Statistics and Applications {ISSN 2454-7395 (online)}
Volume 16 No. 2, 2018 (New Series), pp 37-47

Diagnostics of Multicollinearity in Multiple Regression Model for Small Area Estimation

Neeraj Tiwari and Ankuri Agarwal
Kumaun University, SSJ Campus, Almora, ,Uttarakhand

Received: February 27, 2017; Revised: November 30, 2017; Accepted: January 09, 2018

Abstract

The purpose of this paper is to discuss the multicollinearity problem in regression models for small area estimation and propose Ridge Regression Model (RRM) to deal with the problem of multicollinearity .The proposed model has been empirically compared with the existing Multiple Linear Regression (MLR) Model. Analysis of data obtained from a survey carried out from Directorate of Economics and Statistics, Uttarakhand, India revealed that RR methodology performs better as compared to MLR model in terms of the criterion of MSE. The approach does not require any additional survey or conducting extra crop cutting experiments (CCE) for crop production estimate at the district level.

Key words : Small area estimation, MLR Model, Multicollinearty, RR Model.

---

1. Introduction

The concept of Small area estimation has received considerable attention in recent years due to growing demand for reliable small area statistics. The term "small area" generally refers to a small geographical area such as a county or a community development block or still smaller units like Mandals or Village Panchayats in India for which limited information is available from the primary source of data. Small Area Estimation has been used to produce effective estimates of characteristics of interest such as means, counts, quartiles, etc and to assess their precision. Some of the other terms used synonymous to small area are "small domain", "minor domain", "local areas" and "small sub-domain". The history of using small area statistics goes back to the eleventh century. For operative social services most countries of the world (both developed and developing) have adopted a decentralized approach. Local level statistics have become increasingly important in policy decisions, resource allocation, monitoring of programs and evaluation of initiatives. Small area estimation methodologies are also beneficial for business organizations and researchers who are interested in estimates for regional small domains but lack adequate funds for a large-scale survey.

Corresponding Author: Neeraj Tiwari
E-mail: kumarn_amo@yahoo.com

Small Area Estimation (SAE) techniques assist state and local government in making various decisions such as how to allocate resources in small areas. This technique overcomes the problem of small sample sizes to produce small area estimates and also improves the quality of direct survey estimates obtained from the sample in each small area. SAE has emerged as a sound statistical procedure to create the estimates and an evaluating system to ensure the reasonable estimates for small areas.

Various researchers have attempted to deal with the different techniques of small area estimation. Earlier reviews on the topic of small area estimation focused on demographic methods for population estimation in post censual years. Panse et al. (1966) introduced two phase sampling to provide crop yield estimates. Singh (1968) introduced the method of double sampling in agriculture. Purcell and Kish (1979) reviewed demographic methods as well as statistical methods of estimation for small domains. An excellent review provided by Zidek (1982) introduces a criterion that can be used to evaluate the relative performance of different methods for estimating the populations of local areas. McCullagh and Zidek (1987) elaborated this criterion more precisely. Stasny et al. (1991) presented a technical report on county estimates of wheat production in Ohio state university, Ohio. Reviews by Rao (1986) and Chaudhuri (1992) covered more recent techniques as well as traditional methods of small area estimation. Sisodia and Singh (2001) developed an estimator for crop production at block level using crop cutting and other related information at district level. Bellow and Lahiri (2011) discussed an empirical best linear unbiased prediction approach to small area estimation of crop parameters assuming linear mixed model. Chandra et al. (2011) employed small area estimation technique to derive model based estimates of proportion of poor households at district level in the state of U.P in India. Daniel Elazer (2012) used the Lagrange’s Multiplier method to adapt the PQL (Penalised quasi likelihood) approach applied to random effects logistic models. Tzavidis et al. (2012) described an application of small area estimation to agriculture business survey data by using empirical best linear unbiased predictor (EBLUP) and model based direct estimators. Chen and Lahiri (2012) discussed an alternative method to overcome the inefficiency of Design-Based methods for making inferences about small area proportions for rare events. Sud et al. (2015) demonstrated an application of spatial information to estimate the production at district level for the state of Uttar Pradesh. Abdelgadir and Eledum (2016) discussed the comparative study of Ridge Regression and Principle component Regression with Application to deal with the problem of multicollinearity.

