Jour. Ind. Soc. Ag. Statistics
49 (Golden Jub. No.), 1996-97 : 177–188

Test of Hypotheses Based on Cross-validation for Non-nested Linear Regression Models

G.K. Shukla and Nisha Bawa
Indian Institute of Technology, Kanpur, India

SUMMARY

The predicting likelihood (PL) test developed by Bawa [1] is applied to non-nested linear regression models. Two examples are given to show that the use of PL method, which is essentially a cross-validatory method, leads to the same conclusion as obtained by Cox's test based on the likelihood ratio.

Key words: Predicting likelihood test, Cross-validation, Non-nested regression models, Discrimination between competing models, Cox's likelihood ratio test.

1. Introduction

For a given set of data, very often we have more than one alternative models in mind. Then the question arises which one of them best explains the data. To settle down the matter, two approaches are generally considered. One is discrimination i.e., some goodness-of-fit criterion is adopted and the model giving the optimum value is chosen. At a more formal level, the other approach of significance testing is employed. For nested models, tests are available which are very simple in nature. However, for non-nested models the situation becomes a bit complex. By non-nested models it is meant that one model cannot be obtained as a special case of the other.

The general problem of non-nested models from testing point of view was first considered by Cox ([2], [3]). His work concentrates on selecting the best model, hence not assuming that one of the hypotheses contains the best one. Another variant of the problem had been considered before Cox but they are very different in nature.

For regression models, the Cox test statistic was simplified which was based on the simple difference between the two variance estimates. Pesaran [10] considered the test statistic based on the difference of the logarithms of the two variance estimates and the statistic is called the Pesaran-Cox test statistic.

Afterwards much work has been done based on Cox's approach and artificial nesting procedures.

Data splitting, or cross-validation, is a very commonly used technique for validation of the models, which is of great relevance when the purpose of a model is prediction. Geisser and Eddy [6] made use of this technique for model selection. They also gave a test criterion for tests of nested models. Based on the predicting density, Bawa [1] developed a test for separate families of hypotheses which we call Predicting Likelihood (PL) Test. The asymptotic distribution of the test statistic was obtained for some distributions. In this paper the basic techniques of that test are applied to non-nested regression models considering the Cox's framework.

In Section 2, the PL test criteria is briefly discussed. The notations used are essentially those of Bawa [1]. In Section 3, the basic concepts of the PL test is used for non-nested regression models. In Section 4, the test is applied to the linear regression models. Section 5 discusses two examples.

2. *Predicting Likelihood (PL) Test Criteria*

In this section we consider the case when $Y_1, \ldots, Y_n$ are independent but not necessarily identically distributed and we are interested in testing the null hypothesis :

$$H_f : \text{ density of Y is } f(y, \underline{\alpha}), \alpha \in \Omega_{\underline{\alpha}}$$

against the alternative

$$H_g : \text{ density of Y is } g(y, \underline{\beta}), \ \underline{\beta} \in \omega_{\underline{\beta}}$$

where $\underline{\alpha}$ and $\underline{\beta}$ are vectors of unknown parameters and can be of different dimensions. Moreover, the p.d.f.'s f and g are separate, that is for any parameter value $\underline{\alpha}_0 \in \Omega_{\underline{\alpha}}$, $f(y, \underline{\alpha}_0)$ cannot be approximated arbitrarily closely by $g(y, \underline{\beta})$.

The Cox's Test statistic for testing H_f against H_g is based on

$$T_f = L_{fg} - E_{\hat{\underline{\alpha}}} \{ L_{fg} \}$$

where

$$L_f(\hat{\underline{\alpha}}) = \sum_{i=1}^{n} \log f(y_i, \hat{\underline{\alpha}}), \quad L_g(\hat{\underline{\beta}}) = \sum_{i=1}^{n} \log g(y_i, \hat{\underline{\beta}})$$

$$L_{fg} = L_f(\hat{\underline{\alpha}}) - L_g(\hat{\underline{\beta}})$$

Here $E_{\hat{\alpha}}$ denotes the expectation taken under the density $f(y, \hat{\underline{\alpha}})$ and $\hat{\underline{\alpha}}$, the maximum likelihood (ML) estimator of $\underline{\alpha}$, is treated as the true value of the parameter. The same is true for $\hat{\underline{\beta}}$.

