Bootstrapping Normal Distribution with Different Group Proficiency Forms

Acha Chigozie Kelechi[1*]

[1] Department of Statistics, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria

Email Address

specialgozie@yahoo.com (Acha Chigozie Kelechi)
*Correspondence: specialgozie@yahoo.com

Received: 30 November 2017**; Accepted:** 22 December 2017

Abstract:

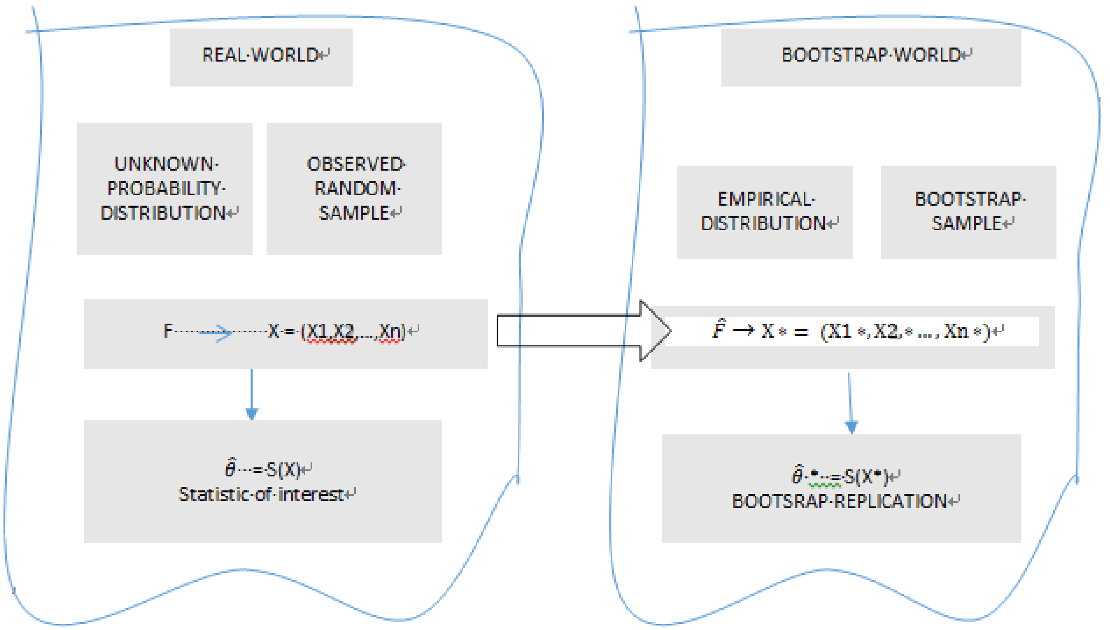
The objective of this paper is to ascertain which group proficiency form (GPF-models) provide more reliable statistical inference by applying nonparametric bootstrap on a normal distribution and define the performance of the GPF-models estimator for forecasting and predicting. The group proficiency forms (ability levels) are very important in a identifying a more reliable statistical model(s) for forecasting and predicting. This objective was achieved by considering seventeen assessment conditions which includes group proficiency forms-upper and lower quartiles of the variance, test lengths, bootstrap levels, sample sizes and the root mean square error. The result as indicated in Table 1 and 2, GPF4 model was associated with the smallest RMSE; 0.0000 in Test 1, followed by GPF6, GPF1, GPF2, GPF5 & GPF3, GPF7 and 0.0001, 0.0002, 0.0004, 0.0009, 0.0017 & 0.0067 respectively under the same assessment conditions. Moreover, the smallest RMSE in all the models is GPF4; 0.0000 while GPF7; 0.0996 gives the largest RMSE in Test 2.GPF4 and GPF7 produced the same RMSE value (0.0130) at B=99, N(0,0.25), n=50000 assessment conditions. The results of the nonparametric bootstrap RMSE with respect to their group proficiency forms showed that the GPF4 model provides more reliable statistical inference among all other models and can be used for forecasting and predicting, though, the effects of the group proficiency forms were quite small compared to sample sizes and bootstrap levels.