The model suggested by Sud et al. (2015) had a drawback that regressor variables included in the model are highly correlated and hence the problem of multicollinearity is said to exist. The present study was carried out to describe computer-based SAE approach to resolve the multicollinearity problem in auxiliary variable. The performance of the proposed model is compared with the existing Multiple Regression Model by making a comparison between predicted and measured values. Results of the study show that the proposed model appears to perform better than the Multiple Linear Regression Model proposed by Sud et al. (2015). In Section 2, preliminaries about MLR and Multicollinearity are discussed. In Section 3 we discuss the effects of multicollinearity on the least square estimates of regression coefficients. Section 4 proposes a Ridge Regression (RR) Model to deal with the problem of multicollinearity. Model specification and analysis on the basis of an empirical study is presented in Section 5. Concluding remarks and future avenues of research are discussed in Section 6.

2. Multiple Linear Regression Model and Multicollinearity

Multiple regression analysis is a multivariate statistical technique used to examine the relationship between a single dependent variable and a set of independent variables, whose values are known to predict the single dependent variable. Several MLR methodologies proposed by different authors for estimation of crop production at Block/Panchayat level have been discussed in Section 1. Sisodia and Singh (2012) described scale down approach using MLR model to obtain block level estimates from the district level crop production as follows:

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \varepsilon_i \qquad (1)$$

where Y_i is the crop production in the i^{th} year (i=1, 2, ...,n), X_{ij} is the value of j^{th} auxiliary variable (j=1, 2,...,p) in the i^{th} year, $\beta' = (\beta_0, \beta_1, \beta_2, ..., \beta_P)$ is the vector of unknown parameters and ε_i is error term. It is assumed that ε_i follows normal distribution with mean 0 and variance σ^2. Let the fitted model be denoted as,

$$\hat{Y}_i = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j \hat{X}_{ij} \qquad (2)$$

The least square estimate of β is :

$$\hat{\beta} = (X'X)^{-1} X'Y$$

It will be convenient to assume that the regressor variables and the responce have been centered and scaled to unit length. Consequently, $X'X$ is a $P \times P$ matrix of correlation between the regressors and X'y is a $P \times 1$ vector of correlations between the regressors and the response. Let the jth column of the X matrix be denoted X_j, so that $X = [X_1, X_2, ..., X_p]$. Thus, X_j contains the n levels of the *jth* regressor variable. We may formally define multicollinearity in terms of the linear dependence of the columns of X. The vectors $X_1, X_2, ..., X_p$ are linearly dependent if there is a set of constants $t_1, t_2, ..., t_p$ not all zero, such that $\sum_{j=1}^{p} t_j X_j = 0$. If this equation holds exactly for the subset of the columns of X, then the rank of the $X'X$ matrix is less than p and $(X'X)^{-1}$ does not exist. However, suppose that given equation is approximately true for some subset of the coloums of X. Then there will be a near-linear dependency in $X'X$ and the problem of ill-conditioning in the $X'X$ matrix will arise. Every dataset will suffer from multicolinearity to some extent unless the coloums of X are orthogonal and the presence of multicolinearity makes the usual least-square analysis of the regression model dramatically inadequate.

3. Effects of Multicollinearity

The presence of multicollinearity has a number of potentially serious effects on the least-squares estimates of the regression coefficients. Some of these effects may be easily demonstrated. Suppose that there are only two regressor variables, x_1 and x_2. The model, assuming that x_1, x_2 and y, are scaled to unit length, is

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{3}$$

and the least-squares normal equations are

$$(X'X)\hat{\beta} = X'y$$

$$\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \times \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

Where r_{12} is the simple correlation between x_1 and x_2 and r_{jy} is the simple correlation between x_j and y, j= 1, 2. Now the inverse of $(X'X)$ is