Thus, T_f compares the log-likelihood ratio L_{fg} with its best estimate under H_f. Let the predicting density of Y_j under H_f be $f(y_j, \hat{\underline{\alpha}}_{(j)})$ and under H_g be $g(y_j, \hat{\underline{\beta}}_{(j)})$ where $\hat{\underline{\alpha}}_{(j)}$ and $\hat{\underline{\beta}}_{(j)}$ are the ML estimator of $\underline{\alpha}$ and $\underline{\beta}$, deleting observation y_j.

The predicting density considered here is essentially a cross-validatory assessment (Stone [11]). Asymptotic equivalence of choice of model using quasi-predicting likelihood and Akaike's criterion was established by Stone [11]. This motivated us to construct a significance test based on the sample reuse method.

The quasi-predicting log-likelihood ratio is given by

$$L^p_{fg}(\hat{\underline{\alpha}}, \hat{\underline{\beta}}) = \sum_{j=1}^{n} \{ \log f(y_j, \hat{\underline{\alpha}}_{(j)}) - \log g(y_j, \hat{\underline{\beta}}_{(j)}) \}$$

It was shown by Stone that for nested models M_1 and M_2 such that $M_1 \subseteq M_2$, the quasi-predicting log-likelihood ratio $L^p_{M_1} - L^p_{M_2}$ is asymptotically equivalent to $\lambda + p_{M_1} - p_{M_2}$ as the sample size $n \to \infty$, where λ is the log-likelihood ratio criterion, and p_{M_i} denotes the dimension of the vector of parameters in the model M_i, $i = 1, 2$. This shows the asymptotic equivalence of the cross-validatory and the Akaike's information criterion. Hence, under general conditions, $-2\{L^p_{M_1} - L^p_{M_2}\} + p$ is asymptotically distributed as χ^2_{2p}, with $p = p_{M_2} - p_{M_1}$. Unfortunately for non-nested models, the above asymptotic theory does not apply. For testing H_f vs H_g, it is proposed to base our test on

$$P_f = L^p_{fg}(\hat{\underline{\alpha}}, \hat{\underline{\beta}}) - E_{\hat{\underline{\alpha}}}\{L^p_{fg}(\hat{\underline{\alpha}}, \hat{\underline{\beta}})\}$$

3. *Asymptotic Distribution of* P_f

The results given in Bawa will be used here and some outlines of the main results are given in the following :

$$L_{fg}^{p}\,(\hat{\underline{\alpha}}, \hat{\underline{\beta}}) = \sum_{j=1}^{n} \{ \log f(Y_j, \hat{\alpha}_{(j)}) - \log g(Y_j, \hat{\beta}_{(j)}) \}$$

$$= L_f^p - L_g^p$$

Under H_f,

$$L_f^p \cong \sum_{i=1}^{n} \log f(y_i, \hat{\underline{\alpha}}) + d_f$$

and

$$L_g^p = \Sigma \log g(y_i, \hat{\underline{\beta}}_{(i)}) \cong \Sigma \log g(y_i, \hat{\underline{\beta}}) + \text{Trace } L_4^{-1} L_3$$

where,
$$L_4 = E_{\underline{\alpha}} \left\{ \frac{\partial^2}{\partial \underline{\beta}^2} \sum_{i=1}^{n} \log g(y_i, \underline{\beta}_{\alpha}) \right\}$$

$$L_3 = E_{\underline{\alpha}} \sum_{i=1}^{n} \left(\frac{\partial}{\partial \underline{\beta}} \log g(y_i, \underline{\beta}_{\alpha}) \right) \left(\frac{\partial}{\partial \underline{\beta}} \log g(y_i, \underline{\beta}_{\alpha}) \right)^T$$

and d_f is the number of parameters under H_f , and $\hat{\underline{\beta}}$ converges in probabilities to $\underline{\beta}_{\alpha}$ under H_o.

