



Restricted and Unrestricted Methods of Bootstrap Data Generating Processes

Acha Chigozie K.^{*}, Nwabueze Joy C.

Department of Statistics, Michael Okpara University of Agriculture, Umudike, Nigeria

Email address:

specialgozie@yahoo.com (Acha C. K.)

^{*}Corresponding author

To cite this article:

Acha Chigozie K., Nwabueze Joy C. Restricted and Unrestricted Methods of Bootstrap Data Generating Processes. *International Journal of Theoretical and Applied Mathematics*. Vol. 2, No. 2, 2016, pp. 121-126. doi: 10.11648/j.ijtam.20160202.24

Received: October 30, 2016; **Accepted:** November 17, 2016; **Published:** December 21, 2016

Abstract: This study compares the restricted and unrestricted methods of bootstrap data generating processes (DGPs) on statistical inference. It used hypothetical datasets simulated from normal distribution with different ability levels. Data were analyzed using different bootstrap DGPs. In practice, it is advisable to use the restricted parametric bootstrap DGP models and thereafter, check the kernel density of the empirical distributions that are close to normal (at least not too skewed). In fact, 21600 scenarios were replicated 200 times using bootstrap DGPs and kernel density methods. This analysis was carried out using R-statistical package. The results show that in a situation where the distribution of a test is skewed, all the scores need to be taken into account, no matter how small the sample size and the bootstrap level are. Across all the conditions considered, models HR5UR and HPN5UR yielded much larger bias and standard error while the smallest bias values were associated with models HR5R (0.0619) and HPN5R (0.0624). The result confirms the fact that bootstrap DGPs are very vital in statistical inference.

Keywords: Restricted, Bootstrap DGPs, Simulation, Unrestricted, Functional Model

1. Introduction

The recent increase in computer performance has made it possible to base many statistical inferences on simulated or bootstrapped distributed rather than on distributions obtained from asymptotic theory. Even though, bootstrap is not simulation, but in practice only trivial cases of bootstrap do not require simulation. In this paper a model is said to be unrestricted when it is not transformed otherwise it is a restricted model. This paper aims at examining the impact of residual and parametric bootstrap DGP procedure in terms of their ability level, bootstrap level, sample size, standard errors and bias on the statistical inference. The objective would be achieved by analytically examining the theorized relationships to see if they hold in the simulated datasets from the normal distribution. To achieve this objective which this paper has set for itself, the next section theoretically reviews bootstrap, the third section describes the method to be adopted in data analysis. In the fourth section data is analyzed using bootstrap DGPs and kernel density methods added by the R-statistical package. The paper is concluded in

the fifth section. Finally, references and appendix were included.

2. Theoretical Overview

The continuing development of bootstrap methods has been motivated by the increasing progress in computational speed and efficiency. According to [1], [2], [3], the major aim of bootstrap testing is the characterization of a test statistic of interest with an unknown distribution under the null hypothesis using information in the data set that is being analyzed. [4], viewed bootstrap as a technique for estimating standard errors. His idea was to use simulation, based on a nonparametric estimate of the underlying error distribution. The model was fitted by three-stage least squares and applied to an econometric model describing the demand for capital, labor, energy, and materials. He concluded that the coefficient estimates and the estimated standard errors performed very well.

According to [5], [2], there are many bootstrap methods that can be used for econometric analysis. In certain circumstances, such as regression models with independent

and identically distributed error terms, appropriately chosen bootstrap methods generally work very well. A large number of bootstrap methods are useful in econometrics. Applications to bias, standard error and hypothesis testing are emphasized, and simulation results are presented for many illustrative cases as included in [6-20], [2], [21], [22].