C=

$$(X'X)^{-1} = \begin{bmatrix} \frac{1}{1-r_{12}^2} & \frac{-r_{12}}{1-r_{12}^2} \\ \frac{-r_{12}}{1-r_{12}^2} & \frac{1}{1-r_{12}^2} \end{bmatrix} \tag{4}$$

and the estimates of the regression coefficients are

$$\hat{\beta}_1 = \frac{r_{1y} - r_{12} r_{2y}}{\left(1 - r_{12}^2\right)}, \; \hat{\beta}_2 = \frac{r_{2y} - r_{12} r_{1y}}{\left(1 - r_{12}^2\right)}, \tag{5}$$

If there is strong multicollinearity between x_1 and x_2, then the correlation coefficient r_{12} will be large. From Equation (4) we see that as $|r_{12}| \to 1$, $Var(\hat{\beta}_j) = C_{jj}\sigma^2 \to \infty$ and $Cov(\hat{\beta}_1, \hat{\beta}_2) = C_{12}\sigma^2 \to \pm\infty$ depending on whether $r_{12} \to +1$ or $r_{12} \to -1$. Therefore, strong multicollinearity between x_1 and x_2 results in large variances and covariance's for the least-squares estimators of the regression coefficients. This implies that different samples taken at the same *x* levels could lead to widely different estimates of the model parameters.

When there are more than two regressor variables, multicollinearity produces similar effects. It can be shown that the diagonal elements of the $C = (X'X)^{-1}$ matrix are

$$C_{jj} = \frac{1}{1 - R_j^2}, \qquad j=1, 2, \ldots, \text{p} \tag{6}$$

Where R_j^2 is the coefficient of multiple determinations from the regression of x_j on the remaining *p*-1 regressor variables. If there is strong multicollinearity between x_j and any subset of the other *p*-1 regressors, then the value of R_j^2 will be close to unity. Since the variance of $\hat{\beta}_j$ is $Var(\hat{\beta}_j) = C_{jj}\sigma^2 = (1 - R_j^2)^{-1}\sigma^2$, strong multicollinearity implies that the variance of the least-squares estimate of the regression coefficient β_j is very large. Generally, the covariance of $\hat{\beta}_i$ and $\hat{\beta}_j$ will also be large if the regressors x_i and x_j are involved in a multicollinearity relationship.

Multicollinearity also tends to produce least-square estimates $\hat{\beta}_j$ that are too large in absolute value. To see this, consider the squared distance from $\hat{\beta}$ to the true parameter vector β, for example,

$$L_1^2 = (\hat{\beta}-\beta)'(\hat{\beta}-\beta) \tag{7}$$

Multicollinearity is not only responsible for the large variances and covariance of regression coefficients but also effects the predicted value of the estimator.

The expected squared distance, $E(L_1^2)$, is

$$\begin{aligned} E(L_1^2) &= E(\hat{\beta}-\beta)'(\hat{\beta}-\beta) \\ &= \sum_{j=1}^{p} E(\hat{\beta}_j-\beta_j)^2 \\ &= \sum_{j=1}^{p} Var(\hat{\beta}_j) \\ &= \sigma^2 Tr(X'X)^{-1} \end{aligned} \tag{8}$$

Where the trace of a matrix (abbreviated by Tr) is just the sum of the main diagonal elements. When there is multicollinearity present, some of the eigenvalues of $X'X$ will be small. Since the trace of a matrix is also equal to the sum of its eigenvalues, Equation (8) becomes

$$E(L_1^2) = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j} \tag{9}$$

Where $\lambda_j > 0$, j=1, 2,…,p, are the eigen values of $X'X$. Thus, if the $X'X$ matrix is ill-conditioned because of multicollinearity, at least one of the λ_j will be small, and Equation (9) implies that the distance from the least-squares estimate $\hat{\beta}$ to the true parameter β may be large. Equivalently, we can show that,

$$E(L_1^2) = E(\hat{\beta}-\beta)'(\hat{\beta}-\beta)$$

$$= E(\hat{\beta}'\hat{\beta} - 2\hat{\beta}'\beta + \beta'\beta)$$

and $E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 Tr(X'X)^{-1}$ T (10)

That is, the vector $\hat{\beta}$ is generally longer than the vector β. This implies that the method of least squares produces estimated regression coefficients that are too large in absolute value.