The Predicting Likelihood test is based on

$$P_f = L_f^p - L_g^p - E_{\hat{\underline{\alpha}}} \{ L_f^p - L_g^p \}$$

Hence,
$$P_f \cong T_f - \frac{1}{n} \Sigma L^T F_{i, \underline{\alpha}}$$

where T_f is the numerator of the Cox's test statistic,

$$L = L_2^{-1} \left\{ \frac{\partial}{\partial \underline{\alpha}} \text{Trace } L_4^{-1} L_3 \right\}; \; F_{i,\underline{\alpha}} = \log f(y_i, \underline{\alpha})$$

and
$$L_2 = E_{\underline{\alpha}} \left\{ \frac{\partial^2}{\partial \underline{\alpha}^2} \log f(y, \underline{\alpha}) \right\}$$

The asymptotic variance of P_f is given in terms of the variance of T_f by the expression

$$V_{\underline{\alpha}}(P_f) \cong V_{\underline{\alpha}}(T_f) + \frac{1}{n} L^T L_2^{-1} L$$

$$= V_{\underline{\alpha}}(T_f) - \frac{1}{n}\left(\frac{\partial}{\partial \underline{\alpha}} \text{Trace } L_4^{-1} L_3\right)^T L_2^{-1}\left(\frac{\partial}{\partial \underline{\alpha}} \text{Trace } L_4^{-1} L_3\right)$$

$P_f^* = P_f / \sqrt{V_{\underline{\alpha}}(T_f)}$ is the PL test statistic for testing H_f against H_g, which under the null hypothesis has a standard normal distribution asymptotically. This test is consistent under certain conditions.

4. *Test Applied to Linear Regression Problem*

In this section, the problem of testing non-nested regression models is considered. The null hypothesis

$$H_0 : \underline{Y} = X\, \underline{b}_0 + \underline{u}_0$$

is to be tested against the alternative

$$H_1 : \underline{Y} = Z\, \underline{b}_1 + \underline{u}_1$$

where $\underline{X}$ and $\underline{Z}$ are assumed to be fixed and all columns of one can not be obtained from those of the other, which essentially means that the models are non-nested. Let k_0 and k_1 be the dimensions of $\underline{b}_0$ and $\underline{b}_1$, respectively. $\underline{u}_0$ and $\underline{u}_1$ are assumed to be i.i.d. with $N(0, \sigma_0^2)$ and $N(0, \sigma_1^2)$, respectively. The log-likelihood under H_0 is

$$\text{Log } f(y \mid b_0, X) = \frac{-1}{2\sigma_0^2}(\underline{Y} - \underline{X}\, \underline{b}_0)^T (\underline{Y} - \underline{X}\, \underline{b}_0) - \frac{n}{2}\log 2\pi\sigma_0^2$$

and under H_1 is

$$\text{Log } g(y \mid b_0, \underline{Z}) = \frac{-1}{2\sigma_1^2}(\underline{Y} - \underline{Z}\underline{b}_0)^T (\underline{Y} - \underline{Z}\underline{b}_1) - \frac{n}{2}\log 2\pi\sigma_1^2$$

$$= \frac{1}{2}\sum_{i=1}^{n} [\, (y_i - \underline{z}_i^1 \hat{\underline{b}}_{1\,(i)})^2 / \hat{\sigma}^2_{1\,(i)} - (y_i - \underline{x}'_i \hat{\underline{b}}_{0\,(i)})^2 / \hat{\sigma}^2_{0\,(i)}$$

$$+ \log (\hat{\sigma}^2_{1\,(i)} / \hat{\sigma}^2_{0\,(i)})\,]$$

For estimating the variances, consider the unbiased estimators rather than the maximum likelihood estimators.

Let $e_{0(i)}$ and $e_{1(i)}$ be the prediction error for the deleted observation y_i under H_0 and H_1 respectively, given by

$$e_{0(i)} = y_i - \underline{x}_i^T \hat{\underline{b}}_{0(i)} \text{ and } e_{1(i)} = y_i - \underline{z}_i^T \hat{\underline{b}}_{1(i)}$$

where $\hat{\underline{b}}_{0(i)}$ and $\hat{\underline{b}}_{1(i)}$ are least squares estimates of $\underline{b}_0$ and $\underline{b}_1$, based on $(n-1)$ observations, excluding the ith observation,

$$\hat{\underline{b}}_{0(i)} = (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T \underline{y}_{(i)} ; \; \hat{\underline{b}}_{1(i)} = (Z_{(i)}^T Z_{(i)})^{-1} Z_i^T \underline{y}_{(i)}$$

