Rank Set Sampling in Improving the Estimates of Simple Regression Model

M. Iqbal Jeelani
Division of Agricultural Statistics, SKUAST-K, India
jeelani.miqbal@gmail.com

S.A. Mir
Division of Agricultural Statistics, SKUAST-K, India
mir_98@msn.com

I. Khan
Division of Agricultural Statistics, SKUAST-K, India
ik400123@gmail.com

S.Maqbool
Division of Agricultural Statistics, SKUAST-K, India
showkatmaq@gmail.com

Nageena Nazir
Division of Agricultural Statistics, SKUAST-K, India
nazir.nageena@gmail.com

Fehim Jeelani
Division of Agricultural Statistics, SKUAST-K, India
faheemwani1@gmail.com

Abstract

In this paper Rank set sampling (RSS) is introduced with a view of increasing the efficiency of estimates of Simple regression model. Regression model is considered with respect to samples taken from sampling techniques like Simple random sampling (SRS), Systematic sampling (SYS) and Rank set sampling (RSS). It is found that R^2 and Adj R^2 obtained from regression model based on Rank set sample is higher than rest of two sampling schemes. Similarly Root mean square error, p-values, coefficient of variation are much lower in Rank set based regression model, also under validation technique (Jackknifing) there is consistency in the measure of R^2, Adj R^2 and RMSE in case of RSS as compared to SRS and SYS. Results are supported with an empirical study involving a real data set generated of *Pinus Wallichiana* taken from block Langate of district Kupwara.

Keywords: Rank set sampling, Simple Regression Model, *Pinus Wallichiana.*

1. Introduction

Cost-effective sampling methods are of major concern in statistics, especially when the measurement of the characteristic of interest is costly and time consuming. Environmental monitoring and assessment, Forest surveys etc; require observational data as opposed to data obtained from controlled experiments. Obtaining such data requires identification of sample units to represent the population of interest, followed by selection of particular units to quantify the desired characteristics. The most expensive

part of this process is laboratory analysis, while identification of potential sample units is comparatively simple. We can therefore achieve great observational economy if we are able to identify a large number of sample units to represent the population of interest, yet only have to quantify a carefully selected subsample. This potential for observational economy was recognized for estimating mean pasture and forage yields in the early 1950s, when McIntyre (1952) proposed a method, later coined *Rank set sampling* (RSS) by Halls and Dell (1966). McIntyre (1952) developed the procedure of RSS to find a more efficient method to estimate the yield of pastures. Measuring yield of pasture plots requires mowing and weighing the hay which is time-consuming. However experience can be used to rank by eye inspection to a large extent accurately the yields of a small number of plots without actual measurement. McIntyre (1952) adopted the sampling scheme, where, each time a random sample of k pasture lots is taken and the lots are ranked by eye inspection with respect to the amount of yield from the first sample, the lot with rank 1 is taken for cutting and weighing. From the second sample, the lot with rank 2 is taken, and so on. When each of the ranks from 1 to k has an associated lot being taken for cutting and weighing, the cycle repeats over again and again until a total of m cycles are completed. McIntyre (1952) observed that the relative efficiency, defined as the ratio of the variance of the mean of a simple random sample and the variance of the mean of a ranked set sample of the same size, is not much less than $(k+1)/2$ for symmetric or moderately asymmetric distributions, and that the relative efficiency diminishes with increasing asymmetry of the underlying distribution but is always greater than 1. He observed that by using the same sample size RSS provides an increased precision as compared to simple random sampling (SRS).

