Nonlinear Analysis: Modelling and Control, 2009, Vol. 14, No. 3, 357–388

Saddlestrapping

Sarjinder Singh

Department of Mathematics
Texas A&M University-Kingsville
Kingsville, TX 78363, USA
sarjinder@yahoo.com

Received: 2009-02-18 **Revised:** 2009-06-30 **Published online:** 2009-09-11

Abstract. A method of weighted bootstrapping in the presence of auxiliary information has been studied and named as saddlestrapping because there exists a saddlepoint. Comparisons of saddlestrapping with the bootstrapping under different situations are performed and discussed. FORTRAN code for doing bootstrapping and saddlestrapping are provided. A huge scope of further studies has been suggested.

Keywords: bootstrapping, saddlestrapping.

1 Introduction

Bradley Efron is a statistician best known for proposing the bootstrap re-sampling technique, which has had a major impact in the field of statistics and virtually every area of statistical application. The bootstrap was one of the first computer-intensive statistical techniques, replacing traditional algebraic derivations with data-based computer simulation. He received numerous awards on this and related contributions in the field of statistics as cited in the free online encyclopedia entitled Wikipedia. It was named bootstrapping because it involves resampling from the original data set. The bootstrap is a form of a larger class of methods that resample from the original data set and thus also called resampling procedure. For details, one could also refer to the books on bootstrapping by Efron and Tibshirani [1] and Chernick [2].

Casella [3] provides an introduction to the Silver Anniversary of the Bootstrap. Efron [4] discusses a second thought on bootstrapping. Davison et al. [5] have a critical review on recent developments in bootstrap methodology during the year 2003. Beran [6], Lele [7], Shao [8], Lahiri [9], and Politis [10] explain the impact of bootstrap on statistical algorithms and theory, estimating functions, sample surveys, small area estimation and time series, respectively. Ernst and Hutson [11] and Rueda et al. [12, 13] discuss application of bootstrapping for quantile estimation. Holmes [14] and Soltis and Soltis [15] discuss applications of bootstrapping in phylogenetic trees and phylogeny reconstruction respectively. Holmes et al. [16] provide an overview of a conversation on bootstrap between Bradley Efron and other good friends.

The use of auxiliary information in survey sampling has an eminent role to improve methods of sample selection and use them to estimate various parameters of interest. In this paper, we are introducing a new method called saddlestrapping which makes the use of auxiliary information correlated with the study variable to select different bootstrapping samples from the first given original sample selected by any sampling scheme. It is also shown that the bootstrapping method due to Efron [17] is a special case of the proposed saddlestrapping method. Barbe and Bertail [18] have published a monograph on weighted bootstrapping which presents an account of the asymptotic behavior of the weighted bootstrap – a new and powerful statistical technique. Researchers and advanced graduate students studying bootstrap method will find this a valuable technical survey which is thorough and rigorous. The main aim of their monograph is to answer two questions: how well does the generalized bootstrap work? What are the differences between all the different weighted schemes? Lee and Young [19] have discussed pre-pivoting by weighted bootstrap iteration. Johnson [20] has given a nice introduction to bootstrapping. The proposed saddlestrapping is a kind of weighted bootstrapping. The purpose of this paper aims to see the effect of correlation between the study and auxiliary variable on the distributions of sample mean and sample variances; this kind of study has recently been performed by Johnson [20]. Likewise, Hesterberg [21] taught a very valuable course entitled, "Bootstrap methods and permutation tests" during the conference of Statisticians at San Antonio, TX.

In the next section, we develop a theory that will be used in the subsequent sections to show comparisons of the proposed saddlestrapping method with the bootstrapping method due to Efron [17].

2 Proposed methods

Let Y and X be the study and auxiliary variables in a finite population Ω of N units having positive linear correlation with each other. Consider the problem of estimation of population mean $\overline{Y} = \frac{1}{N}\sum_{i\in\Omega} y_i$ of the study variable y while using information on the auxiliary variable x. Note that the distribution of the study variable Y will depend on the value of its correlation with the auxiliary variable X.

Let $s_{original} = \{(y_i, x_i)\colon i = 1, 2, \ldots, n\}$ be a random sample taken from the population Ω by using any sampling scheme among the list of 50 sampling schemes available in Brewer and Hanif [22]. Let us define

$$p_i = \frac{x_i}{x} \tag{1}$$

where $x = \sum_{i\in s_{original}} x_i$ is the total of the auxiliary variable in the original sample $s_{original}$, as the probability of selecting the ith unit from the original sample. We call such a method saddlestrapping. Obviously, the maximum number of saddlestrapping samples will be $\Theta = n^n$ if each saddlestrap sample is of the size of original sample size of n units. Note that for $p_i = 1/n$ the proposed saddlestrapping method leads back to the original bootstrapping method. We suggest the use of Cumulative Total Method to select saddlestrapping samples. However, in case of large sample size, the use of Lahiri [23] method could be beneficial. To learn more about the probability proportional to size and with replacement (PPSWR) sampling one can refer to Hansen and Hurvitz [24].

Now we have the following theorems:

Theorem 1. *An unbiased saddlestrapping (ss) estimator of the population mean of the study variable Y is given by:*

$$\overline{y}_{ss} = \frac{1}{n^2} \sum_{i \in s_{saddle}} \frac{y_i}{p_i}. \tag{2}$$

Proof. Let E_1 and E_2 be the expected values over all possible original samples Ⓢ that could be taken with any Inclusion Probability Proportional to Size (IPPS) design $p(s_{original})$, $original = 1, 2, \ldots,$ Ⓢ from the given population Ω, and over all possible saddlestrapping samples s_{saddle} that could be taken from the given original sample $s_{original}$, respectively. Note that total number of original samples Ⓢ could be either N^n or $\binom{N}{n}$ depending upon whether the original sample $s_{original}$ is selected using with replacement or without replacement sampling from the given population Ω of N units.

