On stochastic linear regression model selection

Cite as: AIP Conference Proceedings **2177**, 020068 (2019); https://doi.org/10.1063/1.5135243
Published Online: 04 December 2019

J. Peter Praveen, B. Mahaboob, Ranadheer Donthi, S. Vijay Prasad, and B. Venkateswarlu

AIP Conference Proceedings **2177**, 020068 (2019); https://doi.org/10.1063/1.5135243 **2177**, 020068

On Stochastic Linear Regression Model Selection

J.Peter Praveen[1], B.Mahaboob[1], Ranadheer Donthi[2], S.Vijay Prasad[1],
B.Venkateswarlu[3, a)]

[1]Department of Mathematics, Koneru Lakshmaih Education Foundation, Vaddeswaram,
Guntur Dist.AP-522502-India
[2]St. Martin's Engineering College, Dhulapally, Kompally, Hyderabad, Telangana, India 500100
[3]Department of Mathematics, Vellore Institute of Technology, Vellore, Tamilnadu

[a)] Corresponding author: venkatesh.reddy@vit.ac.in

Abstract. The research article primarily focuses on the criteria for selecting best stochastic linear regression model namely C_p -conditional mean square error prediction, Generalized Mean Squared Error criterion (GMSE) which comes out of the deficiencies of R^2 and $\overline{R}^2$ criteria. The most uncomfortable aspect of both R^2 and $\overline{R}^2$ measures is that they do not include a consideration of losses associated with choosing an incorrect model. C.L.Cheng et al, in 2014, in their research paper proposed the goodness of fit statistics based on the variants of R^2 for multiple measurement errors and also studied the asymptotic properties of the conventional R^2 and the proposed variants of R^2 like goodness of fit statistics analytically and numerically. M.HasheemPesaran et al, in 1994, in their paper discussed why both R^2 and $\overline{R}^2$ are inappropriate as a measure of fit and for model selection in the sense that their use does not guarantee that true model is chosen even asymptotically .D.Wallach et.al, in 1987, in their paper used the mean square error of prediction (MSEP) as a criterion for evaluating models for studying ecological and agronomic systems. M.Revan Ozkale, in 2009, in his paper introduced a new estimator by combining ideas underlying the mined and the ridge regression estimators under the assumption that the errors are not independent and identically distributes when there are stochastic linear restrictions on the parameter vector. David A. Mc Allester, in 2003, in his article, gave a PAC-Bayesian performance guarantee for stochastic model selection that is superior to analogous guarantees for deterministic model selection.

INTRODUCTION

Stochastic modelling is the art and science of using statistical techniques for the measurement of relationships between the variables. The formulation or specification of a stochastic model is an art just as using knowledge or architecture to design a building. In the specification of a stochastic model the most important variables are selected while the nonessential variables are discarded. The crucial relationships are formulated and incorporated in the model. Best stochastic models are like best architectural designs and can serve as prototype to be followed in the future investigation. A set of mathematical equations concerns with two or more variables refers to a mathematical model. By introducing an error random variable or a disturbance term the mathematical becomes statistical model or a stochastic model. Now- a- days modelling is a new and fertile area of research in mathematical and statistical sciences. Selection of the best model is an important part of stochastic model building. A large number of methods have been developed in the literature for selecting bets stochastic linear regression model.

Recent Trends in Pure and Applied Mathematics
AIP Conf. Proc. 2177, 020068-1–020068-5; https://doi.org/10.1063/1.5135243
Published by AIP Publishing. 978-0-7354-1924-7/$30.00

THE C_p -CONDITIONAL MEAN SQUARE ERROR PREDICTION CRITERION (OR) C_p -CRITERION FOR STOCHASTIC LINEAR REGRESSION MODEL SELECTION

There are deficiencies in the criteria R^2 and $\overline{R}^2$ need not be the most powerful of the criteria involving the quadratic form of residuals that have as their property: The expected value is minimized by the true model. The most uncomfortable aspect of the both R^2 and $\overline{R}^2$, measures is that they do not include a consideration of losses associated with choosing an incorrect model. That is they do not consider within a decision context the purpose for which the model is to be used. With the goal of eliminating this deficiency, a criterion based on Mean Square Prediction Error for stochastic linear regression model selection has been suggested and it is known as C_p -criterion.