RSS has been used to estimate shrub phytomass, Martin *et al.,* (1980), mass herbage in a paddock, Cobby *et al.,* (1985) in order to achieve observational economy and increased precision over simple random sampling (SRS). Gilbert (1995) recommended it for environmental research quieries such as estimating plutonium soil concentrations and Nussbaum and Sinha (1997) discussed the problem of quality testing reformulated gasoline with reference to RSS. Samawi (1997) proposed a regression-type estimator based on RSS. They demonstrated that this estimator is always more efficient than the regression estimator using SRS. You (2009) suggested two practical examples from fishery research that RSS incorporates information on concomitant variables that are correlated with the variable of interest in the selection of samples, to demonstrate the approach: site selection for a fishery-independent monitoring survey in the Australian northern prawn fishery (NPF) and fish age prediction by simple linear regression modeling a short-lived tropical clupeoid. Both the strategies were based on RSS. The relative efficiencies of the new designs were derived analytically and sampling strategies were developed based on the idea of ranked set sampling (RSS) to increase efficiency and reduce the cost of sampling in fishery research. Many sampling methods have been suggested for estimating population median or second quartile in a situation when sampling units in a study can be ranked easily than quantified. Kamarulzaman (2011) illustrated the superiority of RSS over SRS through simulation studies. Ranked set sampling is used for obtaining the sub-sample from the set of non-respondents to tackle non-response situations. The use of RSS appears as the best alternative as compared to SRS, Gajendra and Bouza (2012). The efficiency of Rank set sampling has been very well demonstrated under Stratification by Jeelani *et al* (2014,a). Some new allocation

schemes for handling Non-Response problems in Rank set sampling under stratification has been suggested by Jeelani *et al* (2014,b).

In this paper data on Pinus Wallichiana is utilized. The data on *Pinus Wallichiana* was taken from block Langate of District Kupwara from Forest department J&K. *Pinus Wallichiana* is a coniferous evergreen tree native to the Himalaya, Karakoram and Hindu Kush mountains, from eastern Afghanistan east across northern Pakistan and India to Yunnan in southwest China. It grows in mountain valleys at altitudes of 1800–4300 m (rarely as low as 1200 m), between 30 m and 50 m in height. It favours a temperate climate with dry winters and wet summers. This tree is often known as 'Bhutan pine', (not to be confused with the recently described Bhutan white pine, *Pinus bhutanica*, a closely related species). Other names include 'blue pine', 'Himalayan white pine' and 'Himalayan blue pine'. In the past, it was also known by the invalid botanic names *Pinus griffithii* McClelland or "*Pinus excelsa*" Wall., *Pinus chylla* Lodd. when the tree became available through the European nursery trade in 1836, nine years after Dr Wallich first introduced seeds to England. The leaves ("needles") are in fascicles (bundles) of five and are 12–18 cm long. They are noted for being flexible along their length, and often droop gracefully. The cones are long and slender, 16–32 cm, yellow-buff when mature, with thin scales; the seeds are 5–6 mm long with a 20–30 mm wing. Typical habitats are mountain screes and glacier forelands, but it will also form old growth forests as the primary species or in mixed forests with deodar, birch, spruce, and fir. In some places it reaches the tree line. The wood is moderately hard, durable and highly resinous. It is a good firewood but gives off a pungent resinous smoke. It is a commercial source of turpentine which is superior quality than that of *P. roxburghii* but is not produced so freely. It is also a popular tree for planting in parks and large gardens, grown for its attractive foliage and large, decorative cones. It is also valued for its relatively high resistance to air pollution, tolerating this better than some other conifers.

2. Material Methods

In this paper simple linear regression model is considered with respect to samples taken from the sampling techniques like simple random sampling (SRS), systematic sampling (SYS) including rank set sampling (RSS).The method of estimation used in this paper is the ordinary least squares method (OLS). Also, bivariate ranked set sample is introduced, Al-Saleh and Zheng (2002). Finally regression models based on different identified sampling schemes are compared with each other based on validation technique (Jackknifing), which is a sample reuse technique, Quenouille (1949). A bivariate rank set sampling given by Al-Saleh and Zheng (2002) can be obtained as follows:

Suppose *(X, Y)* is a bivariate random vector with the joint probability density function *(jpdf) f (x, y).*

1. A random sample of size k^4 is identified from the population and randomly allocated into k^2 pools of size k^2 each, where each pool is a square matrix with k rows and k columns.

2. In the first pool, identify by judgment the minimum value w.r.t. the first characteristic, for each of the *k* rows.

3. For the *k minima* obtained in Step 2, choose the pair that corresponds to the minimum value of the second characteristic, identified by judgment, for actual quantification. This pair, which resembles the label (1, 1), is the first element of the bivariate rank set sample.

