Applied Soft Computing 11 (2011) 3859–3869

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

# The predictive accuracy of feed forward neural networks and multiple regression in the case of heteroscedastic data

Mukta Paliwal, Usha A. Kumar*

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

ARTICLE INFO

*Article history:*
Received 17 May 2010
Received in revised form 2 December 2010
Accepted 24 January 2011
Available online 3 March 2011

*Keywords:*
Monte Carlo simulation
Heteroscedaticity
Prediction
Regression
Neural networks

ABSTRACT

This paper compares the performances of neural networks and regression analysis when the data deviate from the homoscedasticity assumption of regression. To carry out this comparison, datasets are simulated that vary systematically on various dimensions like sample size, noise levels and number of independent variables. Analysis is performed using appropriate experimental designs and the results are presented. Prediction intervals for both the methods for the case of nonconstant error variance are also calculated and are graphically compared. Two real life data sets that are heteroscedastic have been analyzed and the findings are in line with the results obtained from experiments using simulated data sets.



## 1. Introduction

Neural networks are mathematical tools that resulted from attempts to model the capabilities of human brains. Literature [1–4] points out the potential of neural networks for prediction problems which has led to a number of studies comparing the performance of neural networks and regression analysis. However, there are very few studies addressing the consequences of deviation from the underlying assumptions of regression technique on the comparative performances of these two techniques. Pickard et al. [5] have used simulation to create data sets with known underlying model and with non-normal characteristics such as skewness, unstable variance and outliers that are frequently found in software cost modeling problem. Multiple regression and classification and regression tree have been compared and viability of simulation to allow such comparisons under controlled mechanism is demonstrated. Standard multiple regression techniques are found to be the best when the data exhibit moderate non-normality and under more extreme condition such as heteroscedasticity, nonparametric techniques are found to have the best performance.

Gaudart et al. [6] have done comparison of the performance of multilayer perceptron and linear regression for epidemiological data when the data deviate from some of the underlying assumptions of regression. One of the functional forms considered in this study pertains to heteroscedasticity in the error variance. For the heteroscedastic errors, authors have concluded that predictions from both the models are of the same size and order. However, the findings are limited to the characteristics of the functional form considered in this study.

The available literature is sparse, particularly comparing the performance of these two techniques when the data is heteroscedastic. Thus there is a need for extensive and systematic study that takes into account various patterns of heteroscedasticity and a large number of data characteristics while comparing the performance of the two techniques.

The aim of the present study is to systematically compare the performance of neural network and regression technique when the data deviate from the assumption of homoscedasticity. This comparison is carried out by simulating data sets for different patterns of heteroscedasticity possessing various data characteristics that vary in sample size, amount of noise and number of independent variables. Appropriate prediction intervals for both the methods in presence of heteroscedasticity are also obtained and are graphically compared. To gain insight on how the comparative performance of both the techniques under heteroscedasticity translates to real life situations, two data sets have also been analyzed and results are discussed.

The next section describes the regression model with nonconstant error variances. Weighted least squares method, a remedial measure to deal with unequal error variances is also discussed in this section. Section 3 presents the construction of prediction intervals for regression and neural network techniques when the data is heteroscedastic. The experimental design, data generation

* Corresponding author. Tel.: +91 22 25767786; fax: +91 22 25722872.
*E-mail addresses:* mukta.paliwal@iitb.ac.in (M. Paliwal), usha@som.iitb.ac.in (U.A. Kumar).

1568-4946/$ – see front matter 
doi:10.1016/j.asoc.2011.01.043

procedure and performance evaluation criterion are discussed in Section 4. Section 5 presents the data analyses and results from the simulation study are presented in Section 6. Section 7 presents analysis of two real life data sets and the last section concludes this paper.

## 2. The model

Consider the multiple linear regression model given by the equation,

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

where $Y_i$ is the dependent variable, the $X_{ji}$ are independent variables, $\beta_j$ are unknown parameters and $\varepsilon_i$ is the error term. In order for estimates of $\beta$'s to have desirable properties, we need the following assumptions, called Gauss–Markov conditions:

$$E(\varepsilon_i) = 0; \quad E(\varepsilon_i^2) = \sigma^2; \quad E(\varepsilon_i \varepsilon_k) = 0, \quad \text{when} \quad i \neq j, \quad \text{for all} \quad i, j = 1, \ldots, n$$

These conditions, if they hold, assure that ordinary least squares (OLS) estimates are best linear minimum variance estimates among the class of unbiased estimators. Violation of the second condition is called heteroscedasticity in which the error term of a regression model does not have constant variance over all the observations, i.e. $Var(\varepsilon_i) = \sigma_i^2$. OLS estimates of regression coefficients under heteroscedasticity are still unbiased and consistent but they are no longer the best linear unbiased estimators. Heteroscedasticity causes variances of parameter estimates to be large and thus can affect $R^2$, estimate of $\sigma^2$, and various inferential procedures substantially.