Keywords:

Root Mean Square Error, Prediction, Nonparametric Bootstrap, Group Proficiency forms

1. Introduction

The bootstrap estimators of bias, standard error and other estimators are easily derived from the empirical distribution F. Resampling improves the information in the original sample. Thus, the advantage of resampling methods like the bootstrap must be the result of the way the sample information is processed. Many studies have worked extensively on bootstrap and its prospects, Acha (2012a), Cover and Thomas (2006), Efron and Tibshirani (1993), Chernick, (1999), Acha and Acha (2015), Chernick, and LaBudde (2011), Acha and Omekara (2016), Freedman (1981), Lehikoinen, et al (2010), Acha and Acha (2011), Chatfield (2004) and MacKay (2003), Acha (2014a). For instance, in the case of samples from a normal distribution, all the information about the distribution of the sample mean is summarized in the sample mean, bias, and variance (standard error for samples), root mean square (RMSE) which are jointly sufficient statistics (Acha, 2014b). In most econometric applications, where there is no readily available finite sample distribution of the test statistics that's' when one gets the most mileage out of the bootstrap methods, though, it is computationally more demanding than other sampling technique.



Source: Efron and Tibshirani (1993)

Figure 1. *A flow chart showing real and bootstrap worlds.*

According to Cover and Thomas (2006) and Park and Bera (2009), the normal distribution is a very commonly occurring continuous probability distribution. It is also function that tells the probability that any real observation will fall between any two real limits or real numbers. Normal distributions are extremely important in collecting, analyzing and organizing in statistics and are often used in other sciences' real-valued random variables whose distributions are not known. Acha (2010) laid down the normal distribution assumptions and how to test them. She also emphasis that normal distribution is useful because of the central limit theorem, that is, become normally distributed when the number of random variables is sufficiently large. However, many other distributions are bell-shaped.

A normal distribution is;

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

where parameter μ in this definition is the mean. The parameter σ is its standard deviation; and σ^2 is variance.

The group proficiency forms from the normal distribution help to select the best statistical inference for a model that will enable further research. Acha and Acha (2011), lay emphasis on the importance of statistical inference in any research work. Statistical inference is the process of drawing conclusions about population or other collections of objects about which we have only partial knowledge from the samples. It is often beneficial to represent the distribution of random variables using a convenient approximation, facilitating mathematical and statistical analysis. Acha (2012b), point out that a major criterion in judging a certain distribution is the use of a few parameters in defining the distribution. This was done by estimating the coefficients of the explanatory variables in the analysis. It is not always easy to choose the model solely on theoretical and biological grounds. However, knowing the of nature of the data, for example by plotting, can change the beliefs in what model may be the most appropriate and what modeling decision to make.

According to Acha (2014b; 2015), bias and standard error together define the performance of an estimator, and the root mean square error (RMSE) of the SLR. As the bootstrap method is a resampling method for the purpose of reducing error and providing more reliable statistical inference.

In the bootstrap (parametric, semi-parametric and nonparametric) methods, for regression models; if the null and alternative hypotheses are regression models, the simplest approach is to estimate the model that corresponds to the null hypothesis and then use the estimates to generate the bootstrap samples, under the assumption that the error terms are normally distributed.

$$y_t = X_t \beta + u_t, \qquad \sim \mathrm{NID}(0, \sigma^2) \tag{2}$$

The model (2) is a fully specified parametric model, which means that each set of parameter values for β, σ^2 defines just one model. The simplest type of bootstrap method for fully specified models is given by the parametric bootstrap. The first step in constructing a parametric bootstrap model is to estimate (2) by OLS, yielding the restricted estimates $\tilde{\beta}$, and $\tilde{s}^2$. Then the bootstrap DGP is given by

$$y_t^* = X_t\tilde{\beta} + \mu_t^*, \qquad \mu_t^* \sim \mathrm{NID}(0, \tilde{s}^2) \tag{3}$$

which is just the element of the model (2) characterized by the parameter estimates under the null, with stars to indicate that the data are simulated. In order to draw a bootstrap sample from the bootstrap DGP (3), we first draw an n--vector u^* from the N(0, , $\tilde{s}^2$I) distribution. The rest of the procedure for computing a bootstrap P value is identical to the one for computing a bootstrapped P value for exact test. For each of the B bootstrap samples, θ_k^* , a bootstrap test statistic θ_k^* is computed from y_k^* in just the same way as $\hat{\theta}$ was computed from the original data, y.