4. Repeat Steps 2 and 3 for the second pool, but in step 3, the pair that corresponds to the second minimum value w.r.t the second characteristic is chosen for actual quantification. This pair resembled the label (1, 2).

5. The process continues until the label (*k*, *k*) is resembled from the $(k^2)^{th}$ (last) pool. The above procedure produces a Bivariate rank set sample of size k^2. Thus we have k^2 observations denote by: $(X[i](j), Y(i)[j])$, $i=1,2\ldots k$ and $j=1,2,\ldots k$.

6. The procedure can be repeated *m* times to obtain a sample of size $n = k^2m$ which will be denoted by $(X[i](j)k, Y(i)[j]k)$, $i=1,2\ldots k$ and $j=1,2,\ldots k$, $k=1,2,m$.

In this article, effect of BVRSS on simple regression model is investigated utilizing Y and X as random variables. An important feature of BVRSS is that ranking is done on both variables Y and X simultaneously**.** Inference for simple regression model parameters (α, β) using asymptotic results are given. The simple regression model of the two variables *Y* and *X* is defined by: $y_{(i)[j]t} = a + \beta X_{[i](j)t} + E_{ijt}$ where *a* is the model intercept, β is the model slope and E_{ijt} is the random error. The assumptions needed here for the purpose of parameters estimation are the mean of the error is zero, its variance is finite and they uncorrelated. Also X_i and E_i are independent. Then the least squares estimators of *a* and β are given by:

$$\hat{\alpha}_{bvrss} = \overline{Y}_{bvrss} - \hat{\beta}_{bvrss}\,\overline{X}_{bvrss} \tag{2.1}$$

$$\hat{\beta}_{bvrss} = \frac{\sum_{t.j.i}(X_{[i](j)t} - \overline{X}_{bvrss})(y_{(i)[j]t} - \overline{Y}_{bvrss.}}{\sum_{t.j.i}(X_{[i](j)kt} - \overline{X}_{bvrss})^2})\overline{X}_{bvrss} \tag{2.2}$$

where

$$\overline{X}_{bvrss} = \frac{\sum_{t.j.i} X_{[i](j)t}}{n} \text{ and } \overline{Y}_{bvrss} = \frac{\sum_{t.j.i} Y_{(i)[j]t}}{n} \tag{2.3}$$

$$\text{var}\,(\hat{Y}_{(i)[j]t}) = \frac{\sigma_e^2}{n}[1 + E(\frac{(X_{[i](j)t} - \overline{X}_{bvrss})^2}{S^2_{X\,bvRSS}})] + \beta^2\sigma^2_{X[i](j)t} \tag{2.4}$$

Then the fitted model is

$$\overline{Y}_{(i)[j]t} = \hat{a}_{bvrss} + \hat{\beta}_{bvrss}\,\overline{X}_{[i](j)t} \tag{2.5}$$

$$e_{(i)[j]t} = \hat{\in}_{(i)[j]t} = Y_{(i)[j]t} - \hat{Y}_{[i](j)t} \tag{2.6}$$

Also, a consistent unbiased estimator for

$$\sigma_e^2 \; is \; \hat{\sigma}_e^2 = \frac{\sum_{t.j.i} e_{(i)[j]t}^2}{n-\gamma} \tag{2.7}$$

where γ is number of parameters to be estimated in simple regression model

Assuming the conditions of the regression model above, then

$$E(\hat{a}_{bvrss}) = a \; E(\hat{\beta}_{bvrss}) = \beta \tag{2.8}$$

$$Var(\hat{a}_{bvrss}) = \frac{\sigma_e^2}{n}[1 + E(\frac{\overline{X}^2{}_{bvrss}}{S^2_{X,\,bvrss}})] \tag{2.9}$$

where

$$S^2_{X,\,bvrss} = \frac{\sum_{t,j,i}(X_{[i](j)t} - \overline{X}_{bvrss})^2}{n} \tag{2.10}$$

