Comparative Performance of Feed Forward Neural Networks and Multiple Regression Using Simulation

Mukta Paliwal[1], Usha A. Kumar[2]

Shailesh J. Mehta School of Management, Indian Institute of Technology, Bombay, Mumbai-400 076, India

[1]mukta.paliwal@iitb.ac.in; [2]usha@som.iitb.ac.in

***Abstract*- Neural network models seem to have potential for prediction purposes and this has led to a number of studies comparing the performance of neural networks and regression analysis. Regression technique is based on certain basic assumptions and validity of these assumptions is critical to its performance. This issue does not appear to have been considered in most of the comparative studies. In the present study, we intend to focus on this aspect by comparing the performance of both the techniques using simulation when all the assumptions of regression are met. This study reveals that the performance of regression analysis and neural network are comparable for large and medium sample sizes and suggests the need for careful implementation of neural network when the sample size is small.**

***Keywords*- Assumption Validity; Evaluation Criterion; Levenberg-Marquardt Training Algorithm; Prediction Intervals; Simulation**

I. INTRODUCTION

Neural networks and Regression analysis have been studied and compared by various authors (eg. Korkut et al. 2011; Guo, W.W. (2010); Yilmaz, I. (2009); Liu, X. et al. (2009); Pendharkar, 2006; Hardgrave et al. 1994). However, most of these comparative studies focus on specific applications where the main objective is to achieve best results on a given application data set. As the application data sets are generally fixed and their characteristics are unknown, it is difficult to evaluate the effect of various data characteristics like sample size, number of variables, noise, etc. on the performance of the technique under study. Simulation allows better assessment of the performance of analytical techniques by generating data with the required characteristics especially when the objective is to compare the performances of the two techniques.

Few studies in literature (eg. Marquez, Hill, Worthley and Remuse, 1991; Marquez and Hill, 1993; Cherkassky, Gehring and Mulier, 1996; Banks, Olszewski and Maxion, 2003) have used simulated data sets for studying comparative performance of various techniques. Marquez, Hill, Worthley and Remuse (1991) have used simulation to carry out an empirical study for evaluating the performance of regression and neural networks by varying some data characteristics like functional form of the linear model, amount of noise and sample size of the data. The results of this study indicate that neural network models perform best under conditions of high noise and low sample size and in case of less noise and larger sample sizes, neural network becomes less competent. Also, Marquez and Hill (1993) have conducted simulation study where they have compared the performance of back_propagation neural network and general regression neural network in approximating various functions in the presence of noise. They have concluded that back propagation model showed significant advantage over general regression model in approximating functions under condition of low random noise. The accuracy in the performance of the back propagation model diminishes as the level of random noise increases. This observation limits the claim that neural networks are more robust in case of noise in the data.

Cherkassky, Gehring and Mulier (1996) have compared six adaptive methods (Nearest neighbour, generalized memory based learning, projection persuit, artificial neural networks, multivariate adaptive regression splines and constrained topological mapping) for function estimation by creating artificial datasets to provide some insights on applicability of these methods. No single method proves to be the best among the methods considered in this study. They have observed that statistical methods using greedy and fast optimization procedures tend to be less robust than neural network methods using iterative optimization methods for parameter estimation. Banks, Olszewski and Maxion (2003) have presented extensive simulation experiment comparing the performance of ten different multivariate nonparametric regression techniques: linear regression, stepwise linear regression, additive models, projection pursuit regression, recursive partitioning regression, multivariate adaptive regression splines, alternating conditional expectations, additivity and variance stabilization, locally weighted regression, and neural networks. Datasets used in the experiment are constructed to have a range of characteristics by varying the dimension of the data, the sample size, the amount of noise, and the complexity of the embedded structure. Analyses of the results show that all of these properties affect the accuracy of each regression technique under investigation.

The studies carried out by Marquez, Hill, Worthley and Remuse (1991) and Marquez and Hill (1993) pertain to comparison of neural networks and regression analysis using simulated data sets but these studies are restrictive in terms of the sample size and the number of independent variables. The need for a more rigorous statistical comparison of neural networks with other traditional techniques has been emphasized by Sharda and Wilson (1993) as many studies in the literature have reported the superior performance of neural networks on an anecdotal basis leading to much skepticism. The statistical techniques are based on certain fundamental assumptions and this issue does not seem to have been considered in many of the past comparative studies involving neural networks and

statistical techniques (Paliwal and Kumar, 2009). Further, the comparative performance of the two techniques has not been studied properly with respect to a variety of data characteristics in the past literature.

In this paper, we intend to compare the performance of regression analysis and multi layered feed forward neural networks using simulation when all the assumptions of regression technique hold true. Data sets with different sample size, number of independent variables and amount of noise are simulated in order to carry out this comparison. In this study, Levenberg-Marquardt (LM) algorithm (Bishop, 1995) has been used for training neural networks though most of the past studies have used back propagation algorithm to train the networks. Due to the slow convergence rate of back propagation algorithm, there seems to be an increasing interest in the neural network research community for new and improved training methods. In recent years, the LM algorithm is becoming increasingly popular and considered to be an efficient algorithm while training neural networks that have up to a few hundreds of weights (Hagan and Menhaj, 1994). Appropriate experimental design is used to compare the predictive accuracy of regression analysis and multi layered feed forward neural networks techniques and mean squared prediction error is used as the performance evaluation criterion. Further, asymptotic prediction intervals for neural networks are also calculated and are graphically compared with the prediction intervals of regression analysis.

In Section II, the experimental design, the process of simulation and the prediction intervals are discussed. Section III presents the analysis and the results of the data analysis are given in Section IV. Section V presents the conclusion of this study.