The parametric bootstrap procedure that we have just described, based on the functional model (3), does not allow us to relax the strong assumption that the error terms are normally distributed. How can we construct a satisfactory bootstrap model if we extend the models (3) to admit non-normal errors? If we knew the true error distribution, whether or not it was normal, we could always generate the $\mu*$ from it. Since we do not know it, we will have to find some way to estimate this distribution. Under the null hypothesis, the OLS residual vector $\tilde{\mu}$ for the restricted model is a consistent estimator of the error vector u. According to Davidson and MacKinnon (2003) is an immediate consequence of the consistency of the OLS estimator itself. From the Fundamental Theorem of Statistics, we know that the empirical distribution function of the error terms is a consistent estimator of the unknown CDF of the error distribution. Because the residuals consistently estimate the errors, it follows that the EDF of the residuals is also a consistent estimator of the CDF of the error distribution. Thus, if we draw bootstrap error terms from the empirical distribution of the residuals, we are drawing them from a distribution that tends to the true error distribution as $n \rightarrow \infty$. This is completely analogous to using estimated parameters in the bootstrap DGP that tend to the true parameters as $n \rightarrow \infty$. Drawing simulated error terms from the empirical distribution of the residuals is called resampling. In order to resample the residuals, all the residuals are, metaphorically speaking, thrown into a hat and then randomly pulled out one at a time, with replacement. Thus each bootstrap sample contains some of the residuals exactly once, some of them more than once, and some of them not at all. Therefore, the value of each drawing must be the value of one of the residuals, with equal probability for each residual. This is precisely what we mean by the empirical distribution of the residuals. To resample concretely rather than metaphorically, we can proceed as follows. First, we draw a random number ′ from the N(0, 1) distribution. Then we divide the interval into *n* subintervals of length 1/*n* and associate each of these subintervals with one of the integers between 1 and *n*. When ′ falls into the *l* th subinterval, we choose the index *l*, and our random drawing is the *l* th residual. Repeating this procedure *n* times yields a single set of bootstrap error terms drawn from the empirical distribution of the residuals. Assuming we have n=10, some of the residuals appear just once in this particular sample, some of them (numbers 2, 3, and 9) appear more than once, and some of them (numbers 1, 4, 5, and 6) do not appear at all. On average, however, each of the residuals appears once in each of the bootstrap samples. If we adopt this resampling procedure, we can write the bootstrap model as

$$y_t^* = X_t\tilde{\beta} + \mu_t^*, \qquad \mu_t^* \sim EDF(\tilde{\mu}_t) \tag{4}$$

Where EDF $(\tilde{\mu}_t)$ denotes the distribution that assigns probability 1/*n* to each of the elements of the residual vector $(\tilde{\mu}_t)$. The model (4) is one form of what is usually called a **Nonparametric bootstrap**, although, since it still uses the parameter estimate $\tilde{\beta}$, it should really be called semiparametric rather than nonparametric. The expectation of the EDF of the residuals is simply their sample mean. Thus, if the bootstrap error terms are drawn from a distribution with nonzero mean, the bootstrap model lies outside the null hypothesis. According to Davidson and MacKinnon (2003) it is, of course, simple to correct this problem. We just need to center the residuals before throwing them into the hat, by subtracting their mean. But if the distribution of the error terms displays substantial skewness (that is, a nonzero third moment) or excess kurtosis (that is, a fourth moment greater than $3\sigma_0^4$), then there is a good chance that the EDF of the re-centered and rescaled residuals does so as well, said Davidson and MacKinnon (2003). More research on bootstraps are still going on, for example, Acha (2016) , examines and discusses a comparative analysis of hypothetical data by using the residual and wild bootstrap methods, including their rescaled versions were applied on the data collected from a normal distribution with different ability levels to check whether they are significant at various assessment conditions. The result showed that rescaled residual functional model out performed all other functional models considered in this paper. Also, the trends at the lower ends of the distributions for each bootstrap model shows that the empirical distributions of true distributions follow the chi-square distribution and also tends to normal distribution as sample size tends to infinity. Gel, Lybchich, and Ramirez (2017), showedthat the new bootstrap method outperforms competing approaches by providing sharper and better-calibrated confidence intervals for functions of a network degree distribution than other available approaches, including the cases of networks in an ultra-sparseregime.