$$Var(\hat{\beta}_{bvrss}) = \frac{\sigma_e^2}{n}[1 + E(\frac{1}{S^2_{X,\,bvrss}})] \tag{2.11}$$

$$E(\hat{Y}_{(i)[j]t} / X_i = x_i) = a + \beta x_{(i)[j]t} \tag{2.12}$$

$$Var(\hat{Y}_{(i)[j]t}) = \frac{\sigma_e^2}{n}[1 + E(\frac{(x_{[i](j)t} - X_{bvrss})^2}{S^2_{X,\,bvrss}})] \qquad + \beta^2 \sigma^2_{X(i)[j]t} \tag{2.13}$$

$$E(e_{(i)[j]t}) = 0 \tag{2.14}$$

$$Var(e_{(i)[j]t}) = \sigma_e^2[1 - [\frac{1}{n} + E(\frac{(x_{(i)[j]t} - \overline{X}_{bvrss})^2}{nS^2_{X,\,bvrss}})] \tag{2.15}$$

$$\mathrm{cov}(\hat{a}_{bvrss}, \hat{\beta}_{bvrss}) = -\frac{\sigma_e^2}{n} E(\frac{(\overline{X}_{bvrss})^2}{\overline{Y}_{bvrss}})] \tag{2.16}$$

$$E(\hat{\sigma}_e^2) = \sigma_e^2 \tag{2.17}$$

$$Var(\hat{\sigma}_e^2) = \frac{2\sigma_e^4}{n-\gamma} \tag{2.18}$$

where γ is number of parameters to be estimated in simple regression model

$$\hat{\sigma}_e^2 \to \sigma_e^2 \tag{2.19}$$

From the above conditions, we can derive the efficiencies of the estimators of α and β using *bvrss* (Bivariate rank set sampling) relative to the estimators using *bvsrs* (Bivariate Simple random sampling) and *bvsys* (Bivariate Systematic sampling) as follows:

$$\text{efficiency } (\hat{a}_{bvrss}, \hat{\beta}_{bvh}) = \left(\frac{[1+E\frac{(\bar{X}^2{}_{bvh})}{S^2_{X,bvh}}]}{[1+E\frac{(\bar{X}^2{}_{bvrss})}{S^2_{X,bvrss}}]}\right) \tag{2.20}$$

Where *bvh = bvsrs, bvsys*

$$\text{and efficiency } (\hat{\beta}_{bvrss}, \hat{\beta}_{bvh}) = \left(\frac{E\frac{1}{S^2_{bvh}}}{E\frac{1}{S^2_{X,bvrss}}}\right) \tag{2.22}$$

where;

$E(Y) = \mu_y, var(Y) = \sigma_y^2, E(X) = \mu_x, var(X) = \sigma_x^2, var(Y) = \sigma_y^2, \rho = cov(Y,X)/\sigma_y\sigma_x, E(X_{[i](j)}) = \mu_{X[i](j)}, E(Y_{(i)[j]}) = \mu_{Y(i)[j]}, E(X^2_{[i](j)}) = \mu^{(2)}_{X[i](j)}, E(Y^2_{(i)[j]}) = \mu^{(2)}_{Y(i)[j]}, var(X_{[i](j)}) = \sigma^{(2)}_{X[i](j)}, var(Y_{(i)[j]}) = \sigma^{(2)}_{Y(i)[j]}, cov(X_{[i](j)}, Y_{(i)[j]} = \sigma_{(X_{[i](j)}, Y_{(i)[j]}}$

3. Numerical illustration

Assume that (X, Y) follow a bivariate normal distribution, the performance of simple regression model using BVSRS, BVSYS and BVRSS was judged with the help of a data set. The original data was collected on two variables of *Pinus Wallichiana*: where, "*X*" is the diameter in centimeters at breast height and "*Y"* is the entire height in feets. The regression model is analyzed assuming that the population consists of *275* trees. The summary statistics of the data is reported in Table-1. A sample size of 55 was fixed in all the sampling designs to make comparisons. Regression analysis and regression diagnostics in all the three sampling designs was carried out in SAS software using the function POC REG. The layout of RSS is given in Table-2. The relative efficiency of RSS with SRS and SYS along with R^2 and Adj R^2 are given the Table-3. The performance of RSS with SRS and SYS is also judged with the help of validation technique i.e. Jack-knifing carried out in SAS using function PROC JACKREG in Tables-5 to 7.