II. METHODOLOGY

The data sets are simulated using Monte Carlo simulation such that all the assumptions of multiple regression model are satisfied. Factors explored in this study are (1) number of independent variables (2) sample size (3) levels of noise and (4) methods of analysis. The effect due to these four factors and their interaction effects will be analyzed using appropriate experimental design. The following subsections describe the experimental design, the process of simulation and the prediction intervals.

A. Experimental Design

The experiment is a 3 x 3 x 3 x 2 factorial design in which there are repeated measures on the last factor. The four factors considered in this study are the number of independent variables, sample sizes, levels of noise and methods of analysis. The choice of the levels of these three factors depends on previous literature and also on applications that arise in practice. The last factor represents the two methods of analysis, namely regression analysis and neural networks that are used to analyze the same data under different experimental conditions. Each of the experimental conditions is replicated 30 times to average the sampling fluctuations. The different levels (or values) that are considered for each factor in the experiment are as follows:

Number of Variables (p): three levels considered for the dimension of the number of independent variables are 2, 4 and 10.

Sample Size (n): the sample size is obtained using the formula given by Sawyer (1982).

$$n = \frac{2K^2 - 1}{K^2 - 1} + \frac{K^2}{K^2 - 1} p \quad (1)$$

where p is the number of independent variables in the model and K denotes the amount of inflation in the measure of performance due to estimating the coefficients and is called the inflation factor. Three values of this inflation factor are used to get small, medium and large samples corresponding to each level of number of independent variables p. Table 1 shows the choice of n for different values of p and K.

TABLE I SAMPLE SIZES FOR DIFFERENT VALUES OF P AND K.

K	***p* = 2**	***p* = 4**	***p* = 10**
1.1	18	30	65
1.01	150	255	560
1.003	500	840	1840

Levels of Noise: Signal-to-noise ratio (SNR) is used to measure the variation present in the data generated. The squared length of the regression parameter vector is referred to as Signal-to-noise ratio. The chosen values of SNR are 1, 4 and 25 corresponding to high noise, medium noise and low noise respectively.

Methods of Analysis: To compare the performance of the two techniques, least squares multiple regression analysis and multilayer feed forward neural network using Levenberg-Marquardt training algorithm are used.

B. Process of Simulation

Data matrices containing the required number of independent variables (Xi) were generated using SAS IML (SAS Institute, Inc. 2007). The number of design cells in this experiment is 27 pertaining to our choice of independent variables, sample size and random noise, all being at three levels. The number of replications chosen for this experiment is 30. Accordingly, thirty data matrices were generated for each of the design cells and analysed by both regression analysis and neural network technique. The data sets are simulated from the model of the Form (2) such that all the assumptions of the multiple regression model hold true.

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_j X_{ji} + ... + \beta_p X_{pi} + \varepsilon_i, \quad \text{for } i = 1, 2, ..., n \text{ and } j = 0, 1, 2, ..., p \quad (2)$$

where Y_i is the dependent variable, $X_{ji}s$ are p independent variables and generated from Normal distribution with mean zero and variance 1, ε_i is the random error component generated from Normal

distribution with mean zero and constant variance σ^2 and $\beta_j s$ are the parameters of Model (2). Using Delaney and Chatterjee (1986), the regression coefficients are computed as

$$\beta_j = \left(\|\boldsymbol{\beta}\| / r_u\right) u_j \quad (3)$$

where $r_u^2 = \sum u_j^2$ and $u_j, j = 1,2,....p$ are uniform random numbers in the interval [-10,10] in order to have enough variation in the magnitude of individual coefficients in the regression analysis. The squared length of the regression parameter vector ($\|\boldsymbol{\beta}\|^2 = \boldsymbol{\beta}'\boldsymbol{\beta}$) chosen for this study is 100 and is a measure of the strength of the signal. The error variances are computed for the selected value of $\|\boldsymbol{\beta}\|^2$ (=100) and the SNR (=1, 4 and 25) by using the relation SNR= $\boldsymbol{\beta}'\boldsymbol{\beta}/\sigma^2$ resulting in σ^2 = 4, 25 and 100 corresponding to the three levels of SNR. Descriptive statistics were examined on the resulting data matrices to verify the desired data characteristics. The resultant data are analysed using both regression and neural network in order to compare the techniques when all the assumptions of regression model are satisfied. To obtain asymptotic prediction intervals, data sets containing 1000 observations were generated for each of the levels of number of independent variables and the levels of noise.

C. Prediction Intervals

This subsection describes the prediction intervals for predicting the future observation Y_0 using the regression model and the neural network model.

Prediction Interval for the Regression Model:

An appropriate $100(1-\alpha)\%$ prediction interval for Y_0 is given by

$$[\hat{Y}_0 \pm t_{\alpha/2}^{(n-p-1)} s\sqrt{1+\mathbf{X_0'(X'X)^{-1}X_0}}] \quad (4)$$

where, the columns in the matrix X contain the observed values of the independent variables corresponding to the *j* parameters $\beta_0, \beta_1, ..., \beta_p$, $\mathbf{X_0} = [1\ x_{10}\ x_{20} \ldots x_{k0}]$ is a row vector containing the values of independent variables for the new observation Y_0 and $t_{\alpha/2}$ denotes the $\alpha/2$ quantile of a t distribution with $n-p-1$ degrees of freedom and $s\sqrt{1+\mathbf{X_0'(X'X)^{-1}X_0}}$ is the standard error of the prediction error $(Y_0 - \hat{Y}_0)$.