The objective of this paper is to ascertain which group proficiency form (GPF-models) provide more reliable statistical inference by applying nonparametric bootstrap on a normal distribution and define the performance of the GPF-models estimator for forecasting and predicting. The group proficiency forms (ability levels) are very important in a identifying a more reliable statistical model(s) for forecasting and predicting. This objective was achieved by considering seventeen assessment conditions which includes group proficiency forms-upper and lower quartiles of the variance, test lengths, bootstrap levels, sample sizes and the root mean square error. It is pertinent to note that the RMSE is an index that takes both factors- bias and precision together into account.

2. Research Methodology

This paper is based on descriptive research and bootstrap methods as well as different group proficiency forms from statistical normal distributions. R- Statistical package is used to generate the hypothetical data set in this study.

According to Fapas (2017), the determinants in test materials of proficiency form may either be at natural levels, incurred or spiked at a particular requested formulation level. Here, in order to select the group proficiency test or form that will give a more reliable statistical model(s) for forecasting and predicting, this study considered different assessment conditions.

Group proficiency form 1 denotes the population θ distribution for Form X with the standard normal distribution, which is used as the baseline for comparison; and group proficiency form 2 and form 3 stand for the population θ distribution for Form M with $\theta \sim N(0, \sigma^2)$. In bias and RMSE test by Wang, Lee, Brennan, & Kolen, (2008); Wang & Brennan, (2009), a difference of 0.1 standard deviation units is generally considered relatively large, whereas a difference of 0.25 is regarded as very large. This style will be adopted in this study. The seventeen assessment conditions were the result of different combinations of the followings: seven functional models, three proficiency levels (ability levels) including the variance quartiles, two test lengths, two bootstrap levels, three sample sizes and the model RMSE were considered.

3. Results and Interpretation

Table 1, gives the summarized results of bootstrap root mean square error (RMSE) with respect to their group proficiency forms. It went further to show the smallest value in each row, overall smallest in each model, overall smallest and largest in the whole model and group proficiency form (GPF-models). The results with the factor of group proficiency levels and test lengths were mixed for all the models. In Table 2, the following models produced the minimum RMSE at different assessment conditions.

GPF4 - when B=1999, PF=N (0, 1), n=1000, $GPF4_{RMSE}$ = 0.0000 while GPF7- B=1999, PF=N (0, 1), n=1000, $GPF7_{RMSE}$= 0.0996. Even though, GPF4 and GPF7 produces the minimum and maximum of the RMSE respectively, both produce the same RMSE value (0.0130) at B=99, N (0, 0.25), n=1000 assessment conditions.

Regardless of the test length, and group proficiency forms, the factor of sample size and bootstrap level had a significant effect on the standard error. As the sample size, and the bootstrap level increased, the standard error from all these bootstrap models decreased at almost all estimated values. For the factor of group proficiency form, there was no evidence showing its effects on the bootstrap models in Test 1; however for Tests 2, slight effects were detected for the group proficiency difference increased. Sample size also had a significant impact on the RMSE. As the sample size and bootstrap level increased, the RMSE decreased at all estimated values and the differences among the models became smaller. Meanwhile, a longer test had larger RMSE. No substantial differences were observed among the three group proficiency forms for different bootstrap methods in Test 1, but for Tests 2 and 3, the RMSE increased slightly as the group proficiency differences increased. A longer test tended to be associated with larger error. In fact group proficiency form, sample size and bootstrap level were the main factors in estimating the RMSE by bootstrap method and model

Table 1. *Summarized Results of Bootstrap RMSE with respect to their Group Proficiency Forms.*