4. Conclusion

It is found that the coefficient of determination obtained from regression model based on Rank set sample is higher than rest of two sampling schemes, also the parameters of comparison like root mean square error, p-value and coefficient of variation is much lower in rank set based regression model than the schemes considered. On using validation technique (Jackknifing) for comparing the regression model based on the considered schemes, it is observed that there is consistency in the measure of R^2, Adj R^2 and RMSE in case of Rank Set Sampling as compared to Simple Random Sampling and

Systematic Sampling. The above results occurred because rank set samples are more regularly spaced than those obtained from Simple Random Sampling and Systematic Sampling and therefore more representative of the population. Because of Ranking the Rank Set Sampling procedure induces stratification at sample level which involves the gained precision in this scheme. Obtaining a sample in this manner maintains the un-biasedness of simple random sampling; however, by incorporating outside information about the sample units, we are able to contribute a structure to the sample that increases its representativeness of the true underlying population. If we quantified the same number of sample units, by a simple random sample or a systematic sample then we have no control over which units enter the sample. Perhaps all the units would come from the lower end of the range, or perhaps most would be clustered at the low end while one or two units would come from the middle or upper range. With other sampling techniques, the only way to increase the prospect of covering the full range of possible values is to increase the sample size. Rank set sampling has a balanced nature in the sense that equal number of observations will be obtained from each rank. It can be easily shown that the sample mean using Rank set sampling has a smaller variances than its counter parts when the number of observations are the same.

Acknowledgment

The first author wishes to record his gratitude and thanks to University Grants Commission, Govt. of India, for providing the UGC National Fellowship for perusing the Doctorate prorgramme in Statistics. Also the authors would like to thank Mr. Irfan Rasool (IFS) DFO, District Kupwara for providing the data for the preparation the manuscript.

References

1. Al-Saleh, M.F. and Zhengh, G. (2002). Estimation of Bivariate characteristics using rank set sampling. *Australian and New Zealand Journal of Statistics.,* 44(1), 221-232.
2. Cobby, J.M., Ridout, M.S., Bassett, P.J. and Large, R.V. (1985). An investigation into the use of ranked set sampling on grass and grass-clover sward. *Grass and Forage Science* 40: 257-263.
3. Cochran, W.G. (1977). Sampling Techniques. *John Wiley and Sons, New York*.
4. Dell, T.R. and Clutter, J.L. 1972. Ranked set sampling theory with order statistics background. *Biometrika*. 28: 545-555.
5. Dell, T.R. and Clutter, V. (1972). Ranked set sampling theory with order statistics background. *Biometrics* 28: 545-555.
6. Gaajendra, K.A. and Bouza, C. (2012). Double sampling with rank set selection in the second phase with non-response: Analytical results and Monte carlo experiments. *Journal of Probability and Statistics* 23: 45-53.
7. Jeelani, M.I., Mir, S.A., Khan, I., and Jeelani, F. (2014). Non-Response Problems in Rank Set Sampling. *Pakistan Journal of Statistics.* 30(4), 555-562.

8. Jeelani, M.I., Mir, S.A., Maqbool, S., Khan, I., Singh, K.N., Zaffer, G., Nazir, N. and Jeelani, F. (2014). Role of Rank Set Sampling in Improving the Estimates of Population Mean under Stratification. *Amer. J. Math. and Statist.*, 4(1), 46-49.

9. Kamarulzaman, I. (2011). On comparison of some variation of rank set sampling. *Sains Malaysiana* 40: 397-401.

10. Martin, W.L., Shank, T.L., Oderwald, R.G. and Smith, D.W. (1980). Evaluation of ranked set sampling for estimating shrub phytomass in Appalachian Oak forest. Technical Report No.FWS-4-80, School of Forestry and Wildlife Resources VPI & SU Blacksburg, VA.