Suppose $X = \lfloor X_{1_{n\times k1}} \quad X_{2_{n\times k2}} \rfloor$

where X_1 is $n \times k_1$ matrix of included variables X_2 is $n \times k_2$ matrix of excluded variables $k = k_1 + k_2$ is the total number of variables in the model .

Write the linear model as

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon \tag{1}$$

Where β_1 is $k_1 \times 1$ and β_2 is $k_2 \times 1$ vectors of parameters. Consider the subset model (restricted model) of (1)

$Y = X_1\beta_1 + u$ where $u = X_2\beta_2 + \varepsilon$

Define the mean square error loss in prediction as

$$\rho\left[X\beta^*, X\beta\right] = E\left[X\beta^* - X\beta\right]'\left[X\beta^* - X\beta\right]$$

Since we consider a subject model, $\beta_2 = 0$ and thus

$$\beta^* = \begin{bmatrix} \beta_1^* \\ O_{k_2\times 1} \end{bmatrix} = \begin{bmatrix} \left(X_1'X_1\right)^{-1} X_1'Y \\ O_{k_2\times 1} \end{bmatrix}$$

Here β_1^* is the restricted least squares estimator of β_1 .

$$\therefore \rho\left[X\beta^*, X\beta\right] =$$

$$E\left\{\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \beta_1^* \\ 0 \end{bmatrix} - \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}\right\}'$$

$$\left\{\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \beta_1^* \\ 0 \end{bmatrix} - \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}\right\}$$

$$= E\left\{\begin{bmatrix} X_1\beta_1^* - X_1\beta_1 \\ X_2O - X_2\beta_2 \end{bmatrix}'\begin{bmatrix} X_1\beta_1^* - X_1\beta_1 \\ X_2O - X_2\beta_2 \end{bmatrix}\right\}$$

$$= E\left\{\begin{bmatrix} \beta_1^* - \beta_1 \\ O - \beta_2 \end{bmatrix}'\left(X'X\right)\begin{bmatrix} \beta_1^* - \beta_1 \\ O - \beta_2 \end{bmatrix}\right\}$$

$$\text{Here } \beta_1^* = \left(X_1'X_1\right)^{-1} X_1'\left(X_1\beta_1 + X_2\beta_2 + \varepsilon\right)$$

$$= \beta_1\left(X_1'X_1\right)^{-1} X_1{}' X_2\beta_2 + \left(X_1'X_1\right)^{-1} X_1'\varepsilon$$

Under linear restriction $R\beta = \begin{bmatrix} O & I_{k_2} \end{bmatrix}\begin{bmatrix}\beta_1 \\ \beta_2\end{bmatrix} = r = O_{k_2.}$ if β^* be the restricted least squares estimator of β then it can be shown that

$$E\left[\beta^* - \beta\right]'\left(X'X\right)\left(\beta^* - \beta\right) = \sigma^2 k_1 + \beta_2' X_2' \left[I - X_1\left(X_1' X_1\right)^{-1} X_1'\right] X_2 \beta_2$$
$$= \sigma^2 k_1 + \left(Bias\right)^2$$

It can be shown that

$$\rho\left[X\beta^*, X\beta\right] = \sigma^2 k_1 + \left(Bias\right)^2 \qquad (2)$$

Where $(Bias)^2 = \beta'_2 X_2' \left[I - X_1\left(X_1' X_1\right)^{-1} X_1'\right] X_2 \beta_2$
=sum of the squares of the bias.

The criterion for stochastic linear model selection is to estimate the unknown parameters σ^2 and β_2 and choose the model with the smallest estimated mean square prediction error.

