Mysore J. Agric. Sci., 50 (4) : 712-715, 2016

Estimation of Regression Parameters in Bootstrapping - An Application on Non-farm Income

SADASHIVA ACHARI, M. GOPINATH RAO AND P. S. SRIKANTHA MURTHY
Dept. of Agricultural Statistics, Applied Mathematics and Computer Science,
College of Agriculture, UAS, GKVK, Bengaluru - 560 065

ABSTRACT

An attempt has been made to estimate the regression parameters using bootstrap procedure for the response variables non-farm income,investmentmadeand employment in Cauvery command area of Vishweshwaraiah canal region of Mandya district of Karnataka.The explanatory variables such as age, farm income, type of nonfarm activity, seasonality of nonfarm produce and other income sources are observed to be having significant effect on employment generation. Among these the type of nonfarm activity and seasonality of the nonfarm produce are found to be more effective on employment. The income is influenced by seasonality of the nonfarm produce and the reason for taking nonfarm activityas a source of additional income for the house hold. The investment is affected by farm income and seasonality of the non farm produce.

REGRESSION models are very useful and widely used tool. They allow relatively simple analysis of complicated situations, where it is attempted to sort out the effects of many possible explanatory variables on a response variable (Fox, 2008). This regression procedure can also be used in resampling procedures. One of the frequently used resampling techniques is bootstrapping. Bootstrapping (Efron and Tibshirani,1993) is a general approach to statistical inference based on building a sampling distribution for a statistic by resampling from the data on hand. This approach uses computer based methods to provide estimates and measures of precision without theoretical models or restrictive assumptions of the sampled populations. Regression procedure used with this resampling technique is known as bootstrapping regression (Bose and Chatterjee, 2002; Sahinler and Gorgulu, 2003). There are very few studies (Assif and Mathwie, 2010, Berezowski, *et.al.,* 2004; Sahinler and Gorgulu, 2003) on application of this technique in agriculture and hardly any study for the socio-economic data. This paper is about the application of this technique for the socio-economic data (Schroeder, 1992; Smith, 2000; Takeshita, 1999). The study makes use of the data generated on non-farm income, investment and employment generated for Cauvery command area in Karnataka the details of which are given in the next section.

This study focuses on illustration and application of resampling techniques inregression analysis. Some hierarchical algorithms of concerning techniques in regression analysis are demonstrated. The basics of the bootstrap techniques and their applications to the real numerical example that can be described by linearregression model are discussed.

MATERIAL AND METHODS

The secondary data used is from the project on role and contribution of irrigation towards rural nonfarm activity-A case of Cauvery command area in Karnataka. The study area has 418 households having non farm activities and the data consist following socio economic characters from each of the households as the independent variable. They are X_1: Age, X_2: Family type (categorical variable), X_3: Family size, X_4: Number of working adults, X_5: Annual farm income, X_6: Type of nonfarm activity (categorical variable), X_7: Seasonality of the produce (categorical variable), X_8: Otherincome sources (categorical variable), X_9: Additional income (categorical variable), X_{10}: Resource diversification (categorical variable). The dependent variables are Y: Employment generation, Income generation and investment made on nonfarm activity.The statistical packageSAS(Version 9.3 Module Base SAS) is usedfor the statistical analysis of these data.

To describe the resampling methods start with an n sized sample *Wi=(Yi,Xi)* and assume that w_i's are drawn independently and identically, where $Y_i = (y_1, y_2, ..., y_n)'$ contains the responses, $X_{ji} = (x_{j1}, x_{j2}, ..., x_{jn})$ is a matrix of dimension n x k, where j=1,2,...k, i=1,2,3,...,n.

Method of drawing bootstrap samples : Multiple linear regression was fitted to response variable Y and all the explanatory variables considered. 1000 samples each of size 418 are drawn with replacement.

