. (1985). Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress. *The Journal of Finance 40*, 269-291.

Hauser, R. P., & Booth, D. (2011, September). Predicting Bankruptcy with Robust Logistic Regression. *Journal of Data Science*, 565-584. Retrieved January 20, 2015, from http://www.jds-online.com/files/JDS-716.pdf

Ibric, S., Djuris, J., Parojcic, J., & Djuric, Z. (2012). Artificial Neural Networks in Evaluation and Optimization of Modified Release Solid Dosage Forms. *Pharmaceutics*, 531–550. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3834927/

____(2010). *Industrial Energy Use in Canada: Emerging Trends.* National Energy Board.

Industry sector data. (2015, February 27). Retrieved from Government of Canada: http://www.canadabusiness.ca/eng/page/2827/

King, G. (1998). Unifying Political Methodology: The Likelihood Theory of Statistical Interference. New York: Cambridge University Press.

Krose, B., & Smagt, P. V. (1996). *An Introduction to Neural Network* (8th ed.). The University of Amsterdam. Retrieved from http://www.infor.uva.es/~teodoro/neuro-intro.pdf

Li, F., & Tkacz, G. (2004). Combining Forecasts with Nonparametric Kernel Regressions. *Studies in Nonlinear Dynamics and Econometrics*, Article 2.

Long, S. J. (1997). Regression Models for Categorized and Limited Dependent Variables. Sage Publications.

Mearns, E. (2013, January 30). *Oil Production Remains the Same, but Rig Counts Continue to Grow*. Retrieved from OILPRICE.com: http://oilprice.com/Energy/Crude-Oil/Oil-Production-Remains-the-Same-but-Rig-Counts-Continue-to-Grow.html

____(2011). *Oil and Gas Fiscal Regimes Western Canadian Provinces and Territories.* Canada's New West Partnership. Retrieved from http://www.energy.alberta.ca/Tenure/pdfs/FISREG.pdf

Petroleum Resources Branch. (2011). *Canadian Crude Oil, Natural Gas and Petroleum Products: Review of 2009 & Outlook to 2030.* Natural Resources Canada. Retrieved from http://www.nrcan.gc.ca/sites/ www.nrcan.gc.ca/files/energy/pdf/eneene/sources/crubru/revrev/pdf/revrev-09-eng.pdf

____ (2008). *Quarterly Monitor of the Canadian ICT sector - Fourth Quarter 2007.* Canada Industry.

____ (2011). *Quarterly Monitor of the Canadian ICT sector - Fourth Quarter 2010.* Canada Industry.

____ (2008). *Quarterly Monitor of the Canadian ICT sector - Second Quarter 2008.* Canada Industry. Retrieved from http://publications.gc.ca/collections/collection_2009/ic/Iu62-1-2008-2E.pdf

Shrivastava , A., & Bhandakkar, A. (2013). Short- Term Load Forecasting Using Artificial Neural Network techniques. *Int. Journal of Engineering Research and Applications, 3*(5), 1524-1527. Retrieved from http://www.ijera.com/papers/Vol3_issue5/IV3515241527.pdf

Tan, C. N. (1997). *An Artificial Neural Networks Primer with Financial Applications Examples in Financial Distress Predictions and Foreign Exchange Hybrid Trading System.* Bond University.

____ (2015). *The Fiscal Pulse of Canada's Oil and Gas Industry: First Quarter 2015.* ARC financialcorp energy capital partners. Retrieved from http://www.arcfinancial.com/assets/672/2015_Q1_Industry_Fiscal_Pulse _-_ARC_Financial.pdf

Whittington, D., Briscoe, J., Mu, X., & Barron, W. (1990). Estimating the willingness to pay for water services in developing counrties: A case study of the use of contingent valuation surveys in Southern Haiti. *Economic Development and Cultural Change*, 293-311.

Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 59-82.

Appendix:

Table 6: Financial ratios used in this study

Financial Ratio	Definition
Cash/TA	Cash/ Total Assets
Cash/TS	Cash/ Net Sales
CF/TD	Operating cash flow ratio = Cash Flow of Operating Income/Total Liabilities
CA/CL	Current Ratio = Total Current Assets/Total Current Liabilities
CA/TA	Current Asset/Total Asset
CA/TS	1/ current asset turnover = Total Current Assets/Net Sales
EBIT/TA	Return on total assets = (Interest Expense + Income Before Tax)/Total Assets
Log (INT + 15)	Log ((EBIT/Total Assets) +15)
Log (TA)	Log (Total Assets)
MVE/TK	Shareholder Equity/(Total Assets – Total Current Liabilities)
NI/TA	Return on Assets = Net Income/Total Assets
QA/CL	Quick Ratio = (Total Current Assets – Inventories)/Total Current Liabilities
QA/TA	(Total Current Assets – Inventories)/Total Assets
QA/TS	(Total Current Assets – Inventories)/Net Sales
RE/TA	Retained Earnings/Total Assets
TD/TA	Total Liabilities/Total Assets
TS/TA	Asset Turnover Ratio = Net Sales/ Total Assets
WK/TA	Working Capital / total assets
WK/TS	1/ Working Asset Turnover = 1/ (Net Sales/ Working Capital)

Source: Dorsey, Robert E, Robert O Edmister and John D Johnson. Bankruptcy Prediction Using Artificial Neural Systems. The Research Foundation of The Institute of Chartered Financial Analysts, 1995.