In terms of standardized risk, under squared error loss (2) can be written as

$$\frac{\rho\left[X_1\beta_1^*, X\beta\right]}{\sigma^2} = \frac{\sigma^2 k_1 + \left(Bias\right)^2}{\sigma^2} = k_1 + \frac{\left(Bias\right)^2}{\sigma^2}$$

Also we know that

$$E\left[Y - X_1\beta_1^*\right]'\left[Y - X_1\beta_1^*\right]$$
$$= E\left[\left(n - k_1\right)\sigma_1^{*2}\right] = \sigma^2\left(n - k_1\right) + \left(Bias\right)^2 \qquad (3)$$

$$(3) \Rightarrow \left(Bias\right)^2 = E\left[\left(n - k_1\right)\sigma_1^{*2}\right] - \left(n - k_1\right)\sigma^2 \qquad (4)$$

By substituting (4) in (2) we get

$$\frac{\rho\left(X_1\beta_1^*, X\beta\right)}{\sigma^2} = \frac{E\left[\left(n - k_1\right)\sigma_1^{*2}\right]}{\sigma^2} = \left(n - k_1\right) + k_1$$
$$= \frac{E\left[\left(n - k_1\right)\sigma_1^{*2}\right]}{\sigma^2} + \left(2k_1 - n\right) \qquad (5)$$

If the unknown parameters are replaced with the unbiased estimates in (5) we get

$$\frac{\rho\left(X_1\beta_1^*, X_1\beta\right)}{\hat{\sigma}^2} = C_p = \frac{\left(n - k_1\right)\sigma_1^{*2}}{\hat{\sigma}^2} + \left(2k_1 - n\right) \qquad (6)$$

Where $\hat{\sigma}^2 = \dfrac{\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)}{n - k}$

Now $(6) \Rightarrow C_p = \dfrac{\left(n - k_1\right)\left(1 - \bar{R}_1^2\right)}{1 - \bar{R}^2} + \left(2k_1 - n\right)$

$$\text{Since } \overline{R}_1^2 = 1-\left(\frac{n-1}{n-k_1}\right)\left(1-R_1^2\right) = 1-\frac{\hat{\sigma}_1^2}{\frac{Y^1Y}{n-1}}$$

$$\text{i.e.} \frac{1-\overline{R}_1^2}{1-\overline{R}^2} = \frac{\hat{\sigma}_1^2 / Y'Y / n-1}{\hat{\sigma}^2 / \frac{Y'Y}{n-1}}$$

When the subset model has small bias then σ_1^{*2} is approximately equal to $\hat{\sigma}^2$ and C_p is approximately equal to K_1 i.e C_p □k$_{1.}$

Under C_p criterion we calculate C_p values for the possible 2^k subsets of models and chose the model in which C_p □k$_{1.}$

GENERALIZED MEAN SQUARED ERROR (GMSE) CRITERION FOR STOCHASTIC LINEAR REGRESSION MODEL SELECTION

In computing two estimates which are both unbiased but have different variances one may prefer estimator with the smaller variance. In a class of all unbiased estimators one may find an estimator with minimum variance is known as Best Unbiased Estimator or Minimum Variance Unbiased Estimator (MVUBE) for the given parameter.

A different problem arises in comparing two biased estimators with different variables. In this case the Mean Squared Error (MSE) criterion may be used to compare these two estimators.

Suppose that $\hat{\theta}$ be a biased estimator of parameter θ. The MSE of $\hat{\theta}$ may be defined as

$$MSE(\hat{\theta}) = E\left(\hat{\theta}-\theta\right)^2, E\left(\hat{\theta}\right) \neq \theta$$

$$= E\left[\hat{\theta} - E(\theta) + E\left(\hat{\theta}\right) - \theta\right]^2$$

$$= E\left[\hat{\theta} - E\left(\hat{\theta}\right)\right]^2 + E\left[E\left(\hat{\theta}\right) - \theta\right]^2 + 2E\left[\hat{\theta} - E\left(\hat{\theta}\right)\right]\left[E\left(\hat{\theta}\right) - \theta\right]$$

$$\Rightarrow MSE(\hat{\theta}) = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2 + O$$

Thus under MSE criterion, a biased estimator may be preferred to one with a smaller or zero bias if its variance is sufficiently smaller to offset the larger bias.