Prediction Interval for the Neural Network Model:

Asymptotic prediction intervals for multilayer feed forward neural network based on certain assumptions have been obtained by Hwang and Ding (1997). This approach is used to construct the prediction intervals as detailed below:

Let us assume that the observations $(\mathbf{X}_i, Y_i), i = 1,2,...,n$ satisfy the model

$$Y_i = g_\theta(\mathbf{X}_i) + \varepsilon_i \quad (5)$$

where Y_i is the output of the neural network $g_\theta(\mathbf{X}_i)$ and ε_i is the error. Further assume that Y_0 is the future observation that satisfy

$$Y_0 = g_\theta(\mathbf{X}_0) + \varepsilon_0 \quad (6)$$

Assuming ε_0 is normally distributed, the asymptotic prediction interval for the observation Y_0 can be constructed as

$$g_{\hat{\theta}}(X_0) \pm t_{1-\alpha/2, n-(p+2)h-1}\, \hat{\sigma}\sqrt{1+S(\hat{\theta})} \quad (7)$$

with asymptotic coverage probability $(1-\alpha)$. Here, $t_{1-\alpha/2, n-(p+2)h-1}$ denotes the $1-\alpha/2$ quantile of a t-distribution with $n-(p+2)h-1$ degrees of freedom,

$$\hat{\sigma}^2 = \frac{1}{n-(p+2)h-1}\sum_{i=1}^{n}\left(Y_i - g_{\hat{\theta}}(\mathbf{X}_i)\right)^2, \text{ and}$$

$$S(\hat{\theta}) = \frac{1}{n}\left\{\left[\nabla_\theta g_\theta(X_0)\right]_{\theta=\hat{\theta}}^t \hat{\Sigma}^{-1} \left[\nabla_\theta g_\theta(X_0)\right]_{\theta=\hat{\theta}}\right\},$$

where

$$\hat{\Sigma}(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left[\nabla_\theta g_\theta(X_i)\nabla_\theta g_\theta(X_i)^t\right]_{\theta=\hat{\theta}},$$

p is the number of input nodes and h is number of neurons in the hidden layer and it is assumed to be known.

III. DATA ANALYSIS

Initially, data sets containing twice the sample size given in Table 1 were generated for each of the design cells. One half of this data set was used to train the data and the other half was used to validate the trained model and to compare the predictive performance of both the techniques. These two data sets are referred as training sample and hold out sample respectively. The procedure mentioned above is replicated 30 times. Replication refers to the repetition of the experiments at each combination of levels of different factors. This provides a precise estimate of the sample mean of the factors of the experiment. Analysis of variance for this repeated measure factorial experiment is carried out in order to evaluate significant differences in the performance of the two techniques with respect to various design factors.

Appropriate prediction intervals are also constructed to predict the future observation Y_0 from the regression model and the neural network model for all the designs corresponding to each of the three levels of number of independent variables and three levels of noise. Asymptotic prediction intervals for the multilayer feed forward neural network are graphically compared with the prediction intervals of regression analysis and the results are presented in the next section.

A. Performance Evaluation Criterion

In order to compare the performance of these two techniques, mean squared error is used as the performance evaluation criterion.

Mean square error (MSE) is an unbiased estimate of σ^2 and is given by

$$s^2 = \frac{1}{n-p-1}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (8)$$

This estimate is appropriate to be used on the training data but provides an underestimate when used on a hold out sample. An appropriate measure to be used on test data is mean squared prediction error (MSPR) given by

$$MSPR = \frac{1}{n^*}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (9)$$

where n^* is the number of observations in the test data set and where the predicted values $\hat{Y}_i$ are obtained using the trained model. This measure indicates the performance of the trained model on future observations.

Square root of the above measures, namely root mean square error (RMSE) and root mean square prediction error (RMSP) are respectively used for the training data and test data so that the error is expressed in the same units as the observations.

B. Neural Network Architecture Selection

Feed forward network architecture is considered and is trained using LM algorithm. The network was considered with one input layer and one output layer. Initial experiments were carried out to compare the performance of feed forward network with one, two and three hidden layers in order to decide the number of hidden layers for each of the 27 experimental conditions. No improvement in the prediction accuracy was observed with the addition of hidden layers. As a result, we have used only one hidden layer for all the networks trained in our experiments. The number of nodes in the input layer and the output layer corresponds to the number of independent variables (p) and one dependent variable respectively. The number of nodes in the hidden layer is usually determined by trial and error so as to obtain the best performance for the data under consideration. The activation functions that were used in the hidden layer and in the output layer are hyperbolic tangent function and identity function respectively.

The purpose of training neural networks is to build a statistical model that generalizes well to the new data. The process of regularization can be used to improve the generalization ability of the network. Regularization involves constraining or penalizing the solution of the estimation problem to improve the network generalization ability by smoothing the predictions (Girosi, Jones and Poggio, 1995; Wu and Moody, 1996). A simple form of regularization called "weight decay" as discussed in Bishop (1995) is used in the present study. Trial and error was used to choose the parameter values of the weight decay parameter and the number of nodes of the hidden layer. The values of weight decay parameter considered were ranging from no weight decay to a high value of 0.14 at intervals of 0.02 and the number of hidden nodes was varied from 1 to 10 for each of the 27 designs of the simulation experiment. Optimum generalization of the test data set was obtained with 1 node in hidden layer for all the experimental conditions. This could be because of the data being generated from a linear model and hence no obvious nonlinearities are present in the data to be trained. To choose these parameters, error values for the test data set corresponding to 30 replications are plotted as box plots shown in Fig. 1. For clarity of presentation, only a selected range of hidden units (1, 2, 3) and weight decay (WD= 0.02.0.04, 0.06 and 0.08) parameter are shown in these box plots.