Bootstrap level (B)	Proficiency form(PF)	Sample Size (n)	GPF1	GPF2	GPF3	GPF4	GPF5	GPF6	GPF7
		120	0.1029	0.1193	0.1123	0.1199	0.1898	**0.0822**	0.1988
	N(0,0.25)	200	0.1776	0.1589	0.1541	**0.0284**	0.1365	0.1376	0.1321
		1000	0.1314	0.1330	0.1307	***0.0130**	0.1172	0.1174	***0.0130**
		120	**0.0028**	0.1190	0.1900	0.1832	0.1978	0.1043	0.1988
B=99	N(0,0.75)	200	0.1758	**0.0008**	0.1340	0.1274	0.1362	0.1380	0.1315
		1000	0.1308	0.1325	0.1163	**0.0011**	0.1069	0.1044	0.1133
		120	0.2011	+**0.0004**	0.1945	0.1922	0.1045	0.1090	0.1060
	N(0,1)	200	0.1762	0.1590	0.1358	0.1371	+**0.0009**	0.1384	0.1320
		1000	+**0.0002**	0.1337	0.1316	0.1140	0.1181	0.1179	0.1135
		120	0.1118	0.1905	0.1709	0.1689	0.1498	**0.0462**	0.1530
	N(0,0.25)	200	**0.0033**	0.1449	0.1340	0.1318	0.1176	0.1149	0.1190
		1000	0.1139	**0.0039**	0.1182	0.1170	0.1082	0.1067	0.1083
		120	0.1128	0.1896	+ **0.0017**	0.1676	0.1495	0.1461	0.1528
B=1999	N(0,0.75)	200	0.1395	0.2051	0.1343	0.1320	0.1178	+ **0.0001**	0.1183
		1000	0.1150	0.1237	0.1179	0.1083	0.1167	0.1083	**0.0067**
		120	0.1131	0.1894	0.1696	**0.0442**	0.1505	0.1478	0.1547
	N(0,1)	200	0.1401	0.1448	0.1342	0.1318	**0.0028**	0.1153	0.1188
		1000	0.1160	0.1250	0.1183	+***0.0000**	0.1082	0.1066 *	**0.2996**

Note. The bold is the smallest value in each row.
**Indicates overall smallest in each model.*
+ Indicates overall smallest and largest in the whole model.

GPF - group proficiency form (GPF-models).

Table2. *Extract result from table 1.*

Functional Model	Bootstrap Levels (B)	Group Proficiency FORM(GPF)	SAMPLE SIZE(n)	Model RMSE
GPF1	99	N(0,1)	1000	0.0002
GPF2	99	N(0,0.75)	120	0.0004
GPF3	1999	N(0,0.25)	120	0.0017
*GPF4	1999	N(0,1)	1000	0.0000
GPF5	99	N(0,1)	200	0.0009
GPF6	1999	N(0,0.75)	200	0.0001
GPF7	1999	N(0,0.75)	1000	0.0067

4. Summary and Conclusions

All the nonparametric bootstrap GPF-models with different proficiency forms and test lengths in RMSE were similar across most of the estimated values, especially when the sample size was equal to or larger than 1,000. Generally it was observed that the larger the sample size, the smaller the bias. Across all combinations of the factors, the standard error from model GPF4 was almost always smaller than those from the other GPF bootstrap models. Thus, model GPF4 also produced the smallest standard error among all the models considered. The RMSE is an evaluation index which reflects both bias and standard error. As indicated in Table 1 and 2, GPF4 model was associated with the smallest RMSE; 0.0000 in Test 1, followed by GPF6; 0.0001 while GPF7; 0.0996 gave the largest RMSE in Test 2. Thus, conclusion was drawn for Tests 1 and 2 when 2% of the estimated at the lower end were excluded that GPF4< GPF6< GPF1< GPF2<GPF5< GPF3< GPF7 and 0.0000< 0.0001< 0.0002< 0.0004< 0.0009< 0.0017< 0.0067 respectively. When all the score points were taken into account, the results for test lengths- Test 1 and Test 2 did not change. Across all models in a hypothetical data set (2% of the estimated value excluded) in the extract result from Table 1 and 2, shows that model GPF4 is the best when B=1999, PF=N(0,1), n=1000.