Then,

$$E_{\underline{\alpha}} \left(e_{0(i)}^2 \mid \underline{y}_{(i)} \right) = \sigma_0^2 + (\underline{b}_0 - \hat{\underline{b}}_{0(i)})^T \underline{x}_i x_i^T (\underline{b}_0 - \hat{\underline{b}}_{0(i)})$$

$$E_{\underline{\alpha}} \left(e_{0(i)}^2 / \sigma_{0(i)}^2 \right) = E_{\underline{\alpha}, \underline{y}_{(i)}} [E_{\underline{\alpha}, \underline{y}_{(i)}} (e_{0(i)}^2 / \hat{\sigma}_{0(i)}^2) \mid \underline{y}_{(i)}, \underline{X}_i)]$$

$$= (n - k_0 - 1) \{ 1 + \text{Trace } \underline{x}_i^T \underline{x}_i (X_{(i)}^T X_{(i)})^{-1} \} / (n - k_0 - 3)$$

Therefore,

$$e1 = \frac{1}{2} E_\alpha \{ \sum_{i=1}^{n} (y_i - \underline{x}_i^T \hat{\underline{b}}_{0(i)})^2 / \sigma_{0(i)}^2 \}$$

$$= (n - k_0 - 1) \{ n + \sum_{i=1}^{n} \text{Trace } \underline{x}_i \underline{x}_i^T (\underline{X}_{(i)}^T \underline{X}_{(i)})^{-1} \} / 2 (n - k_0 - 3)$$

$(n - k_1 - 1) \sigma_{0(i)}^2 / \sigma_0^2$ has a non-central chi-square distribution with $(n - k_1 - 1)$ degrees of freedom and non-centrality parameter δ. Hence,

$$E [1 / \sigma_{1(i)}^2] = \frac{n - k_1 - 1}{\sigma_0^2} \exp(-\delta/2) \sum_{j=0}^{\infty} \frac{\delta^j}{2^j j!} \frac{1}{n - k_1 - 2 + 2j}$$

which gives

$$E_{\underline{\alpha}} \{ (y_i - \underline{z}_i^T \hat{\underline{b}}_{1(i)})^2 \mid \bar{y}_{(i)}' \underline{x} \}$$

$$= E_{\underline{\alpha}} \{ (y_i - \underline{x}_i^T \underline{b}_0) + (x_i^T \underline{b}_0 - \underline{z}_i^T \hat{\underline{b}}_{10(i)}) + \underline{z}_i^T (\hat{\underline{b}}_{10(i)} - \hat{\underline{b}}_{1(i)}) \}^2$$

$$= \sigma_0^2 + l_i^2 + [\underline{z}_i^T (\hat{\underline{b}}_{10(i)} - \hat{\underline{b}}_{1(i)})]^2$$

where

$$l_i = \underline{x}_i^T \underline{b}_0 - \underline{z}_i^T \hat{\underline{b}}_{10\,(i)}, \quad \text{and} \quad \hat{\underline{b}}_{10\,(i)} = (Z_{(i)}^T \underline{Z}_{(i)})\, \underline{Z}_{(i)}^T X_{(i)} \hat{\underline{b}}_{0\,(i)}$$

is the limit of $\hat{b}_{1\,(i)}$ in probability under H_0.

Hence,

$$E_{\underline{\alpha}} \{ (y_i - \underline{z}_i^T \hat{\underline{b}}_{1\,(i)})^2 \} = \sigma_0^2 + l_i^2 + \sigma_{10}^2 \; \underline{z}_i^T \; (Z_{(i)}^T Z_{(i)})^{(-1)} \underline{Z}_i$$

Therefore,

$$e2 = \frac{1}{2} E_{\underline{\alpha}} \{ \sum_{i=1}^{n} (y_i - \underline{z}_i^T \; \hat{b}_{1\,(i)})^2 \; / \, \sigma_{1\,(i)}^2 \; \}$$

$$= \frac{1}{2} \left(\frac{n-1}{\sigma_0^2} \exp\,(-\delta/2) \sum_{j=0}^{\infty} \frac{\delta^j}{2^j \; j\,!} \; \frac{1}{n-q-3+2j} \right) \left\{ \sigma_0^2 + l_i^2 + \sigma_{10}^2 \, \underline{z}_i^T \, (Z_{(i)}^T Z_{(i)})^{-1} \, \underline{z}_i \right\}$$