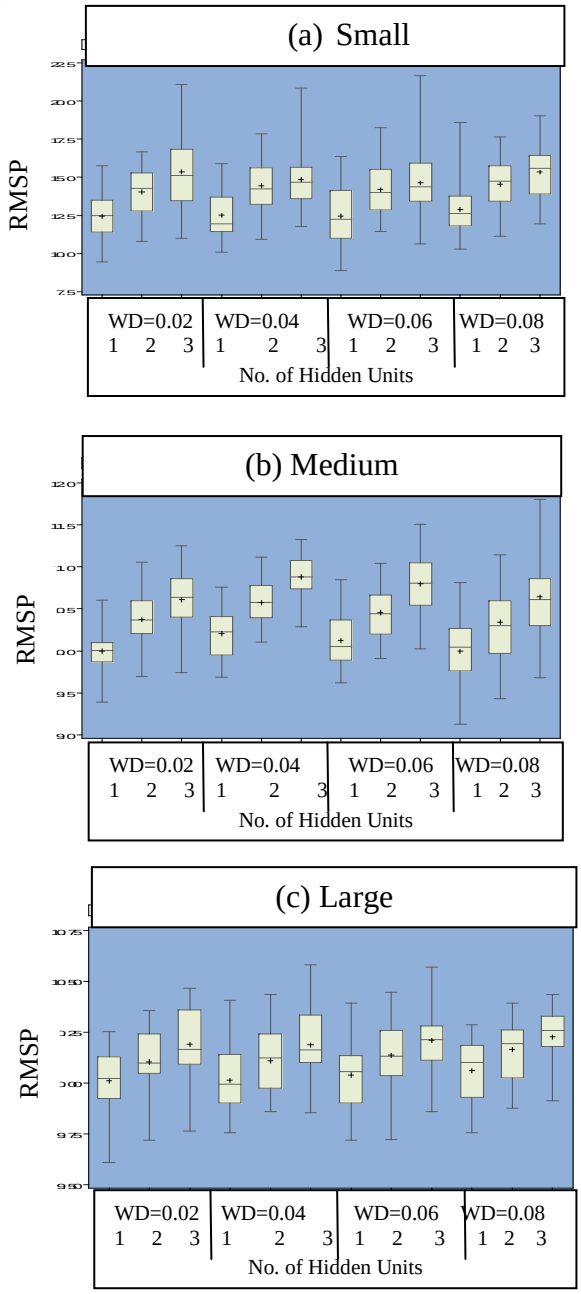


Figure1 Box Plots showing error values of test data set for 30 replications for selected data conditions

IV. RESULTS

To compare the performance of both the techniques, RMSE is calculated for the training data set and RMSP is calculated for the test data set for each of the design cells for 30 replications and mean and standard deviations of these error values are presented in Table 2. It can be observed from this table that the mean error for neural network and regression are almost the same for training data. However for the test data, mean error for the regression technique is always lesser than that of neural networks. As it was of interest to see how these two techniques perform when the sample size is quite high, relevant experiments are carried out and the results are shown in Table 2. In this regard, the sample sizes were determined using the same formula as given in eq. (1) with K= 1.0003 for p=2, 4 and 10 resulting in 10000, 16670, 36680 respectively. The performance of the two techniques can be found to be identical and hence this factor is not considered for further analysis.