4. Methods for dealing with Multicollinearity: Ridge Regression

Ridge Regression (RR) has been introduced by Hoerl and Kennard (1970). They suggested a small positive number K>0 to be added to the diagonal element of the $X'X$ matrix from the multiple regression and resulting estimator is obtained as

$$\hat{\beta}_{RR} = \left(X'\hat{X} + KI'\right)^{-1} X'Y, \tag{11}$$

where I is a unit matrix and K is a constant selected by the analyst, K>0. It is to be noted that if K=0, then Ridge estimator is the Least Square estimator. According to Hoerl, Kennard and Baldwin (1975), the appropriate choice for K is,

$$K = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}, \tag{12}$$

where $\hat{\beta}$ and $\hat{\sigma}^2$ are found from the least square solution.

Following Montgomery and Peck (1982), the sum of square due to regression is decomposed to define weight for the j^{th} auxiliary variable on the basis of its relative contribution in the model, given by

$$w_j = \frac{\text{sum of square due to j}^{\text{th}} \text{ auxiliary variable}}{\text{sum of square due to regression}}$$

Using these weights, an estimator of crop production for a block in a given year is constructed as follows:

$$\hat{Y}_q = \left[\sum_{j=1}^{p} w_j\, x_j(q)\right]\overset{\wedge}{\bar{Y}}; \quad q = 1, 2, ..., Q,$$

where Q is the total number of blocks in the given district and $x_j(q)$ is the value of j^{th} auxiliary variable in q^{th} block in the given year and $\overset{\wedge}{\bar{Y}}$ is the estimated average yield of a crop based on the fitted model (2) in the same year .

In order to find out the relative contribution of individual auxiliary variable, the auxiliary variable named as total area under rice (X_1), irrigated area under rice (X_2) and fertilizer used (X_3) were included in the model depending upon the order of magnitude of correlation coefficients between auxiliary variables and crop production (Y) .On the basis of their related contribution the value of weight were calculated using, SPSS16.0.

5. Model Specification and Analysis: An Empirical Study

5.1. Data

Time series data on production of rice, area under rice, irrigated area under rice and fertilizer consumption pertaining to the period 1990-91 to 2002-03 for Uttarakhand state of India were taken from the Bulletin of Agricultural Statistics, published by the government of Uttarakhand, India. The district wise data on auxiliary variables in Uttarakhand state were also taken for the year 2000-01 from Agricultural bulletin, published by the government of Uttarakhand.

5.2. Method

First, the correlations matrix among variables was calculated to test the linear relationship between regressor variables, then least squares method was conducted to construct a linear model between production of rice and total area under rice (*TA*), irrigated area under rice (*IR*) and fertilizer consumption (*FC*). Table 1 represents the correlation matrix among explanatory variables. From Table 1, we find that the variable total area (*TA)* is highly positive correlated (r = 0.75) with variable irrigated area (*IR)*.

Table 1: Correlation matrix of explanatory variables

Column1	TA	IR	FC
TA	1	0.75	0.156
IR		1	0.385
FC			1

Since two regressor variables are highly correlated, there are potential multicollinearity problem in data. The full quadratic model for the Rice production at district level in Uttarakhand is

$$y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \beta_7 X_1^2 + \beta_8 X_2^2 + \beta_9 X_3^2 + \varepsilon$$

Table 2 represents the least square estimates of the model parameters in simple linear case while Table 3 represents the model parameters in quadratic form.