11. McIntyre, G. A. (1952). A method for unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research* 3: 385-390.

12. McIntyre, G. A. (1978). Statistical aspects of vegetation sampling: Measurement of Grassland Vegetation and Animal Production. *Commonwealth Bureau of Pastures and Field Crops*. Hurley, Berkshire, UK. 45: 8-21.

13. Nussbaum, B.D. and Sinha, B.K. (1997). Cost effective gasoline sampling using ranked set sampling. In a Proceedings of the Section on Statistics and the Environment. *American Statistical Association* 41:

14. Quenouille, M.H. 1956. Notes on bias in estimation. *Biometrik*, 43: 353-360.

15. Samawi, H.M and Abu-Dayyeh, W. (1997). Estimating the population mean using bivariate rank set sampling, *Biometrical Journal*., 38 (5), 577-586.

16. You, G. (2009). Efficient designs for sampling and sub sampling in fisheries research based on ranked sets. *Journal of Marine Science*. 66: 928-934.

Table 1: Summary statistics of the Pinus data

	DBH (cm)	***Height (m)***
Mean	21.44	15.66
Standard Deviation	20.95	17.06
Range	216.80	70.87
Minimum	2.20	0.90
Maximum	219	71.77
Count	275	275

Table 2: Layout of RSS based on ranking (dbh 'cm' and height 'feets') simultaneously

Cycles	Set size = k= 5 (N= 275, n = 55, means we have to repeat the process of ranking (m=11) 11 times i.e. 11×5 = 55					
	Tree Number	*1*	*2*	*3*	*4*	*5*
	Height	15.9	22	56.9	**9.6**	24.6
Cycle 1	(dbh)	28.0	26.0	119.0	**16.0**	43.0
	Tree Number	*6*	*7*	*8*	*9*	*10*
	Height	3.3	11.4	**4.7**	21.3	16.8
	(dbh)	7.0	21.0	**6.0**	40.0	28.0
	Tree Number	*11*	*12*	*13*	*14*	*15*
	Height	**5.1**	7.5	3.1	4.9	6.1
	(dbh)	**12.0**	22.0	7.0	7.0	9.0
	Tree Number	*16*	*17*	*18*	*19*	*20*
	Height	5.5	**6.5**	5.6	6.9	3.8
	(dbh)	12.0	**11.0**	14.0	11.0	6.0
	Tree Number	*21*	*22*	*23*	*24*	*25*
	Height	9.7	6.9	4.1	**58.5**	46
	(dbh)	27.0	16.0	8.0	**192.0**	203.0
Cycle 11	Tree Number	*251*	*252*	*253*	*254*	*255*
	Height	10.9	3.5	**2.5**	10.9	8.9
	(dbh)	33.0	6.0	**4.0**	26.0	24.0
	Tree Number	*256*	*257*	*258*	*259*	*260*
	Height	21	44.1	7.0	9.4	**8**
	(dbh)	67.0	107.0	16.0	27.0	**17.0**
	Tree Number	*261*	*262*	*263*	*264*	*265*
	Height	23	11.6	33	7.5	**17.5**
	(dbh)	59	35.0	90.0	17.0	**46.0**
	Tree Number	*266*	*267*	*268*	*269*	*270*
	Height	8.9	47.4	**22**	6.8	7.5
	(dbh)	33.0	53.0	**49.0**	18.0	18.0
	Tree Number	*271*	*272*	*273*	*274*	*275*
	Height	**22.2**	19.3	14.5	3.5	10.9
	(dbh)	**32.0**	25.0	22.0	5.0	26.0

For the sake of simplicity only 1st and 11th cycle is presented here, where figures in italics are tree numbers and figures in bold are dbh 'cm' and height 'feets'

Table 4: Relative efficiency of RSS with SRS and SYS along with R^2, Adj R^2 and others measures of comparison