Consider

$$y_t = X_t \beta + \mu_t; E(\mu_t / X_t) = 0, E(\mu_s \mu_t) = 0 \forall s \neq t, \mu_t \sim NID(0; \sigma^2) \quad (1)$$

The corresponding dependent variables from the bootstrap methods are given by;

$$y_b = X \beta^* + \mu_t \quad (2)$$

where the dependent variable, y_t is a linear combination of the parameters, n is the number of observations, β^* is the bootstrap k -vector, β is a k -vector, and the $1 \times k$ vector of regressors X_t , treated as fixed and μ is an $n \times 1$ vector of independent identically distributed errors with mean 0 and variance σ^2 . For each vector y_b the estimator is recomputed and the sampling distribution of the estimator is estimated by the assumed distribution and empirical distribution respectively.

There are many ways to specify bootstrap data generating processes (DGPs) for the model (2). According to [23], some require very strong assumptions about the error terms μ_t , whereas others require much weaker ones. Two types of bootstrap DGP for regression models (2) when the data is independently and identically distributed (iid) are;

(i) The residual bootstrap

$$y_t^* = X_t \hat{\beta} + \mu_t^*, \mu_t^* \sim NID(\hat{\mu}_t) \quad (3)$$

(ii) The parametric bootstrap DGP

$$y_t^* = X_t \hat{\beta} + \mu_t^*, \mu_t^* \sim NID(0, s^2) \quad (4)$$

Here it is assumed that the errors are normally distributed, the usual ordinary least square (OLS) estimate of the error variance for the residual bootstrap DGP. Similar methods can be used with any model estimated by maximum likelihood (for 4), but their validity generally depends on the strong assumptions inherent in maximum likelihood estimation, thereafter, the transformations of the models. There are also frequency domain approaches to bootstrapping econometrics, regression, time series, design of experiments, etc. [6], [24], [25], [26], and [27]. Other author that contributed immensely in bootstrap are [28], [1], [13], [29] and [30]; who treated computing leaf rectangularity index: theory and applications, parametric bootstrap methods for parameter estimation in SLR models, bootstrapping normal and binomial distributions, bootstrap confidence regions for multidimensional scaling solution and selection of A-B procedure for bootstrap level respectively.

3. Research Methodology

Secondary data sets are analyzed using bootstrap DGPs and kernel density methods. This analysis was carried out by

the R-statistical package with several assessment conditions. They bias and standard error are computed as shown below;

3.1. Bias in Bootstrapped Regression Models

The algorithm for estimating standard errors from regression models as suggested by [7].

The bias of $\hat{\theta}(b) = s(x)$ as an estimate of θ is defined to be difference between the expectation of $\hat{\theta}$ and the value of the parameter θ .

$$\text{bias}_F = \text{bias}_F(\hat{\theta}, \theta) = E_F[s(X)] - \theta \quad (F)$$

where;

$$\hat{\theta} = s(X) \text{ and } \theta = t(F) \quad (5)$$

3.2. Standard Error in Bootstrapped Regression Models

Bootstrap can be apply to more general regression models and save time from taking many samples from the population to make statistical inference. [7], suggested bootstrap algorithm for estimating standard errors from regression models as shown below;

a. Select B independent bootstrap samples $x^{*1}, x^{*2}, \dots, x^{*B}$ each consisting of n data values drawn with replacement from x , for estimating a standard error, the number B will ordinarily be in the range 25–200.

b. Evaluate the bootstrap replication corresponding to each bootstrap sample,

$$\hat{\theta}^*(b) = s(x^{*b}) \quad b=1, 2, \dots, B. \quad (6)$$

c. The bootstrap estimate of standard error is the standard deviation of the bootstrap (B) replications:

$$\widehat{se}_{boot} = \left\{ \sum_{b=1}^B [s(x^{*b}) - s(\cdot)]^2 / (B - 1) \right\}^{1/2} \quad (7)$$

where,

$$s(\cdot) = \sum_{b=1}^B s(x^{*b}) / B.$$

4. Data Analysis and Interpretation of the Result

4.1. Data Analysis

Each of the forms of the bootstrap methods were represented by using at least one functional model each from hypothetical data sets of a particular bootstrap DGP method to illustrate how others were estimated before tabulation;

Using (1) to estimate original hypothetical data sets with fixed sample size is as follows;