TABLE II MEAN AND STANDARD DEVIATIONS OF ERRORS FOR TRAINING AND TEST DATA SETS

Sample Size	Variable	Measure	High Noise				Medium Noise				Low Noise			
			Training		Test		Training		Test		Training		Test	
			Reg	NN	Reg	NN	Reg	NN	Reg	NN	Reg	NN	Reg	NN
Small	2	**Mean**	**9.1602**	**9.2515**	**10.8997**	**12.0414**	**4.8849**	**5.2738**	**5.3113**	**5.7909**	**1.8021**	**1.9822**	**2.1123**	**2.7562**
		Std Dev	1.5701	1.8931	2.0902	2.4188	1.0196	1.0892	0.7444	1.7636	0.3081	0.3685	0.3336	0.7639
	4	**Mean**	**9.5479**	**9.2146**	**10.9671**	**12.8996**	**4.713**	**4.8591**	**5.5669**	**5.892**	**1.8684**	**2.0641**	**2.2047**	**2.637**
		Std Dev	1.1497	1.3421	1.6745	1.9083	0.6602	0.6469	0.688	0.8898	0.2131	0.2497	0.2983	0.5747
	10	**Mean**	**9.7636**	**9.6109**	**10.8553**	**12.22**	**5.0453**	**5.1545**	**5.6068**	**5.691**	**1.9427**	**2.0888**	**2.2208**	**2.3808**
		Std Dev	0.8617	0.7928	1.0246	1.7618	0.5358	0.5286	0.5102	0.5461	0.1924	0.1784	0.291	0.2762
Medium	2	**Mean**	**10.1255**	**10.1935**	**10.1233**	**10.1831**	**4.9891**	**5.0401**	**5.0173**	**5.0604**	**1.9879**	**2.0302**	**2.0014**	**2.0656**
		Std Dev	0.559	0.5798	0.6121	0.603	0.3106	0.3082	0.3565	0.3551	0.1172	0.123	0.1301	0.1513
	4	**Mean**	**9.9444**	**9.9826**	**9.983**	**10.0334**	**5.0385**	**5.0717**	**5.0665**	**5.1047**	**1.969**	**1.9996**	**2.0053**	**2.0412**
		Std Dev	0.4551	0.4459	0.3846	0.3694	0.2324	0.2395	0.2646	0.2655	0.086	0.0865	0.0878	0.0847
	10	**Mean**	**9.918**	**9.9321**	**9.9868**	**10.0038**	**4.99**	**5.0041**	**5.1216**	**5.1445**	**1.9994**	**2.0359**	**2.0159**	**2.0671**
		Std Dev	0.3362	0.3372	0.3576	0.3536	0.1481	0.148	0.1552	0.161	0.0632	0.068	0.0633	0.0746
Large	2	**Mean**	**9.9483**	**9.9707**	**10.0489**	**10.0753**	**4.9506**	**4.9719**	**5.0337**	**5.0557**	**1.9872**	**2.0188**	**2.0145**	**2.0435**
		Std Dev	0.3023	0.3047	0.2816	0.2863	0.1291	0.1296	0.1704	0.1707	0.0664	0.0665	0.0599	0.0665
	4	**Mean**	**10.0944**	**10.1106**	**9.9908**	**10.0081**	**4.9718**	**4.979**	**5.0131**	**5.0237**	**1.9945**	**2.0121**	**2.0047**	**2.0229**
		Std Dev	0.2293	0.2307	0.2327	0.2358	0.1139	0.1176	0.13	0.1281	0.0484	0.048	0.0699	0.0711
	10	**Mean**	**9.9763**	**9.9828**	**10.0627**	**10.0686**	**4.9921**	**4.9967**	**5.0114**	**5.0168**	**1.9853**	**1.997**	**2.0038**	**2.016**
		Std Dev	0.1605	0.1612	0.1394	0.1385	0.0988	0.0998	0.0693	0.0687	0.0336	0.0345	0.0392	0.0388
Very Large	2	**Mean**	**10.0284**	**10.0273**	**9.9762**	**9.9881**	**4.9931**	**4.9938**	**4.9909**	**4.9952**	**2.0053**	**2.0065**	**2.0006**	**2.0028**
		Std Dev	0.0977	0.0982	0.11	0.1104	0.0506	0.0512	0.0518	0.0531	0.018	0.0182	0.0247	0.0241
	4	**Mean**	**10.0026**	**10.0065**	**9.9886**	**9.9978**	**4.9902**	**4.9976**	**4.9902**	**4.9976**	**1.9999**	**2.0155**	**1.9999**	**2.0155**
		Std Dev	0.0549	0.0552	0.0907	0.0899	0.032	0.0323	0.032	0.0323	0.0125	0.0126	0.0125	0.0126
	10	**Mean**	**10.0001**	**9.9972**	**9.9863**	**9.9958**	**4.9902**	**4.9976**	**4.9902**	**4.9976**	**1.9999**	**2.0155**	**1.9999**	**2.0155**
		Std Dev	0.048	0.0478	0.0753	0.0731	0.032	0.0323	0.032	0.0323	0.0125	0.0126	0.0125	0.0126

Analysis of variance for the four factorial designs with repeated measure on the last factor is carried out to test the effects of these design factors (independent variable, sample size, noise levels and method of analysis) on the prediction performances. The validity of the usual F-tests in a repeated measure factorial experiment rests on the assumption of covariance matrices being homogeneous across the levels of the between subjects factors (compound symmetry) and variances for all pair wise differences between variables being equal (sphericity). As there are only two levels in the within subject factor, the current experiment does not require any sphericity test. The homogeneity of covariance matrices is tested using Box'M test and the null hypothesis is rejected. The empirical literature indicates that for a balanced design (group sizes are equal), the degrees of freedom adjusted univariate F-tests could be used to obtain a robust test of repeated measures main and interaction effect hypotheses even when the assumption of equality of the covariance matrices is not satisfied (Huynh, 1978; Keselman, Carriere and Lix, 1993). The results from univariate tests for within subject effects are summarized in

Table 3 and Table 4 respectively for training and test data sets.

TABLE III REPEATED MEASURES ANALYSIS OF VARIANCE (TRAINING DATA)

Tests of Hypotheses for Between Subjects Effects					
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
Var	2	0.94	0.47	0.82	0.4411
Noise	2	16854.68	8427.34	14652.1	<.0001**
Var*Noise	4	0.81	0.2	0.35	0.8424
Size	2	16.7	8.35	14.52	<.0001**
Var*Size	4	5.57	1.39	2.42	0.047
Noise*Size	4	25.51	6.38	11.09	<.0001**
Var*Noise*Size	8	6.76	0.84	1.47	0.1646
Error	783	450.35	0.58		
Tests of Hypotheses for Within Subject Effects					
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
Method	1	0.85	0.85	19.4	<.0001**
Method*Var	2	0.59	0.29	6.68	0.0013**
Method*Noise	2	1.04	0.52	11.79	<.0001**
Method*Var*Noise	4	0.29	0.07	1.67	0.1553
Method*Size	2	0.35	0.18	4	0.0187
Method*Var*Size	4	0.68	0.17	3.9	0.0039**
Method*Noise*Size	4	2.19	0.55	12.47	<.0001**
Method*Var*Noise *Size	8	0.55	0.07	1.56	0.1323
Error(Method)	783	34.4	0.04		

Note: ** means the corresponding effects are significant at 1% significance level.