Bootstrapping regression algorithm based on the resampling observations : This approach is usually applied when the regression models built from data have regressors that are as random as the response. Let the (k+1) x1 vector $w_i = (y_i, x_{ji})'$ denote the values associated with i^{th} observation. In this case, the set of observations are the vectors $(w_1, w_2, ..., w_n)$. The bootstrap procedure based on the resampling observations is as follows :

1. Draw a n sized bootstrap sample $(w_1^{(b)}, w_2^{(b)}, \ldots, w_n^{(b)})$ with replacement from the observations giving 1/n probability to each w_i values and label the elements of each vector as follows $w_i^{(b)} = (y_i^{(b)}, x_{ji}^{(b)})'$ where j= 1,2,…,k and i= 1,2,…,n. From these form the vector $Y_i^{(b)} = (y_1^{(b)}, y_2^{(b)}, ..., y_n^{(b)})'$ and the matrix $X_{ji}^{(b)} = (x_{j1}^{(b)}, x_{j2}^{(b)}, ..., x_{jn}^{(b)})'$

2. Calculate the OLS co-efficients from the bootstrap sample: $\hat{\beta}^{(b1)} = (X^{(b)'} X^{(b)})^{-1} X^{(b)'} Y^{(b)}$

3. Repeat the steps 1 and 2 for r = 1, 2,…, B where, B is the number of repetitions.

4. The bootstrap estimate of regression coefficient is the mean of the coefficients and it is given by

$$\hat{\beta}_i^{(b)} = \frac{\sum_{r=1}^{B} \hat{\beta}_i^{(br)}}{B}$$

where, i= 1,2…k

Thus, the bootstrap regression equation is $\hat{Y} = X\hat{\beta}^{(b)}$ where, $\hat{\beta}^{(b)}$ is the unbiased estimator of β(Shao, 1995).

The bootstrap bias, variance, confidence and percentile interval

The bootstrap bias equals,

$$bias_b = \hat{\beta}^{(b)} - \hat{\beta}$$

The bootstrap variance is calculated by

$$\text{var}(\hat{\beta}^{(b)}) = \frac{\sum_{r=1}^{B} [(\hat{\beta}^{(br)} - \hat{\beta}^{(b)})(\hat{\beta}^{(br)} - \hat{\beta}^{(b)})']}{B - 1}$$

where, r = 1,2,...,B

The bootstrap confidence interval by normal approach is obtained by

$$\hat{\beta}^{(b)} - t_{n-p,\alpha/2} * S_e(\beta^{(b)}) < \beta < \hat{\beta}^{(b)} + t_{n-p,\alpha/2} * S_e(\hat{\beta}^{(b)})$$

Where, $t_{n-p,\alpha/2}$ is the critical value of t with probability á/2 the right for n-p degrees of freedom and $S_e(\hat{\beta}^{(b)})$ is the standard error of the $\hat{\beta}^{(b)}$

$$S_e(\hat{\beta}^{(b)}) = \sqrt{\frac{\sum_{r=1}^{B} [(\hat{\beta}^{(br)} - \hat{\beta}^{(b)})(\hat{\beta}^{(br)} - \hat{\beta}^{(b)})']}{B - 1}}$$

Results and Discussion

Summary statistics of bootstrap regression for the entire study area : Table I represents the summary statistics of the bootstrapping regression coefficients for entire study area, where it estimates the effect of 10 explanatory variables with the respective response variables. The age, annual farm income,other income sources are found to be statistically significant at 5 per cent level of significant and type of nonfarm activity, seasonality of nonfarm produce are observed statistically significant at 1 per cent level of significant for the dependent variable employment. When income is taken as dependent variable, the seasonality of nonfarmproduce and additional income are found to be statistically significant at 5 per cent level of significance. Annual farm income and seasonality of nonfarmproduce are found to be statistically significant at 5 per cent level of significance for the response variable investment.

Confidence interval for bootstrapping regression for the entire study area : Table II gives the confidence interval for the coefficients at 95 per cent level of confidence.

Among the ten independent variables considered to find their impact on the response variables such as employment, income and investment some are found to have significant effect, and these are discussed in the following paragraph with respect to the entire command area.

For the entire command area the explanatory variables such as age, farm income, type of nonfarm activity, seasonality of nonfarm produce and other income sources are observed to be having significant effect on employment generation. Among these, the type of nonfarm activity and seasonality of the nonfarm

TABLE I

Summary statistics of bootstrap regression for the entire study area

(n=418, B=1000)