Suppose that $\hat{\beta}$ be a estimator of parametric vector β in the standard stochastic regression model $Y = X\beta + \varepsilon$

The generalized MSE or Risk matrix of $\hat{\beta}$ is defined as

$$GMSE\left[\hat{\beta}, \beta\right] = E\left[\hat{\beta} - \beta\right]\left[\hat{\beta} - \beta\right]'$$

$$\text{such that} \quad E\left(\hat{\beta}\right) \neq \beta$$

$$= E\left[\hat{\beta} - E\left(\hat{\beta}\right) + E\left(\hat{\beta}\right) - \beta\right]\left[\hat{\beta} - E\left(\hat{\beta}\right) + E\left(\hat{\beta}\right) - \beta\right]'$$

$$= E\left[\hat{\beta} - E\left(\hat{\beta}\right)\right]\left[\hat{\beta} - E\left(\hat{\beta}\right)\right]' + E\left[\left(E\left(\hat{\beta}\right) - \beta\right)\right]\left[\left(E\left(\hat{\beta}\right) - \beta\right)\right]'$$

$$\Rightarrow GMSE\left[\hat{\beta},\beta\right] = Cov(\hat{\beta},\beta) + [Bias(\hat{\beta})][Bias(\hat{\beta})]'$$

(7)

$$= \begin{bmatrix} Co\ Variance \\ Matrix\quad for\ \hat{\beta} \end{bmatrix} + \begin{bmatrix} Bias\, squared \\ Matrix\ \ for\ \hat{\beta} \end{bmatrix}$$

The diagonal elements of (7) are the mean squared errors for each element $\hat{\beta}$ and the trace of the MSE matrix is equal to the Mean Squared Error Loss

Mean Squared Error Loss:

$$tr(E\left(\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)\right) = E\left[\left(\hat{\beta}-\beta\right)'\left(\hat{\beta}-\beta\right)\right] = E\left(L\left(\beta,\hat{\beta}\right)\right) = Risk$$

where $L\left(\beta,\hat{\beta}\right)$ is error loss.

Under GMSE criterion, an estimator $\hat{\beta}$ is equal to or superior to another estimator β^* if the MSE for every liner combination of elements of $\hat{\beta}$ is equal to or less than the MSE of the same combination of elements of β^*.

In other words estimator $\hat{\beta}$ is better than β^* under GMSE criterion if

$$E\left[\beta^*\right] \neq \beta\ ,\ E\left[\hat{\beta}\right] \neq \beta \text{ and}$$

$$E\left[\beta^* - \beta\right]\left[\beta^* - \beta\right]' - E\left[\hat{\beta} - \beta\right]\left[\hat{\beta} - \beta\right]' = \Delta$$

where Δ is positive semi definite matrix.

Under GMSE criterion the restricted least squares (RLS) estimator $\beta^*{}_{RLS}$ is a biased estimator and it is superior to unbiased OLS estimator $\hat{\beta}_{OLS}\quad for\ \ \beta$.

CONCLUSIONS

Owing to the deficiencies of the criteria R^2 and $\overline{R}^2$ as they are not most powerful the C_p conditional Mean square error Prediction criterion , generalized Mean squared Error criterion for stochastic linear regression model has been presented in the above research article. In the context of future research some special problems viz. model selection, miss-specification of the model and the selection of regressors specification errors along with their sources can be evaluated.

REFERENCES

1. C. L. Cheng, Shalabh, and G. Garg, J. of Multivariate Analysis, **126**, 131-153 (2014)
2. M. Hashem Pesaran and Richard J. Smith, Econometric a, **62**, 705-710 (1994).
3. D. Wallach and B. Goffnet, Biometrics, **43**, 561-573 (1987)
4. M. Revan Ozkale, J. of Multivariate Analysis, **100**, 1706-1716 (2009).
5. David A. Mc Allester, Machine Learning, **5**1, 5-21 (2003)
6. Akaike .H. “IEEE Transactions on Automatic control (1974)
7. BysonJ. T. Morgan, CRC Press, 978-1-5848-666-2 (2008).
8. Berry L. Nelson, McGraw Hill, 978-0070462137 (1995).
9. Burnham K.P. and Anderson A.R. Springer-Verlag, New-York, (1998).
10. Pinsky, M.A and Samuel Kanlin, Academic Press, 978-0-12-381416-6 (2011).