If X has a central χ^2 distribution with n degrees of freedom, then

$$E\,[\,\log X\,] = \log 2 + \Psi\,(n/2)$$

where Ψ is a diagamma function. Therefore,

$$E_{\underline{\alpha}}\,[\,\log \hat{\sigma}_{0\,(i)}^2\,] = \log\,(2\sigma_0^2 / (n-1) + \Psi\,(n-p-2)/2) = e3 \text{ (say)}$$

If X has a non-central χ^2 distribution with n degrees of freedom and non-centrality parameter δ, then

$$E\,[\,\log X\,] = \log 2 + \sum_{j=0}^{\infty} \frac{\delta^j}{2^j \, j\,!} \exp\,(-\delta/2)\, \Psi\,(n/2 + j)$$

Therefore,

$$E\,[\,\log \hat{\sigma}_{1(i)}^2\,] = \log\,(2\sigma_1^2 / (n-1) + \sum_{j=0}^{\infty} \frac{\delta^j}{2^j \, j\,!} \exp\,(-\delta/2) \;\; \Psi \left(\frac{n-p-2}{2} \right)$$

$$= e4 \text{ (say)}.$$

The numerator of the test statistic is given by

$$P_f = L_{fg}^p - \{ -e1 + e2 - e3 + e4 \}$$

For the computation of the variance under H_0, note that

$$L_3 = \frac{1}{n\,\sigma_{10}^2}\begin{pmatrix} \sigma_0^2 Z^T Z + \sum_i l_i^2 \underline{z}_i \underline{z}_i T & \dfrac{\sum_i l_i^2 \underline{z}_i}{2n\,\sigma_{10}^2} \\ \dfrac{\Sigma\, l_i^3 \underline{z}_i^T}{2n\,\sigma_{10}^2} & \dfrac{[3(1-2c)\,\sigma_0^4 + (1-2c)\,\sigma_{10}^4 + 2(4c-1)\,\sigma_0^2\,\sigma_{10}^2\,]}{4\sigma_{10}^2} \end{pmatrix}$$

where $c = (n - k_1)/n$

$$\text{Trace } L_4^{-1} L_3 = \frac{-1}{\sigma_{10}^2}\Big[\,k_1\,\sigma_0^2 + (\Sigma\, l_i^2\, \underline{z}_i\, \underline{z}_i^T)\,(Z^T Z)^{-1}$$

$$+ \frac{1}{2}\left\{\frac{(1-2c)\,(3\sigma_0^2 + \sigma_{10}^4) - 2\,(1-4c)\,\sigma_0^2\,\sigma_{10}^2}{2\,(1-c)\,\sigma_0^2 - (1-2c)\,\sigma_{10}^2}\right\}$$

Thus one can see that

$$\left\{\frac{\partial}{\partial \underline{\alpha}} \text{trace } L_4^{-1} L_3\right\}^T L_2^{-1} \left\{\frac{\partial}{\partial \underline{\alpha}} \text{ trace } L_4^{-1} L_3\right\}^T$$

$$= -\,\sigma_0^2\, n\, \underline{d}_1^T\,(X^T X)^{-1}\;\underline{d}_1 - 2\,\sigma_0^4\, d_2^2$$

where $d_1 = \dfrac{\partial}{\partial\, b_0}$ trace $L_4^{-1} L_3$

$$= -\frac{1}{2\,\sigma_{10}^8}\left\{\frac{\sigma_{10}^4}{n}\,(-12\sigma_0^2\, X^T M_z\, X\, \underline{b}_0 + 4\,\Sigma\, l_i^3\;(\underline{X}_i - X^T Z\,(Z^T Z)^{-1}\;\underline{z}_i\,))\right.$$

$$\left. - \frac{4}{n}(-3\sigma_0^4 + 6\sigma_0^2\,\sigma_{10}^2 + \Sigma\, l_i^4/n\,)\;\sigma_{10}^2\,\underline{x}'_i\;M_z\,X\,\underline{b}_0\right\}$$

$$-\;\frac{2}{\sigma_{10}^6}\;\{\frac{1}{n}(-3\sigma_0^4 + 3\sigma_0^2\,\sigma_{10}^2 + \Sigma\, l_i^4\;\}\,X^T M_z\, X\,\underline{b}_0$$