Table 2: **The Least Square Parameter estimates for Linear Regression Model**

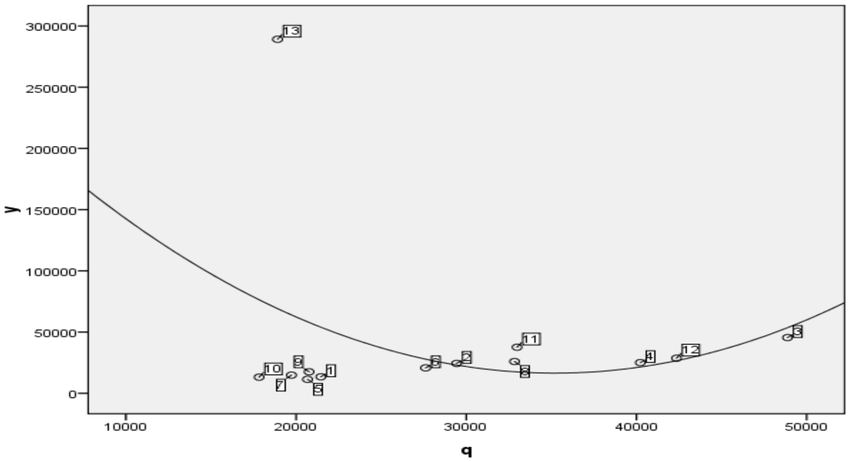
Model		Unstandardized Coefficients		Standardized Coefficients
		B	Std. Error	Beta
1	(Constant)	15203.886	174952.845	
	TA	2.341	.893	.935
	IR	-.704	.834	-.322
	FC	.001	.004	.051

The estimated values of the regression coefficients is presented in Table 3 and graph between predicted and observed values obtained from quadratic model is represented in Figure 1.

Table 3: Parameter estimates for Quadratic Regression Model

Parameters	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
Estimate	-8.371E6	43.638	62.888	-.686	.001	2.550E-6	-3.078E-7	-4.436E-5	5.855E-6	1.078E-9
Std. Error	4.213E6	39.468	54.455	.637	.001	.000	.000	.000	.000	.000

Fig 1: Non Linear *R*- square value curve



From the Table 3, we conclude that when multicollinearity is suspected, the least square estimates of the regression coefficient gives very poor results and limits the usefulness of regression model for inference and prediction.

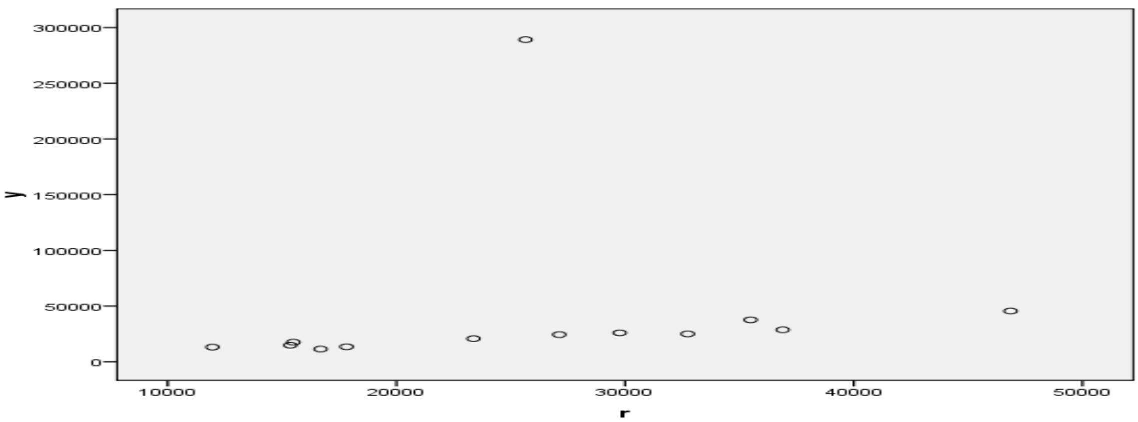
5.3 Application of Ridge Regression

To determine the best model fitted in the data using ridge regression, firstly we present methods of choosing *K*. According to Hoerl, Kennard and Baldwin (1975), the appropriate choice of *K* is given by (12). Applying this formula, we get,

$$K = \frac{3 \times 0.866}{0.98051} = 2.6497$$

Now the value of $\hat{\beta}_{RR}$ can be obtained by putting the value of *K* in (11). The graph between the predicted and actual value is shown in Figure 2.