TABLE IV REPEATED MEASURES ANALYSIS OF VARIANCE (TEST DATA)

Tests of Hypotheses for Between Subjects Effects					
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
Var	2	0.92	0.46	0.51	0.6
Noise	2	19656.73	9828.37	10949.50	<.0001**
Var*Noise	4	1.02	0.25	0.28	0.8887
Size	2	259.78	129.89	144.71	<.0001**
Var*Size	4	4.31	1.08	1.20	0.3087
Noise*Size	4	104.68	26.17	29.16	<.0001**
Var*Noise*Size	8	4.27	0.53	0.59	0.7827
Error	783	702.83	0.90		
Tests of Hypotheses for Within Subject Effects					
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
Method	1	27.96	27.96	135.73	<.0001**
Method*Var	2	1.13	0.57	2.74	0.07
Method*Noise	2	12.81	6.41	31.11	<.0001**
Method*Var*Noise	4	1.67	0.42	2.02	0.09
Method*Size	2	44.14	22.07	107.14	<.0001**
Method*Var*Size	4	1.86	0.46	2.25	0.06
Method*Noise*Size	4	25.49	6.37	30.94	<.0001**
Method*Var*Noise *Size	8	3.32	0.42	2.02	0.04
Error(Method)	783	161.28	0.21		

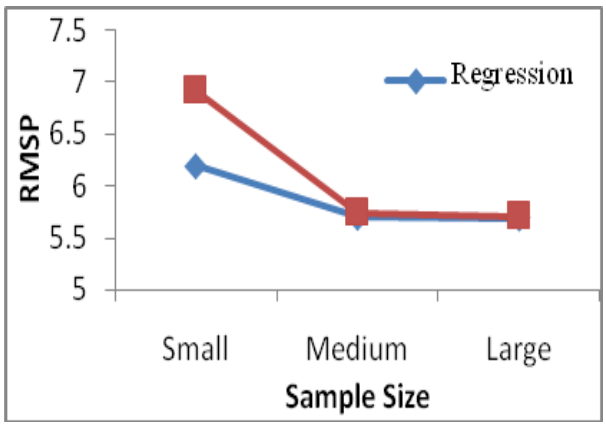
Note: ** means the corresponding effects are significant at 1% significance level.

As can be seen from between effects of the test data set (Table 4), two of the main effects, noise and sample size along with the two way interaction effect of noise × size are significant at 1% level of significance. From the within effects, it can be seen that the main effect of method is significant implying significant difference between the means of MSPR for regression and neural networks. Also, the interaction effects of method × size, method × noise and method × noise × size are all statistically significant.

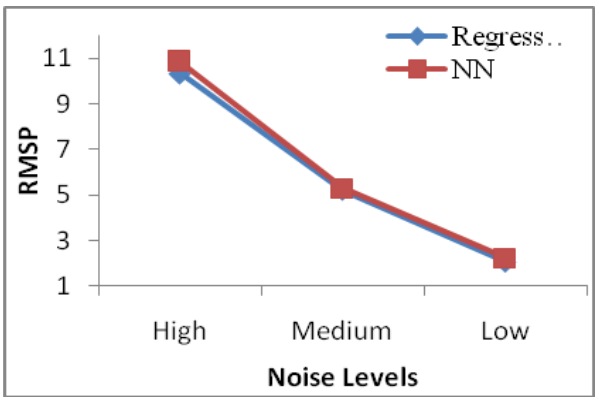
To assist in the practical interpretation of these interaction effects, Figures 2(a), 2(b) and 2(c) present the profile plots for the two-way interaction effects of method × size, method × noise and method × variable. The profile plots give us an indication of the levels of various factors contributing to the differences in the performance measures. It can be seen from Figure 2(a) that the RMSP value for regression is less as compared to neural network technique at all the levels of number of independent variable. From Figure 2(b), it is clear that the RMSP value for regression is less than that of neural network for small sample size. As the sample size increases from small to large, the error values of the two techniques become very close. It can be seen from Figure 2(c) that the RMSP value for regression is less than that of neural network at all levels

of noise but the difference is visible only for high noise level.

(a) Method & Sample Size interaction profile plot



(b) Method & Noise Level interaction profile plot



(c) Method & Variable interaction profile plot

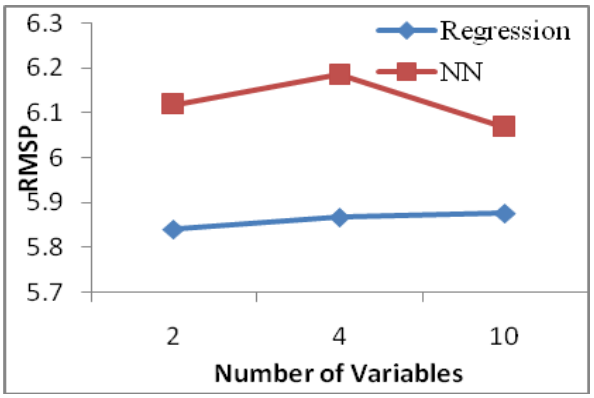
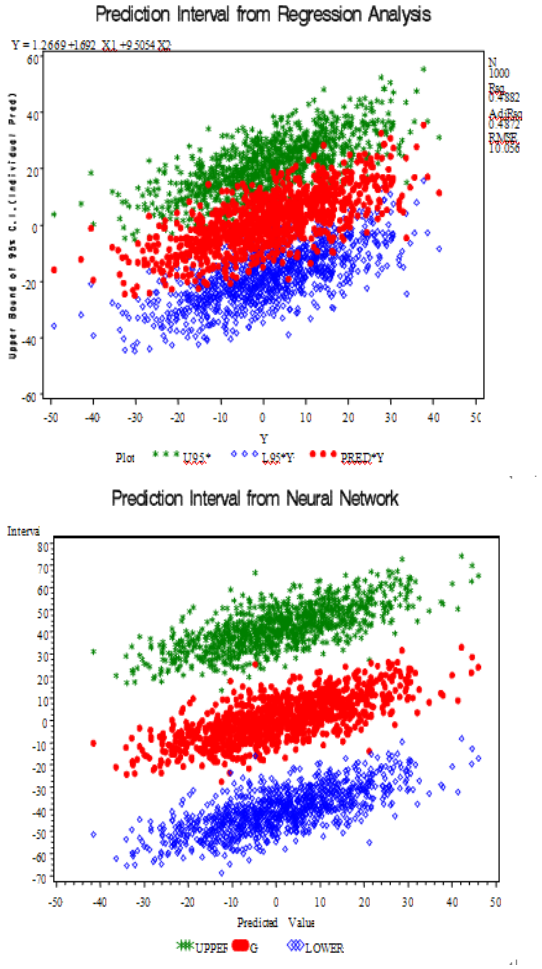