$$+\,\sigma_{10}^2\;\Sigma\, l_i^3\,(\,\underline{x}_i - X^T Z\,(Z^T Z)^{-1}\,\underline{z}_i\,)\}$$

and

$$d_2 = \frac{\partial}{\partial\,\sigma_0^2}\text{ Trace } L_4^{-1} L_3 = \frac{-2}{\sigma_{10}^8}\;\{\,6\sigma_0^2\,\sigma_{10}^2 - 3\,\sigma_{10}^4 - 3\sigma_0^4 + \Sigma\, l_i^4\}$$

and $M_z = I - Z\,(Z'\,Z)^{-1}\,Z'$

$$V_\alpha (P_f) = V_\alpha (T_f) - \sigma_0^2 \underline{d}_1^T (X^T X)^{-1} \underline{d}_1 - 2\sigma_0^2 d_2^2 / n$$

where

$V_{\underline{\alpha}} (T_f)$, as given by Pesaran [10], is

$$V_\alpha (T_f) = \sigma_0^2 \underline{b}_0^T X^T M_z M_x M_z X b_0 / \sigma_{10}^4$$

5. *Examples*

William's Examples: First, consider an example discussed by Williams [12]. It considers data of 42 specimens on maximum compressive strength (Y), density (X) and adjusted density (Z) for Pinus radiata. It was desirable to find a relationship between the maximum compressive strength Y and density X. Due to some reasons, adjusted density Z was thought to be more reasonable variable. Now the problem concentrates on which one should be preferred i.e., we have two competing models :

$$H_0 : y = \alpha_0 + \alpha_1 x$$

$$H_1 : y = \beta_0 + \beta_1 z$$

A test given by Hotelling [8] was applied by Williams to the data to test the significance of difference between the correlation coefficients of Y with X and Z. It was found that the difference is significant at the 5 per cent level. Sum of squares due to regression for the two models H_0 and H_1 are 28.209×10^6 and 29.746×10^6 respectively, with Z having larger of it.

Efron [5] also compared the two models by using Bootstrap method. He found 90% central confidence interval for the difference in mean squared error of Mallow's C_p statistic and found the model H_1 to be superior on this criterion.

In our analysis, we have also compared the Allen's PRESS statistic for these models. They were found to be 50.89×10^6 and 34.31×10^6 for models H_0 and H_1, respectively, which could mean that H_1 has higher predictive accuracy as compared to H_0.

The Cox's test was applied to test the non-nested models H_0 and H_1. It should be noted here that the problem considered by Cox is very different in nature than that of Hotelling and Efron.

For testing H_0 against H_1, for Cox's test, the following results were obtained :

$$T_f = -16.701\ ;\ \hat{V}_\alpha(T_f) = 8.427\ ;\ T_f^* = \frac{T_f}{\sqrt{\hat{V}_\alpha(T_f)}} = -5.75$$

which rejects the null hypothesis. However, reversing the role i.e. for testing H_1 against H_0 one gets for Cox's test

$$T_g = -3.259\ ;\ \hat{V}_\beta(T_g) = 9.4527\ ;\ T_g^* = \frac{T_g}{\sqrt{\hat{V}_\beta(T_g)}} = -1.06$$

which accepts H_1 i.e. regression with Z as independent variable is accepted.

Now coming to the proposed PL test we get

$$P_f = -16.948\ ;\ \hat{V}_\alpha(P_f) = 8.563\ ;\ P_f^* = \frac{P_f}{\sqrt{\hat{V}_\alpha(P_f)}} = -5.79$$

Thus rejecting H_0. Now reversing the role i.e., H_1 vs H_0 we get

$$P_g = -1.904\ ;\ \hat{V}_\beta(P_g) = 9.76\ ;\ P_g^* = \frac{P_g}{\sqrt{\hat{V}_\beta(P_g)}} = -0.61$$

Thus accepting H_1. Thus on the basis of predictive likelihood we reach the same conclusions as obtained by Cox test based on the likelihood ratio.