Coefficients	Dependent variables		
	Employment	Income	Investment
β_0	0.822	-151.9 *	-96.51
	(1.36)	(69.84)	(89.0)
β_1	-0.036 *	1.692	0.950
	(0.014)	(0.876)	(1.045)
β_2	-0.038	-1.255	10.393
	(0.571)	(31.82)	(48.97)
β_3	0.266	-0.170	2.365
	(0.161)	(4.35)	(6.973)
β_4	0.056	10.375	-13.93
	(0.208)	(11.21)	(11.66)
β_5	0.011 *	0.577	2.084 *
	(0.004)	(0.343)	(1.011)
β_6	0.508 **	11.599	-0.601
	(0.119)	(5.955)	(8.499)
β_7	-0.28 **	2.923 *	7.659 *
	(0.08)	(1.436)	(2.966)
β_8	-0.909 *	-31.27	-17.33
	(0.391)	(18.87)	(24.32)
β_9	0.219	29.47 *	23.477
	(0.505)	(14.44)	(26.72)
β_{10}	0.542	36.755	50.920
	(0.408)	(20.15)	(53.52)

Note: *Significant at 5 per cent, **Significant at 1 per cent, n represents sample size; B represents number of bootstrap replications. Values indicated inside the parentheses represent standard error.

TABLE II

Confidence interval for bootstrapping regression for the entire study area

(n=418, B=1000)

C. I. at 95% level of confidence						
Co-efficients	Dependent variables					
	Employment		Income		Investment	
	Lower	Upper	Lower	Upper	Lower	Upper
β_0	-1.852	3.495	-289.2	-14.68	-271.5	78.441
β_1	-0.008	0.065	-0.03	3.413	-1.103	3.003
β_2	-1.162	1.085	-63.81	61.304	-85.88	106.66
β_3	-0.051	0.583	-8.722	8.381	-11.34	16.07
β_4	-0.353	0.466	-11.66	32.416	-36.86	8.989
β_5	0.003	0.019	-0.096	1.251	0.096	4.072
β_6	0.274	0.741	-0.106	23.304	-17.3	16.105
β_7	-0.443	-0.127	0.1	5.747	1.828	13.49
β_8	-1.678	-0.14	-68.38	5.831	-65.15	30.476
β_9	-0.774	1.213	1.074	57.873	-29.05	76.007
β_{10}	-0.26	1.345	-2.867	76.378	-54.29	156.13

Note: C. I. = indicate confidence interval
n = represents sample size
β = represents number of bootstrap replications

produce are found to be more effective on employment. The income is influenced by seasonality of the nonfarm produce and the reason for taking non farm activityas a source of additional income for the house hold. The investment is affected by farm income and seasonality of the non farm produce.

Investment on non-farm activities is influenced significantly by farm income and seasonality of the non farm produce. Income is significantly affected by other income sources and resource diversification. Employment is influenced very significantlyby seasonality of nonfarm produce, other income sources and significantly bytype of nonfarm activity, additional income.

REFERENCES

ASSIF, A. AND MATHWIE, K. M., 2010, Improving the accuracy of DEA efficiency analysis, A bootstrap application to the health care foodservice industry. *Appl. Econ.*, **42** (25/27) : 3547 - 3558.

BEREZOWSKI, J., RENTER, D. AND EVANS, R., 2004, A Bootstrapping method for diagnostic test evaluation and prevalence estimation. *Society for veterinary epidemiology and preventive medicine proceedings of a meeting* held at the Martigny, Switzerland. pp. 21 - 31.

BOSE, A. AND CHATTERJEE, S., 2002, Comparison of bootstrap and jacknife variance estimators in linear regression second order results. *Stat. Sinica.,* **12** : 575 - 598.

EFRON, B AND TIBSHIRANI, B. J., 1993, *An Introduction to the Bootstrap*. (eds: Chapman and hall). Washington, United States of America.

Fox, J., 2008, Applied Regression, Generalised Linear Models and Related Methods, 2nd edition. Thousand Oaks California: Sage Publications

SAHINLER, S. AND GORGULU, M.,2003, Estimation of main carcass components by using bootstrapping regression method. *J. Anim. Feed Sci.*, **12**(4) : 723 - 737.

SCHROEDER, T. C., 1992, Economies of scale and scope for Agricultural supply and marketing cooperatives. *Rev. Agric. Econ.*, **14**(1) : 93 - 103.

SHAO, J., 1995, Bootstrap model selection, *J. Amer. Statist. Assoc.*, pp. 655-665.

SMITH, D., 2000, The spatial dimension of access to the rural non-farm economy. *NRI Draft paper*. Chatham, United Kingdom.

TAKESHITA, H., 1999, Econometric analysis of health information impact on food consumption. *J. Rur. Econ.*, **71**(2) : 61 - 70.

(*Received :* August., 2016 *Accepted :* November, 2016)