Figure 2 Profile plots for different interaction effects

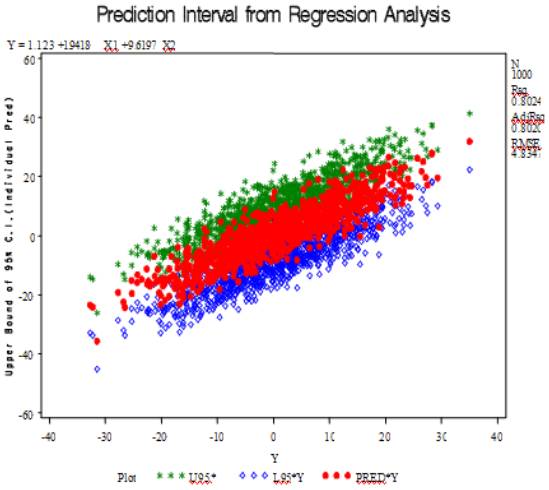
Further, multiple comparison tests are performed to see at what level of one factor, difference exists in the other factor causing interaction effect to be significant. To carry out multiple comparisons of this repeated measure factorial design, we have used F-tests with appropriate numerator and denominator as explained in Winer (1971). From these multiple comparison tests, it can be concluded that significant differences are found between the performances of the two techniques at all levels of noise but only for small sample size.

Prediction intervals for regression analysis and neural network analysis are computed for each of the levels of the number of independent variables and for each of the levels of noise. The scatter plots shown in Figure 3 correspond to the case of number of independent variables being two at various levels of noise.

(a) High Noise



(b) Medium Noise



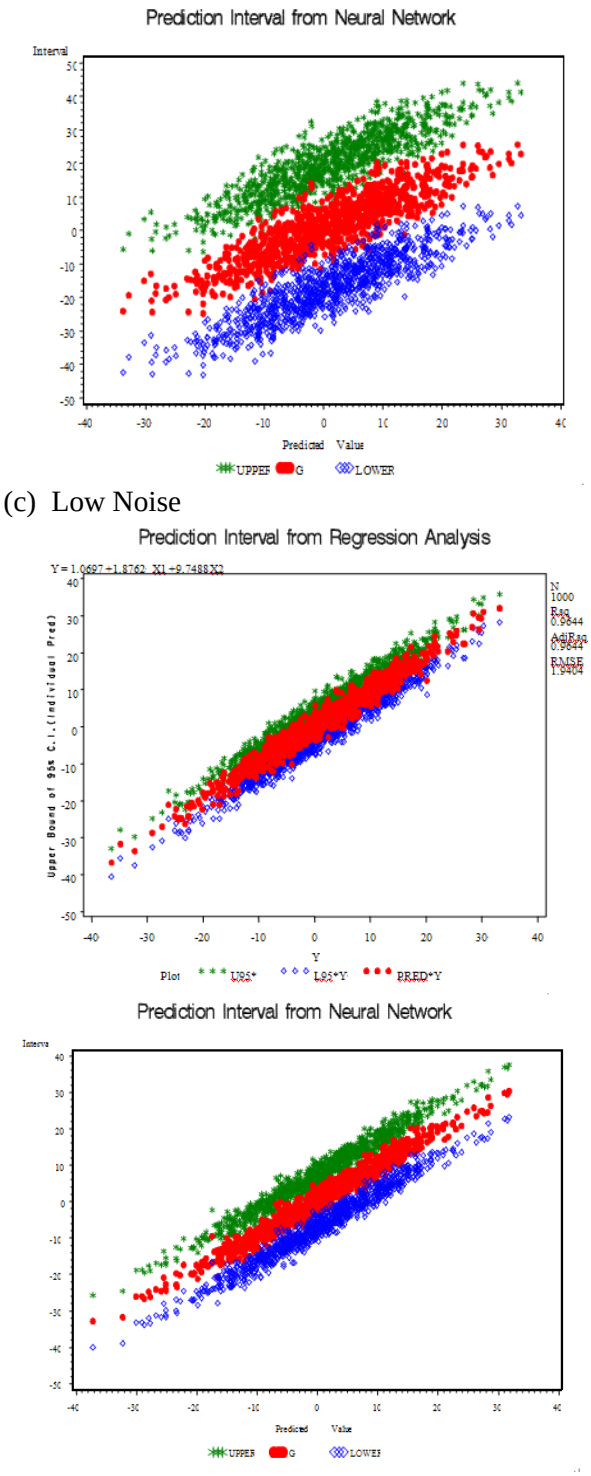


Figure 3 Scatter plots of prediction intervals of regression and NNs for the case of two independent variables at various levels of noise

These plots clearly indicate that the sizes of the prediction intervals of neural networks are much larger as compared to regression analysis as these prediction intervals are well within that of neural networks. It can be seen that the size of the prediction intervals becomes narrower as we move from high noise level to low noise level. Similar inference results are seen when the number of independent variables considered are 4 and 10 and hence the scatter plots are not show

V. CONCLUSION

In this paper, the performance of neural networks is compared with that of regression analysis when all the assumptions of regression analysis are met. This is carried out by systematically generating large number of data sets that vary on various dimensions like number of variables, sample size and noise level. Performance of the two techniques is compared using predictive error values for the independent data set that was not used for training the models.