A. Original Hypothetical Model (H5):

$$\text{HYP}t = b_0 + b_1 A + b_2 B + \mu_t \quad (8)$$

Original Hypothetical Model (H5), $B=1999$, $N(1, 0.25)$, $n_1=1000$

$$\text{HYP}t = 34.14231687 b_1 + 0.05696246 b_2 \quad (9)$$

Standard error (0.00494) (0.02791)
 Bias (0.07031) (0.0333)
 RMSE (0.00221)

B. The unrestricted residual bootstrap DGP;

$$y_t^* = X_t \hat{\beta} + \mu_t^*, \mu_t^* \sim NID(\hat{\mu}_t) \tag{10}$$

The unrestricted residual bootstrap DGP when applied on the hypothetical data sets with fixed sample size is as follows;

Hypothetical Model (HR5UR):

$$HYPt = b_0 + b_1 A + b_2 B + \mu_t \tag{11}$$

Under several assessment conditions; B=99, N (0, 1), n₁=1000 we have:

$$HYPt = 34.1342316 b_1 + 0.06562462 b_2 \tag{12}$$

Standard error (0.00500) (0.05180)
 Bias (0.08755) (0.03280)
 RMSE (0.00243)

C. The restricted (transformed) residual bootstrap DGP using the battle transformations in section 2;

$$y_t^* = Z_t \hat{\beta} + \mu_t^*, \mu_t^* \sim NID(\hat{\mu}_t) \tag{13}$$

where Z_t represents the transformation in the equation.

Transformed Residual Model (HR5R),

$$HYPt = b_0 + b_1 A + b_2 B + \mu_t \tag{14}$$

Under several assessment conditions; B=99, N (0, 1), n₁=1000 we have:

$$HYPt = 34.12231687 b_1 + 0.05696246 b_2 \tag{15}$$

Standard error (0.00486) (0.04852)
 Bias (0.0719) (0.03231)
 RMSE (0.0019)

D. The unrestricted Parametric bootstrap DGP with nuisance parameter

$$y_t^* = X_t \hat{\beta} + \mu_t^*, \mu_t^* \sim NID(0, s^2) \tag{16}$$

Hypothetical Model (HPN5UR):

$$HYPt = b_0 + b_1 A + b_2 B + \mu_t \tag{17}$$

Under several assessment conditions; B=499, N (0, 1), n₁=1000 we have:

$$HYPt = 34.24231667 b_1 + 0.05356962 b_2 \tag{18}$$

Standard error (0.00392) (0.04302)
 Bias (0.0800) (0.033609)
 RMSE (0.0030)

E. The restricted (Transformed) Parametric bootstrap DGP with nuisance parameter

$$y_t^* = Z_t \hat{\beta} + \mu_t^*, \mu_t^* \sim NID(0, s^2)$$

where Z_t represents the transformation in the equation. (19)

Hypothetical Model (HPN5UR):

$$HYPt = b_0 + b_1 A + b_2 B + \mu_t \tag{20}$$

Under several assessment conditions; B=499, N (0, 1), n₁=1000 we have:

$$HYPt = 34.21131652 b_1 + 0.05126522 b_2 \tag{21}$$

Standard error (0.00478) (0.02030)
 Bias (0.0624) (0.0107)
 RMSE (0.0021)

It is pertinent to note that equations (9), (12), (15), (18), and (21) represent the original SLR hypothetical data set before restrictions and different bootstrap DGPs were applied.

4.2. Interpretation of the Result

In this study, five groups of models were selected to represent the hypothetical data sets on restricted and unrestricted bootstrap DGPs. Thereafter more than 200 trials were carried out within each bootstrap level (B). The selection was based on the fact that as n (number of trials) increase, the models maintain the same pattern, and unless there is change in the pattern another model will not be selected. The five equations above represent each of the groups of models selected; results presented in table (H5, through HPN5UR). It is also important to note that in this study the logarithm of the data sets was used among the three (3) transformations because it came out with minimum bias and standard error.