Fig 2: Predicted and observed values curve for Ridge Regression Model



5.4 Comparison among Least Squares Method, Non linear Regression and Ridge Regression

The result of ordinary least square method, non linear regression model and ridge regression method are presented in Table 4. The Multicollinearity problem between the independent variables for the production model has been solved by using Ridge regression (*RR*). From Table 4, we see that the *RR* method yields best result when multicollinearity is present in the data. The MSE for RR model was found to be minimum.

Table 4: The Results of OLS, Non Linear Regression Model and RR estimates of parameters

	OLS	**RR**	**Non Linear Regression Model**
Max VIF	58.21	1.078	2.081
C.V	0.085	0.083	0.089
MSE	45.295	34.006	39.978

The estimated values of rice production at district level in Uttarakhand using three regression methods, in simple linear regression, Non linear regression and Ridge Regression are represented in Table 5 given below.

Table 5: Actual and estimated value of rice production in year 2001

2001								
Sl.No	District	Dependent Variable	Independent Variable			Three Estimates for Rice Production		
		Actual Rice Production(Y) (kg)	Area Under Production(X1) (hac)	Irrigated Area(X2) (hac)	Fertilizer Used(X3) (kg/hac)	Estimated Rice Production using MLR Model (kg)	**Estimated Rice Production using Non Linear Regression Model (kg)**	**Estimated Rice Production using Ridge Regression Model (kg)**
1	Chamoli	13503	13364	1352	27	24936.9	**21456.7**	**17821.2**
2	Dehradun	24396	21616	11724	2235	31692.36	**29432.3**	**27126.4**
3	Haridwar	45520	22633	20334	15696	51604.39	**48875.8**	**46854.62**
4	Pauri	25051	9307	6905	49	51496.18	**40225.5**	**32735.08**
5	Rudraprayag	11419	12750	2351	25	21095.21	**20673.78**	**16681.7**
6	Tahari	20753	10316	7176	71	29553.5	**27612.02**	**23372.9**
7	Uttarkashi	14861	20696	4896	90	23761.5	**19728.12**	**15361.3**
8	Almora	25969	14239	5656	136	46981.04	**32834.2**	**29763.8**
9	Bageshwar	17553	11458	5424	141	32577.19	**20764.4**	**15495.6**
10	Champawat	13187	15546	2118	109	25843.04	**17824.9**	**11956.2**
11	Nainital	37644	23070	13446	3217	36841.05	**32985.9**	**35487.8**
12	Pithodaghar	28672	94753	4501	175	52072.11	**42356.4**	**36892.8**
13	Udham Singh Nagar	289134	11108	94193	42182	226787.2	**18912.6**	**25648.9**

From Table 5, it is clear that the estimated values of the rice production using *RR* model (given in bold letters in the last column of the Table 5) are quite close to the actual values (given in the third column of the Table 5) as compared to the estimated values obtained through the MLR model and Non Linear regression model.

The present study indicates *RR* model provide better small area estimates than *MLR* and non linear model. The *RR* model may be recommended for use in small area estimation.

6. Conclusion and Discussion

Small area estimation is becoming important in survey sampling due to a growing demand for reliable small area statistics from both public and private sectors. Direct survey estimates for small areas are likely to yield unacceptable large standard error due to small sample sizes. This makes it necessary to "borrow strength" form related areas to find more accurate estimates. In the present study, three auxiliary variables viz., total area under production, irrigated area under production and fertilizer used are used to borrow strength for the estimation of rice production at district level in Uttarakhand state of India.

In this paper, we demonstrate an application of small area estimates by using *RR* model. Least squares (*LS*) method is the oldest techniques for estimating the parameters of linear regression model under some assumptions. However, if these assumptions are violated, LS method does not assure the desirable results. The influence of the multicollinearity is one of these problems, which occurs when the number of explanatory variable is relatively large in comparison to the sample or if the variables are almost collinear. The Ridge regression (*RR*) method is used to deal with this problem. Time series data on production of rice, area under rice, irrigated area under rice and fertilizer consumption pertaining to the period 1990-91 to 2002-03 for Uttarakhand state of India is taken from the Bulletin of Agricultural statistics, published by the government of Uttarakhand, India. The district wise data on auxiliary variables in Uttarakhand state were also taken for the year 2000-01 from Agricultural Bulletin, published by the government of Uttarakhand.