This study reveals that the performance of regression analysis and neural network are comparable at all levels of noise and at all levels of the number of independent variables for large and medium sample sizes when all the assumption of regression analysis are satisfied. However, regression analysis outperforms neural network for small sample size irrespective of the levels of noise and the number of independent variables. Neural network being data driven technique, the decline in the performance for small sample size could be due to non availability of enough training samples to complete the learning process. This result points out the need for careful implementation of neural network when the sample size is small.

Further, asymptotic prediction intervals are obtained for neural networks and are graphically compared with the prediction intervals of regression analysis for each of the experimental conditions. The width of the prediction intervals of neural networks is much larger as compared to regression analysis as these prediction intervals are well within that of neural networks. The mean error values of regression are consistently less than the corresponding values of neural networks for the test data set. These observations support the use of regression technique in predicting future observations when the assumptions of regression analysis are met. Regression technique further stands out as it allows interpretation of coefficients of the independent variables and inferences can also be drawn regarding the significance of variables in prediction problems.

ACKNOWLEDGEMENT

This work was supported under the Research Scheme No. 25 (0137)/05/EMR-II from Council of Scientific and Industrial Research, India.

REFERENCES

[1] Banks, D.L., Olszewski, R. T. and Maxion, R. A. (2003). Comparing methods for multivariate nonparametric regression. Communications in Statistics: Simulation and Computation, 32(2), 541-571.

[2] Bishop, C. M. (1995). Neural Networks for Pattern Recognition. London, U.K.: Oxford Univ. Press.

[3] Cherkassky, V., Gehring, D. and Mulier, F. (1996). Comparison of adaptive methods for function estimation from samples, IEEE Transactions on Neural Network, 7(4), 969-984.

[4] Girosi, F., Jones, M. and Poggio, T. (1995). Regularization theory and neural networks architectures. Neural Computation, 7, 219-269.

[5] Guo, W.W. (2010). Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction. Expert Systems with Applications, 37, 3358–3365.

[6] Hagan M. T. and Menhaj M. (1994). Training feedforward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks, 5 (6), 989-993.

[7] Hardgrave, B. C., Wilson, R. L., and Walstrom, K. A. (1994). Predicting graduate student success: A comparison of neural networks and traditional techniques. Computers and Operations Research, 29(30), 49–263.

[8] Hill, T. and Remus, W. (1994). Neural network models for intelligent support of managerial decision making. Decision Support Systems, 11, 449-459.

[9] Hinton, G.E. (1987). Learning Translation Invariant Recognition in Massively Parallel Networks. In Proceedings PARLE Conference on Parallel Architectures and Languages Europe. Bakker, J.W., A.J. Nijman, and P.C. Treleaven (eds). Berlin: Springer Verlag.

[10] Huang, W.Y. and Lippmann, R.P. (1987). Comparisons between neural net and conventional classifiers. IEEE First International Conference on Neural Networks, IV, San Diego, CA, 485-493.

[11] Huynh, H. (1978). Some approximate tests for repeated measurements. Psychometroka, 43, 161-175.

[12] Hwang, J.T.G., Ding, A.A. (1997). Prediction intervals for arti5cial neural networks. Journal of the American Statistical Association, 92, 748–757.

[13] Keselman, H.J., Carriere, K.C. and Lix L.M. (1993). Testing Repeated Measures Hypotheses When Covariance Matrices Are Heterogeneous. Journal of Educational Statistics, 18(4), 305-319.

[14] Korkut, I., Acır, A. and Boy, M. (2011). Application of regression and artificial neural network analysis in modelling of tool–chip interface temperature in machining. Expert Systems with Applications, 38, 11651-11656.

[15] Levenberg, K. (1944). A method for the solution of certain problems in least squares. Quarterly Applied Mathematics, 2, 164-168.

[16] Liu, X., Kang, K. and Li, F. (2009). Simulation of artificial neural network model for trunk sap flow of Pyrus pyrifolia and its comparison with multiple-linear regression. Agricultural Water Management, 96, 939–945.

[17] Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. Journal of Society of Industrial Mathematics, 11, 431-441.

[18] Marquez, L., Hill, T., Warthley, R. and Renues, W. (1991). Neural network models as an alternative to regression. Proceedings of the 24th Annual Hawaii International Conference on System Sciences, IEE Computer Society Press, Kauai, Hawaii, 4,129-135.

[19] Marquez, L. and Hill, T. (1993). Function approximation using backpropagation and general regression neural networks. Proceedings of the 26th Hawaii International Conference on System Sciences, IEE Computer Society Press, Hawaii, 4, 607-615.

[20] Paliwal, M., and Kumar, U.A. (2009). Neural networks and statistical techniques: A review of applications. Expert Systems with Applications, 36, 2-17.

[21] Pendharkar, P. C. (2006). Scale economies and production function estimation for object-oriented software component and source code documentation size. European Journal of Operational Research, 172, 1040–1050.

[22] Sharda, R. and Wilson, R.L. (1993). Performance comparison issues in neural network experiments for classification problems. Proceeding of the Twenty-Sixth Hawaii International Conference on System Sciences, 4, 649-657.

[23] SAS Institute Inc. (2007). Statistical Analysis System, Version 9.1. Cary, NC: SAS Institute Inc.

[24] Sawyer, R. (1982). Sample Size and the Accuracy of Predictions Made from Multiple Regression Equations. Journal of Educational Statistics, 7(2), 91-104.

[25] Winer, B. J. (1971), *Statistical Principles in Experimental Design*, Second Edition, New York: McGraw-Hill, Inc.

[26] Wu, L. and Moody, J. (1996). A smoothing regularizer for feed-forward and recurrent neural networks. Neural Computation, 8(3), 463-491.

[27] Yilmaz, I. (2009). Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Katland slides (Tokat—Turkey). Computers and Geosciences, 35 1125–1138.