This will enable determine the present the effects of the factors of sample size and bootstrap level on a hypothetical data sets. Extreme values in the ranges stated above were truncated and very low estimates were also observed, results in these ranges are presented in order to demonstrate the trends and the performance at the lower ends of the distributions for each bootstrap model. The bootstrap models when bootstrap DGP models with Uncorrelated Error Term from the forms: N (0, 1), N (0, 0.9), and N (1, 0.25) distributions.

The conditional bias for the bootstrap models from hypothetical data sets was considered, in fact, only the correlation between original values and restricted (transformed) values. Although the magnitude of bias varied across the bootstrap models, the pattern of relative effects of these factors was generally consistent within each bootstrap model. It can be seen that sample size and bootstrap level had large effects on bias of the SLR, group proficiency level had relatively small effects under some conditions. Estimation bias decreases as the sample size and bootstrap levels increases.

Across all the conditions considered, models HR5UR and HPN5UR yielded much larger bias and standard error while the smallest bias, standard error and RMSE values were associated with models HR5R and HPN5R when compared to H5 functional model. Therefore, for the bootstrap models, the pattern was clear that lower sample sizes and unrestricted models were associated with larger bias while higher sample sizes, restricted (transformed) were related to lower bias. This is not surprising, because the fitted distribution with the

higher sample sizes even when bootstrapped was more similar to the distribution of the original data. For the three different sample sizes and the three different ability levels, the largest (bias and standard error) estimate were always associated with model HPN5UR while the smallest bias was from model HR5R. They model HR5R can also be used at that stage to predict and forecast.

5. Major Findings and Conclusion

Among all the conditions considered, models HR5UR and HPN5UR yielded much larger bias and standard error than the other models at almost all the estimates. It was also

observed from the results that the parametric bootstrap functional DGP models when transformed with a higher sample size and higher bootstrap level generally yielded smaller total errors in estimating the standard error and bias. Though, parametric bootstrap DGP (residual and parametric) models were similar in standard error estimates across most of the estimated values, especially when the sample size was equal to or larger than 200. The study therefore concludes that the pattern (restricted and unrestricted) parametric bootstrap models was clear, that is, the lower sample sizes and unrestricted models were associated with larger bias and standard error, vice versa.

Appendix

Table A1. Bias of SLR across all Models in a Hypothetical data set.

Bootstrap level	Ability Level	Sample Size	H5	HR5UR	HR5R	HPN5UR	HPN5R
B=99	N (0, 1)	200	0.0390	0.1051	0.0448	0.0301	0.0249
		1000	0.0209	0.0547	0.0100	0.0124	0.0110
		10000	0.0312	0.0115	0.0029	0.0071	0.0057
	N (0, 0.9)	200	0.0341	0.1048	0.0429	0.0283	0.0238
		1000	0.0557	0.0510	0.0071	0.0530	0.0240
		10000	0.0319	0.0296	0.0028	0.0305	0.0124
	N (1, 0.25)	200	0.0207	0.1071	0.1013	0.1037	0.0242
		1000	0.0551	0.0514	0.0533	0.0098	0.0056
		10000	0.0321	0.0297	0.0304	0.0065	0.0028
B=499	N (0, 1)	200	0.0419	0.0901	0.0795	0.0450	0.0237
		1000	0.0463	0.0221	0.0103	0.0217	0.0106
		10000	0.0271	0.0238	0.0031	0.0129	0.0066
	N (0, 0.9)	200	0.0331	0.0880	0.0797	0.0451	0.0246
		1000	0.0450	0.0238	0.0067	0.0218	0.0097
		10000	0.0266	0.0236	0.0030	0.0079	0.0137
	N (1, 0.25)	200	0.0227	0.0878	0.0831	0.0477	0.0268
		1000	0.0451	0.0416	0.0060	0.0240	0.0099
		10000	0.0266	0.0239	0.0244	0.0062	0.0030
B=1999	N (0, 1)	200	0.0200	0.0548	0.0184	0.0122	0.0067
		1000	0.0041	0.0429	0.0036	0.0096	0.0087
		10000	0.0014	0.0231	0.0056	0.0031	0.0020
	N (0, 0.9)	200	0.0592	0.0130	0.0165	0.0174	0.0066
		1000	0.0317	0.0062	0.0093	0.0094	0.0036
		10000	0.0016	0.0154	0.0054	0.0055	0.0022
	N (0, 1)	200	0.0587	0.0533	0.0182	0.0168	0.0069
		1000	0.0427	0.0318	0.0028	0.0055	0.0040
		10000	0.0241	0.0174	0.0011	0.0029	0.0054

Note. The bold is the smallest value in each row. The number 5 in the model name represents the 5 functional models in this study.