	RSS		SRS		SYS	
RMSE	8.221		9.521		9.851	
R^2	0.8029		0.771		0.7458	
Adj R^2	0.7991		0.763		0.7268	
ESS	2942.73		3600.26		4804.16	
F value	414.78		215.85		209.21	
CV	35.41		55.61		48.42	
P value	0.0007		0.0034		0.0048	
Dbh (β_1)	1.175		0.935		0.916	
	Lower 95%	**Upper 95%**	**Lower 95%**	**Upper 95%**	**Lower 95%**	**Upper 95%**
Confidence Interval	1.05	1.29	0.64	1.06	0.60	1.08
	0.24		0.42		0.48	
Relative Efficiency	**$RMSE_{srs}/RMSE_{rss}$**		**$RMSE_{sys}/RMSE_{rss}$**			
	1.15		1.19			

Table 5: Comparison of regression models based on various schemes using Jack-knifing

No. of Models	RSS (Rank set sampling)			SRS (Simple random sampling)			Systematic sampling		
	R^2	Adj R^2	RMSE	R^2	Adj R^2	RMSE	R^2	Adj R^2	RMSE
1	0.8021	0.7984	8.231	0.7972	0.7940	9.525	0.7455	0.7267	9.653
2	0.8019	0.7987	8.225	0.7901	0.7632	9.632	0.7135	0.6995	10.432
3	0.8011	0.7982	8.223	0.7992	0.7165	9.743	0.7194	0.6941	10.932
4	0.8015	0.7993	8.228	0.7832	0.7132	10.146	0.7065	0.6932	11.401
5	0.8011	0.7984	8.235	0.7814	0.7115	10.324	0.7034	0.6911	11.567
6	0.8018	0.7984	8.238	0.7801	0.7135	10.365	0.7010	0.6843	11.537
7	0.8022	0.7983	8.242	0.7832	0.7128	10.378	0.7001	0.6872	11.612
8	0.8024	0.7979	8.239	0.7733	0.7132	9.432	0.6942	0.6845	11.105
9	0.8020	0.7973	8.245	0.7632	0.7109	8.475	0.6837	0.6741	11.653
10	0.8023	0.7971	8.228	0.7774	0.7113	11.372	0.6135	0.6735	11.724
11	0.8029	0.7974	8.223	0.7701	0.6995	11.248	0.6347	0.6611	12.001
12	0.8015	0.7992	8.236	0.7732	0.6943	11.371	0.6451	0.6601	12.345
13	0.8013	0.7993	8.230	0.7632	0.6735	9.506	0.6458	0.6538	9.377
14	0.8016	0.7995	8.227	0.7448	0.6525	9.135	0.6743	0.6527	11.433
15	0.8012	0.7998	8.240	0.7452	0.7103	9.346	0.6748	0.6439	12.165
16	0.8015	0.7988	8.245	0.7456	0.6772	9.732	0.6551	0.6425	12.437
17	0.8021	0.7981	8.237	0.7480	0.6785	9.441	0.6449	0.6417	11.523
18	0.8022	0.7987	8.230	0.7495	0.6742	10.532	0.6442	0.6391	12.453
19	0.8023	0.7983	8.228	0.7501	0.6832	9.632	0.6767	0.6382	11.06
20	0.8025	0.7985	8.229	0.7832	0.6945	9.575	0.6743	0.6311	10.501
21	0.8015	0.7982	8.237	0.7827	0.7343	9.321	0.6859	0.6235	10.425
22	0.8014	0.7983	8.231	0.7773	0.7135	10.242	0.6866	0.6211	11.501
23	0.8016	0.7985	8.234	0.7721	0.7247	11.438	0.6977	0.6171	10.501
24	0.8013	0.7987	8.236	0.7732	0.7135	11.586	0.6542	0.6123	10.425
25	0.8015	0.7988	8.228	0.7560	0.6991	10.788	0.6321	0.6118	11.167
26	0.8030	0.7990	8.240	0.7470	0.6432	9.656	0.6354	0.6112	12.432
27	0.8028	0.7991	8.239	0.7321	0.6135	9.842	0.6366	0.6011	11.937
28	0.8029	0.7993	8.244	0.7215	0.7201	9.747	0.6501	0.5991	10.666
29	0.8023	0.7992	8.233	0.7721	0.7115	9.735	0.6554	0.5932	11.732
30	0.8024	0.7993	8.237	0.7232	0.6432	9.645	0.6932	0.5932	12.406
31	0.8012	0.7995	8.243	0.7245	0.6940	9.748	0.6142	0.5995	9.735
32	0.8017	0.7991	8.240	0.7243	0.6532	10.432	0.5995	0.6013	9.532
33	0.8015	0.7992	8.246	0.7165	0.6348	10.458	0.5812	0.6115	9.567
34	0.8030	0.7984	8.231	0.7237	0.6474	9.638	0.5994	0.5942	9.471
35	0.8017	0.7981	8.237	0.7354	0.6903	9.546	0.6011	0.5711	9.450
36	0.8016	0.7980	8.228	0.7380	0.7103	9.532	0.6045	0.5432	9.448
37	0.8027	0.7976	8.243	0.7410	0.6532	9.437	0.6741	0.5511	9.478
38	0.8023	0.7972	8.245	0.7580	0.6671	9.648	0.6849	0.5432	9.501
39	0.8024	0.7973	8.235	0.7595	0.6643	9.632	0.6978	0.5617	9.548
40	0.8020	0.7975	8.238	0.7610	0.6854	9.644	0.7015	0.6348	9.539
41	0.8023	0.7979	8.242	0.7623	0.6711	9.651	0.7135	0.6556	9.511
42	0.8025	0.7983	8.246	0.7651	0.6535	9.628	0.7143	0.6417	10.235
43	0.8021	0.7981	8.248	0.7659	0.7113	9.645	0.7211	0.6521	11.247
44	0.8026	0.7989	8.229	0.6906	0.7211	9.137	0.7312	0.6695	9.373
45	0.8028	0.7986	8.233	0.6972	0.7013	9.635	0.7154	0.6743	9.381
46	0.8019	0.7988	8.238	0.7643	0.6143	9.875	0.7312	0.6528	9.456
47	0.8020	0.7990	8.240	0.7511	0.6051	9.445	0.7221	0.6818	9.556
48	0.8017	0.7995	8.245	0.7451	0.6543	9.436	0.7145	0.6743	9.247
49	0.8015	0.7993	8.240	0.7559	0.6543	9.635	0.7116	0.6713	10.498
50	0.8013	0.7992	8.239	0.6972	0.6136	10.432	0.7234	0.6855	12.071
51	0.8023	0.7990	8.238	0.6907	0.6142	10.548	0.7112	0.6711	11.247
52	0.8024	0.7993	8.240	0.7643	0.6547	10.456	0.7135	0.6943	11.478
53	0.8025	0.7990	8.243	0.7511	0.6567	11.230	0.7149	0.6843	12.562
54	0.8021	0.7982	8.238	0.7432	0.6549	10.629	0.6235	0.6816	12.164