It has been concluded that the estimates of district level rice production obtained from the *MLR* and Non Linear Regression Models are far away from their actual district level rice production values as compare to the estimated production obtained through the proposed RR model. The RR models are now widely used for obtaining estimates when multicollinearity is present in data. On the basis of the results obtained through the empirical study it may be concluded that the *RR* model perform better than the existing models for small area estimation.

Acknowledgement

The authors are thankful to the Editor and one referee for their constructive comments, which have made considerable improvement in presentation of this manuscript.

References

Abdelgadir, G.A., and Eledum, H. (2016). A comparison study of rigid regression and principle component regression with application. *International Journal of Research,* **3(8).**

Bellow, M.E., and Lahiri, P.S. (2011). An empirical best linear unbiased prediction approach to small area estimation of crop parameters. National Agricultural Statistics Service, University of Maryland.

Chandra, H. (2010). Small area estimation with binary variables. *Journal of the Indian Society of Agricultural Statistics,* **64(3),** 367-374.

Chaudhuri, A. (1992). *Small domain statistics: a review*. Technical report ASC/92/2. Indian Statistical Institute, Calcutta.

Chen, S., and Lahiri, P. (2012). Inference on Small Area Proportions, Joint Program of Survey methodology, University of Maryland, College Park, USA.

Elazar, D. (2012). Small area estimation under additive constraints to published direct survey estimates. *Journal of the Indian Society of Agricultural Statistics,* **66(1)**, 15-30.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression biased estimation for nonorthognal problems. *Technometrics*. **12(1),** 55-67.

Hoerl, A.E., Kennard, R.W., and Baldwin, K.F (1975). Ridge regression: some simulations. *Communications in Statistics*, **4(2),** 105-123.

McCullagh, P. and Zidek, J. (1987). Regression methods and performance criteria for small area population estimation. *Small Area Statistics,* 62-74. Wiley, New York.

Montgomery, D.C. and Peck, E.A. (1982). Introduction to linear regression analysis. *John Wiley & Sons*, Canada.

Panse, V.G, Rajogopalam, M. and Pillai, S. (1966). Estimation of crop yields for small areas. *Biometrics*, **66**, 374-388.

Purcell, N.J. and Kish, L. (1979). Estimation for small domain. *Biometrics*, **35**, 365-384.

Rao, J.N.K. (1986). Synthetic estimators, SPREE and best model based predictors. Conference on Survey Research Methods in Agriculture, **1-16**. US. Dept. Agriculture, Washington, DC.

Singh, D. (1968). Double sampling and its application in agriculture. In *Contributions in Statistics and Agricultural Sciences*, Indian Society of Agricultural Statistics, New Delhi, 213-226.

Sisodia, B.V.S. and Singh, A. (2012). On small area estimation techniques- An application in agriculture. *Journal of the Indian Society of Agricultural Statistics,* **66(3),** 391-400.

Sisodia, B.V.S. and Singh, A. (2001). On small area estimation- an empirical study. *Journal of the Indian Society of Agricultural Statistics,* **54(3)**, 303-316.

Stasny, E.A., Goel, P.K. and Rumsey, D.J. (1991). *County estimates of wheat production*. Technical Report 458, Department of Statistics, The Ohio State University, Columbus, Ohio.

Sud, U.C., Aditya, K. and Chandra, H. (2015). District level crop yield estimation under spatial small area model. *Journal of the Indian Society of Agricultural Statistics,* **69(1)**, 49-56.

Zidek, J.V. (1982). *A review of methods for estimating the populations of local areas*. Technical Report 824, University. British Columbia, Vancouver.

Tzavidis, N., Chambers, R., Salvati, N. and Chandra, H. (2012). Small area estimation in practice: an application to agricultural business survey data. *Journal of the Indian Society of Agricultural Statistics,* **66**, 213-228.