In this study, heteroscedasticity in the error term is introduced by defining $\varepsilon_i$ as $h_i u_i$, where $u_i$ (random error) are independent $N(0, \sigma^2)$ and $h_i$ is some function of one of the independent variables ($X_i$'s) and thus $Var(\varepsilon_i) = \sigma_i^2 = (h_i^2 \sigma^2)$. It is common for heteroscedastic error variance to increase or decrease as the expectation of $Y$ grows larger, or there may be systematic relationship between error variance and a particular $X$. In this situation, the nonconstant error variances $\sigma_i^2 = (h_i^2 \sigma^2)$ are known up to a proportionality constant, where $\sigma^2$ is the proportionality constant. Weighted least squares estimation procedure is one of the methods that can be used to correct such form of heteroscedasticity and is the topic of discussion in the next subsection.

### 2.1. Weighted least squares

Weighted least squares (WLS) regression is useful for estimating the regression parameters when heteroscedasticity is present in the data. Here, we assume that the error variances are known up to proportionality constant i.e. $Var(\varepsilon_i) = \sigma_i^2 = (h_i^2 \sigma^2)$. Estimation of parameters by the method of weighted least squares is closely related to parameter estimation by ordinary least squares. The weighted least squares criterion includes an additional weight, $w_i (= 1/h_i^2)$, where the estimates of parameters are obtained by minimizing

$$Q_w = \sum_{i=1}^{n} w_i (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_p X_{pi})^2 \tag{2}$$

Since the weight $w_i$ is inversely related to the variance $\sigma_i^2$, it reflects the amount of information contained in the observation $Y_i$. Thus, an observation $Y_i$ that has a large variance receives less weight than another observation that has a smaller variance. The weighted least squares estimators of the regression coefficients for model (1) can be easily expressed using matrix terms. The following matrices are defined as:

$$\underset{n\times 1}{\mathbf{Y}} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \qquad \underset{n\times(p+1)}{\mathbf{X}} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{1p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n1} & \cdots & X_{np} \end{pmatrix} \tag{3}$$

Let the matrix $\mathbf{W}$ be a diagonal matrix containing the weights $w_i$:

$$\underset{nxn}{\mathbf{W}} = \begin{pmatrix} w_1 & 0 & \ldots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{pmatrix} \tag{4}$$

The normal equations can be expressed as follows:

$$(\mathbf{X}'\mathbf{W}\mathbf{X})\mathbf{b}_w = \mathbf{X}'\mathbf{W}\mathbf{Y} \tag{5}$$

The weighted least squares estimators of the regression coefficients are:

$$\underset{(p+1)\times 1}{\mathbf{b}_w} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y} \tag{6}$$

where $\mathbf{b}_w$ is the vector of the estimated regression coefficients obtained by weighted least squares. The variance–covariance matrix of the weighted least squares estimated regression coefficients is:

$$\boldsymbol{\sigma}^2 \underset{(p+1)\times(p+1)}{\{\mathbf{b}_w\}} = \sigma^2 (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \tag{7}$$

and this matrix is not known since the proportionality constant $\sigma^2$ is not known. However, it can be estimated as

$$s^2 \underset{(p+1)\times(p+1)}{\{\mathbf{b}_w\}} = s_w^2 (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \tag{8}$$

where $s_w^2$ is based on weighted squared residuals and can be calculated as $\sum w_i (Y_i - \hat{Y}_i)^2 / (n - p - 1)$. Weighted least squares can also be viewed as ordinary least squares on transformed model given by

$$\mathbf{Y}_w = \mathbf{X}_w \boldsymbol{\beta} + \boldsymbol{\varepsilon}_w \tag{9}$$

where $\mathbf{Y}_w = \mathbf{W}^{1/2}\mathbf{Y}$, $\mathbf{X}_w = \mathbf{W}^{1/2}\mathbf{X}$ and $\boldsymbol{\varepsilon}_w = \mathbf{W}^{1/2}\boldsymbol{\varepsilon}$ with $\mathbf{W}^{1/2}$ being a diagonal matrix containing square roots of the weights $w_i$. Here

$$\underset{(p+1)\times 1}{\mathbf{b}_w} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y} = (\mathbf{X}_w'\mathbf{X}_w)^{-1}\mathbf{X}_w'\mathbf{Y}_w \quad \text{and} \quad \sigma^2 \underset{(p+1)\times(p+1)}{\{\mathbf{b}_w\}} = \sigma^2(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} = \sigma^2(\mathbf{X}_w'\mathbf{X}_w)^{-1}$$

It can be observed that

$$\boldsymbol{\sigma}^2\{\varepsilon_w\} = \mathbf{W}^{1/2}\boldsymbol{\sigma}^2\{\varepsilon\}\mathbf{W}^{1/2} = \mathbf{W}^{1/2}\sigma^2\mathbf{W}^{-1}\mathbf{W}^{1/2} = \sigma^2$$

satisfying the constant variance assumption.

### 2.2. Prediction intervals

Prediction intervals to predict the future observation, $Y_0$, from regression model and neural network model in case of heteroscedastic error variance are discussed in this section.

In case of heteroscedastic error variance, a modified form of standard prediction interval needs to be used as usual prediction intervals will tend to be wider for small values of error variance and will tend to be narrower for large values of error variance. Incorporating the nonconstant error variance component $h_i$ in regression technique, an appropriate $100(1-\alpha)\%$ prediction interval for $Y_0$ is given by

$$[\hat{Y}_0 \pm t_{\alpha/2}^{(n-p-1)} s_w \sqrt{h_i^2 + \mathbf{X}_0'(\mathbf{X}_w'\mathbf{X}_w)^{-1}\mathbf{X}_0}] \tag{10}$$