Table A2. Standard Error of SLR across all Models in a hypothetical data set.

Bootstrap Level	Ability Level	Sample Size	H5	HR5UR	HR5R	HPN5UR	HPN5R
B=99	N (0, 1)	200	0.1184	0.0340	0.0479	0.0567	0.0571
		1000	0.0407	0.0115	0.0168	0.0167	0.0115
		10000	0.0150	0.0058	0.0058	0.0064	0.0052
	N (0, 0.9)	200	0.1197	0.0339	0.0475	0.0471	0.0522
		1000	0.0111	0.0146	0.0142	0.0169	0.0173
		10000	0.0056	0.0061	0.0065	0.0066	0.0050
	N (1, 0.25)	200	0.1188	0.0343	0.0398	0.0485	0.0485
		1000	0.0425	0.0162	0.0174	0.0172	0.0110
		10000	0.0170	0.0050	0.0049	0.0065	0.0067
B=499	N (0, 1)	200	0.2403	0.0540	0.0627	0.1136	0.1228
		1000	0.0890	0.0185	0.0283	0.0337	0.0372
		10000	0.0347	0.0089	0.0104	0.0112	0.0081
N (0, 0.9)	200	0.2350	0.0955	0.1037	0.1107	0.0537	

Bootstrap Level	Ability Level	Sample Size	H5	HR5UR	HR5R	HPN5UR	HPN5R
B=1999	N (1, 0.25)	1000	0.0869	0.0186	0.0194	0.0214	0.0294
		10000	0.0344	0.0080	0.0083	0.0088	0.0130
		200	0.2252	0.0543	0.1013	0.0982	0.1104
		1000	0.0858	0.0186	0.0286	0.0331	0.0375
		10000	0.0343	0.0080	0.0105	0.0120	0.0133
		200	0.2360	0.0641	0.1201	0.1265	0.0556
	N (0, 1)	1000	0.0893	0.0299	0.0320	0.0317	0.0194
		10000	0.0360	0.0115	0.0133	0.0135	0.0088
		200	0.2313	0.0968	0.1022	0.1037	0.0560
		1000	0.0196	0.0231	0.0301	0.0363	0.0378
		10000	0.0352	0.0087	0.0114	0.0120	0.0116
		200	0.0570	0.0692	0.0956	0.1184	0.1265
	N (1, 0.25)	1000	0.0873	0.0196	0.0307	0.0372	0.0399
		10000	0.0349	0.0088	0.0115	0.0127	0.0118

Note. The bold is the smallest value in each row. The number 5 in the model name represents the 5 functional models in this study.

References

[1] Acha, C. K. (2014a), Parametric Bootstrap Methods for Parameter Estimation in SLR Models, “*International Journal of Econometrics and Financial Management*”. 2 (5). Doi: 10.12691/ijefm-2-5-2.

[2] Acha, C. K. (2016), Rescaling Residual Bootstrap and Wild Bootstrap. *International Journal of Data Science and Analysis*. 2 (1): 7-14. doi: 10.11648/j.ijdsa.20160201.12.

[3] Acha, C. K. and Acha I. A. (2015) Smooth Bootstrap Methods on External Sector Statistics. *International Journal of Econometrics and Financial Management*, 3 (3), 115–120. Doi: 10.12691/ijefm-3-3-2.

[4] Freedman, D. A., and Peters, S. C. (1984). Bootstrapping a regression equation: Some empirical results. *J. Am. Statist. Assoc.* 79, 97–106.