Table 6: Summary statistics of regression model based on sampling designs using Jacknifing

Jackknife samples (54)	RSS (Rank set sampling)			SRS (Simple random sampling)			Systematic sampling		
	R^2	Adj R^2	RMSE	R^2	Adj R^2	RMSE	R^2	Adj R^2	RMSE
Mean	0.8020	0.7986	8.2264	0.7034	0.6832	9.9631	0.6759	0.6418	10.8037
Standard Deviation	0.0005	0.0007	0.0065	0.0262	0.0388	0.6675	0.0417	0.0434	1.0948
Range	0.0019	0.0027	0.0250	0.1086	0.1889	3.1110	0.1643	0.1835	3.3150
Largest	0.8030	0.7998	8.2480	0.7992	0.7940	11.5860	0.7455	0.7267	12.5620
Smallest	0.8011	0.7971	8.2230	0.6906	0.6051	8.4750	0.5812	0.5432	9.2470

*(Each observation is based on 54 jackknife samples)

Table 7: Parameters of comparison of actual regression model

RSS (Rank set sampling)			SRS (Simple random sampling)			Systematic sampling		
R^2	Adj R^2	RMSE	R^2	Adj R^2	RMSE	R^2	Adj R^2	RMSE
0.8029	0.7991	8.221	0.771	0.763	9.521	0.7458	0.7268	9.851