Hald's Example : The second example considered is that given by Hald [7]. It consists of a response variable Y which is the heat evolved in calories per gram of cement and four predictions X_1, X_2, X_3 and X_4 each of which is the amount of various ingredients in the mix. These data were used by Draper and Smith [4] and Montgomery and Peck [9] to illustrate all possible regressions. The four regressors are highly correlated and it was found that two-regressor models (X_1, X_2) and (X_1, X_4) have nearly the same R^2 values and if other variables are added, then there is only a slight increase in R^2. Since X_4 was found to be the best one-regressor model, Draper and Smith [4] suggested that (X_1, X_4) might be preferred over (X_1, X_2). But on the other hand (X_1, X_2) gives smaller residual mean square. C_p statistics for (X_1, X_2) and (X_1, X_4) are 2.68 and 5.50 respectively. Hence on the basis of all three criteria i.e. R^2, residual mean squares and C_p statistic (X_1, X_2) seems to be a good choice with (X_1, X_4) as a close competitor. PRESS statistic for (X_1, X_2) and (X_1, X_4) are

93.88 and 121.22, respectively. Addition of one more variable decreases the value slightly.

From the above analysis (X_1, X_2) seems to be a good choice, but to reach a definite conclusion, significance testing is necessary. We have the following two hypotheses :

$$H_0 : E(Y) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2$$

$$H_1 : E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_4$$

The summary of the results is as follows :

For Cox's test for H_0 vs H_1

$$T_f = -2.51, \; \hat{V}_\alpha (T_f) = 3.06, \; T_f^* = \frac{T_f}{\sqrt{\hat{V}_\alpha (T_f)}} = -1.44$$

and for H_1 vs H_0

$$T_g = -5.06, \; \hat{V}_\beta (T_g) = 2.96, \; T_g^* = \frac{T_g}{\sqrt{\hat{V}_\alpha (T_g)}} = -2.94$$

For the PL test for testing H_0 vs H'_1

$$P_f = -0.94, \; \hat{V}_\alpha (P_f) = 3.72, \; P_f^* = \frac{P_f}{\sqrt{\hat{V}_\alpha (P_f)}} = -0.49$$

and for H_1 vs H_0

$$P_g = -5.04, \; \hat{V}_\beta (P_g) = 3.56, \; P_g^* = \frac{P_g}{\sqrt{\hat{V}_\beta (P_g)}} = -2.66$$

Therefore, both Cox's and PL test accept H_0 i.e. model with (X_1, X_2) and reject H_1 i.e. model with (X_1, X_4).

The procedure based on PL method, which is a cross-validatory method, is relatively more complex to apply but is advisable when models are to be used for predicting future observations.

REFERENCES

[1] Bawa, N., 1994. A test for separate families of hypotheses based on crossed validation. Unpublished Ph.D. thesis, I.I.T., Kanpur.

[2] Cox, D.R., 1961. Tests of separate families of hypothesis. *Proc. of the fourth Berkeley Symposium on Mathematical Statistics and Probability,* **1,** 105-123.

[3] Cox, D.R., 1962. Further results on tests of separate families of hypotheses. *J. Roy. Statist. Soc.*, **B24**, 406-424.

[4] Draper, N.R. and Smith, H., 1981. *Applied Regression Analysis,* 2nd ed., John Wiley and Sons, New York.

[5] Efron, B., 1984. Comparing non-nested linear models. *J. Amer. Statist. Assoc.,* **79**, 791-803.

[6] Geisser, S. and Eddy, W.F., 1979. A predictive approach to model selection. *J. Amer. Statist. Assoc.*, **74**, 153-160.

[7] Hald, A., 1952. *Statistical Theory with Engineering Applications*. John Wiley and Sons, New York.

[8] Hotelling, H., 1940. The selection of variates for use in prediction with some comments on the general problem of nuisance parameters. *Ann. Math. Statist.,* **11,** 271-283.

[9] Montgomery, D.C. and Peck, E.A., 1982. *Introduction to Linear Regression Analysis.* John Wiley and Sons, New York.

[10] Pesaran, M.H., 1984. Asymptotic power comparisons of separate parametric families by Bahadur's approach. *Biometrika,* **71**, 245-252.

[11] Stone, M., 1977. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. Roy. Statist. Soc.*, **B39**, 44-47.

[12] Williams, E.J., 1959. *Regression Analysis.* John Wiley and Sons, New York.