[5] MacKinnon J. G. (2002), Bootstrap inference in econometrics, *Canadian Journal of Economics Revue Canadienne d'Economique*, 35 (4): 615-645. On-line: 10.1111/0008-4085.00147.

[6] Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. CBMS-NSF Regional.

[7] Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap (Monographs on Statistics and Applied Probability 57). New York: Chapman & Hall.

[8] Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and Their Application*, Cambridge, Cambridge University Press.

[9] Acha, I. A. and Acha, C. K. (2011). Interest Rates in Nigeria: An Analytical Perspective. *Research Journal of Finance and Accounting*, 2 (3); 71-81 www.iiste.org ISSN 2222-1697 (Paper) ISSN 2222-2847 (Online).

[10] Chernick, M. R. (1999). Bootstrap confidence intervals for Cpk : An application to catheter lesion data. In *ASA Proceedings of the Section on Quality and Productivity*, pp. 46–49. American Statistical Association, Alexandria, VA.

[11] Chernick, M. R. (2007). *Bootstrap Methods: A Guide for Practitioners and Researchers*, 2nd Edition Wiley, Hoboken.

[12] Chernick, M. R. and LaBudde, R. (2010). Revisiting qualms about bootstrap confidence intervals. *Am. J. Math. Manag. Sci. Am. J. Math. Manag. Sci.* 29, 437-456.

[13] Acha, C. K. (2014b), Bootstrapping Normal and Binomial Distributions, “*International Journal of Econometrics and Financial Management*”. 2014, 2 (6). Doi: 10.12691/ijefm-2-6-5 Acha, C. K. (2012) Regression and Principal Component Analyses: A Comparison Using Few Regressors. *American Journal of Mathematics and Statistics*, 2 (1), 1–5.

[14] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.

[15] Manly, B. F. J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 2nd Edition, Chapman and Hall, London.

[16] Acha, C. K. (2012) Regression and Principal Component Analyses: A Comparison Using Few Regressors. *American Journal of Mathematics and Statistics*, 2 (1), 1–5.

[17] Lunneborg, C. E. (2000). *Data Analysis by Resampling: Concepts and Applications*. Brooks/Cole, Pacific Grove.

[18] Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. Springer-Verlag, Berlin.

[19] Acha, C. K. and Omekara, C. O. (2016) Towards Efficiency in the Residual and Parametric Bootstrap Techniques. *American Journal of Theoretical and Applied Statistics*. 5 (5) 285-289. doi: 10.11648/j.ajtas.20160505.16.

[20] Shao, J., and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York.

[21] Good, P. (1998). *Resampling Methods: A Practical Guide to Data Analysis*. Birkhauser, Boston.

[22] Good, P. (2004). *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. 3rd Edition. Springer-Verlag, New York.

[23] Davidson, R. and MacKinnon, J. G., (2006). The power of bootstrap and asymptotic tests, *Journal of Econometrics*, Elsevier, 133 (2), 421-441.

[24] Chernick, M. R. and LaBudde, R. (2011). *An Introduction to the Bootstrap with Applications in R* Wiley, Hoboken.

[25] Stefan Van Aelst, Gert Willems (2012) Fast and Robust Bootstrap for Multivariate Inference: The R Package FRB53 (3); 57-62.

- [26] Mackinnon (2001), improving the reliability of bootstrap tests, econometric theory, Taylor and Francis Journals, 19 (1), 55-68.
- [27] Essi (2011) Computing Leaf Regularity Index: Theory and Applications, American Journal of Scientific and Industrial Research, 2 (4), 573-576.
- [28] William G. J. and David A. A. (2014), Bootstrap Confidence Regions for Multidimensional Scaling Solutions. American Journal of Political Science, 58 (1); 264–278.
- [29] Andrews, D. W. K. and Buchinsky, M. (1998). “On the number of bootstrap repetitions for bootstrap standard errors, confidence intervals, confidence regions, and tests,” Cowles Foundation Discussion Paper No. 1141R, Yale University, revised.