In conclusion, summarized results of bootstrap RMSE with respect to their group proficiency forms in table 1 and 2, ascertain that the GPF4 model provides more reliable statistical inference among all the models and can be used for prediction or forecasting. Other model considered in this study can also be used but at the point the produce their minimum RMSE. Generally, the larger the sample size and the bootstrap level at N (0, 1), the smaller the bias, the standard error, and the RMSE though, the effects of the group proficiency forms were quite small compared to sample sizes and bootstrap levels.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

References

[1] Acha, C.K. Impact of Housing and Age of Birds on Egg Production. *International Journal of Numerical Mathematics*, 2010, 5(2), 275-298.

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

[3] Acha, C.K. On two Methods of Analysis Balanced Incomplete Block Designs. *Pakistan Journal of Statistics and Operation Research*, 2012, 8(4), 749-757.

[4] Acha, C.K. Parametric Bootstrap Methods for Parameter Estimation in SLR Models. *International Journal of Econometrics and Financial Management*, 2014, 2(5), 175-179. Available online: http://pubs.sciepub.com/ijefm/2/5/2 (accessed on 30 November 2017).

[5] Acha, C.K. Bootstrapping Normal and Binomial Distributions. *International Journal of Econometrics and Financial Management*, 2014, 2(6), 253-256.

[6] Acha, C.K. Rescaling Residual Bootstrap and Wild Bootstrap. *International Journal of Data Science and Analysis*, 2016, 2(1), 7-14.

[7] Acha, C.K.; Acha I.A. Smooth Bootstrap Methods on External Sector Statistics. *International Journal of Econometrics and Financial Management*, 2015, 3(3), 115-120.

[8] Acha, I.A.; Acha, C.K. Interest Rates in Nigeria: An Analytical Perspective. Research Journal of Finance and Accounting, 2011, 2(3), 71-81.

[9] Acha, C.K.; Omekara, C.O. Towards Efficiency in the Residual and Parametric Bootstrap Techniques. *American Journal of Theoretical and Applied Statistics*, 2016, 5(5), 285-289.

[10] Chatfield, C. The Analysis of Time Series. - An Introduction. Chapman and Hall, London, UK. Sixth Edition, 2004.

[11] Chernick, M.R. Bootstrap Methods: A Practitioners Guide. Wiley, New York, 1999.

[12] Chernick, M.R.; LaBudde, R. An Introduction to the Bootstrap with Applications in R Wiley, Hoboken, 2011.

[13] Cover, T.M.; Thomas, J.A. Elements of Information Theory. John Wiley and Sons, 2006, 254.

[14] Davidson, R.; MacKinnon, J.G. Econometric Theory and Methods. New York: Oxford University Press, 2003.

[15] Efron, B.; Tibshirani, R.J. An Introduction to the Bootstrap. Chapman and Hall, New York, 1993.

[16] Fapas. Protocol for Proficiency Test Schemes, Fera Science Ltd (Fera), Version 6, 2017, April. Available online: https://fapas.com/sites/default/files/2017-05/FeraPTSprotocol_pt1_common.pdf (accessed on 30 November 2017).

[17] Freedman, D.A. Bootstrapping regression models. *Ann. Statist.* 1981, 9(6), 1218-1228.

[18] Gel, Y.R.; Lybchich, V.; Ramirez, R.L. Bootstrap quantification of estimation uncertainties in network degree distributions. *Sci Rep.* 2017, 7, 5807. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5517433/ (accessed on 30 November 2017).

[19] Lehikoinen, A.; Saurola, P.; Byholm, P.; Linden, A.; Valkama, J. Life history event of euasian sparrowhawk in a changing climate. *Journal of Avian Biology*, 2010, 41(6), 627-636.

[20] MacKay, D. Information Theory, Inference and Learning Algorithms. Cambridge University Press; First Edition. 2003, ISBN 978-0521642989.

[21] Wang, T.; Lee, W.; Brennan, R.L.; Kolen, M.J. A comparison of frequency estimation and chained equipercentile methods under the common-item nonequivalent groups design. *Applied Psychological Measurement*, 2008, 32(8), 632-651.