Then taking expected value on the both sides of (2), we have

$$\begin{aligned} E(\overline{y}_{ss}) &= E_1 E_2(\overline{y}_{ss} | s_{original}) = E_1 E_2 \left(\frac{1}{n^2} \sum_{i \in s_{saddle}} \frac{y_i}{p_i} | s_{original} \right) \\ &= E_1 \left(\frac{1}{n} \sum_{i \in s_{original}} y_i \right) = E_1(\overline{y}_{original} | s_{original}) \\ &= \sum_{original=1}^{Ⓢ} p(s_{original})(\overline{y}_{original}) = \overline{Y} \end{aligned}$$

which proves the theorem. □

Note that while taking expected value E_2, the original sample $s_{original}$ is treated as a population of size n and then all possible probability proportional to size with replacement (PPSWR) samples of the same size n are drawn from the population $s_{original}$. It works because in the usual PPSWR sampling, the sample size could be equal to the population size.

Theorem 2. *The variance of the unbiased saddlestrapping (ss) estimator $\overline{y}_{ss}$ of the population mean of the study variable Y is given by:*

$$\begin{aligned} V(\overline{y}_{ss}) = \frac{1}{n^3} &\sum_{original=1}^{Ⓢ} p(s_{original}) \sum_{i \in s_{original}} p_i \left(\frac{y_i}{p_i} - n\overline{y}_{original} \right)^2 \\ &+ \sum_{original=1}^{Ⓢ} p(s_{original}) \left(\overline{y}_{original} - \overline{Y} \right)^2. \end{aligned} \tag{3}$$

Proof. Let V_1 and V_2 be the variances over all possible original samples Ⓢ from the given population Ω and over the all possible "saddlestrapping" samples s_{saddle} from the given original sample $s_{original}$, respectively.

Then the variance of the proposed saddlestrapping estimator is given by:

$$\begin{aligned}
V(\overline{y}_{ss}) &= E_1 V_2(\overline{y}_{ss}|s_{original}) + V_1 E_2(\overline{y}_{ss}|s_{original}) \\
&= E_1 V_2\left(\frac{1}{n^2}\sum_{i\in s_{saddle}}\frac{y_i}{p_i}|s_{original}\right) + V_1 E_2\left(\frac{1}{n^2}\sum_{i\in s_{saddle}}\frac{y_i}{p_i}|s_{original}\right) \\
&= E_1\left[\frac{1}{n^3}\sum_{i\in s_{original}} p_i\left(\frac{y_i}{p_i} - n\overline{y}_{original}\right)^2 |s_{original}\right] + V_1(\overline{y}_{original}|s_{original}) \\
&= \sum_{original=1}^{Ⓢ} p(s_{original})\left[\frac{1}{n^3}\sum_{i\in s_{original}} p_i\left(\frac{y_i}{p_i} - n\overline{y}_{original}\right)^2\right] \\
&\quad + \sum_{original=1}^{Ⓢ} p(s_{original})\left(\overline{y}_{s_{original}} - \overline{Y}\right)^2
\end{aligned}$$

which proves the theorem. □

Note that the second component of the variance, given by

$$\sum_{original=1}^{Ⓢ} p(s_{original})\left(\overline{y}_{original} - \overline{Y}\right)^2 \tag{4}$$

disappears if we replace the population mean $\overline{Y}$ with $\overline{y}_{original}$.

Now we have the following corollary:

Corollary 1. *An estimator of variance of saddlestrapping estimator based on each saddlestrapping sample is given by*

$$\widehat{v}\left(\overline{y}_{ss}\right) = \frac{1}{n^2}\left[\frac{1}{n\left(n-1\right)}\left\{\sum_{i\in s_{saddle}}\frac{y_i^2}{p_i^2} - n^3\overline{y}_{ss}^2\right\}\right]. \tag{5}$$

Proof. Note that the estimator (5) can be written as

$$\widehat{v}\left(\overline{y}_{ss}\right) = \frac{1}{n^3\left(n-1\right)}\left[\sum_{i\in s_{saddle}}\frac{y_i^2}{p_i^2} - n\left(\frac{1}{n}\sum_{i\in s_{saddle}}\frac{y_i}{p_i}\right)^2\right] \tag{6}$$

By following Theorem 4.2.3 from Singh [25, p. 309], taking the expected value E_2 on both sides of (6), we get

$$E_2[\widehat{v}(\overline{y}_{ss})] = \frac{1}{n^3}\sum_{i\in s_{original}} p_i\left(\frac{y_i}{p_i} - n\overline{y}_{original}\right)^2. \tag{7}$$

Now taking expected value E_1 on both sides of (7) we get the first term on the right hand side of equation (3), and it proves the corollary.

Thus, it is possible to look at the distributions of saddlestrapping estimator of sample mean $\overline{y}_{ss}$ and that of the estimator of variance $\widehat{v}\left(\overline{y}_{ss}\right)$ obtained from each saddlestrapping sample for various situations such as: different sample sizes, different values of positive correlation between the study variable y and the auxiliary variable x and outliers in the original sample $s_{original}$. Let r_{xy} be the estimator of the population correlation coefficient between the study variable y and the auxiliary variable x in the given original sample $s_{original}$ of n units selected with any sampling design $p(s_{original})$.

Thus an empirical $(1-\alpha)100\,\%$ confidence interval estimate of the population mean $\overline{Y}$ based on Θ each saddlestrapping sample is given by:

$$\overline{y}_{estimate} \mp t_{\frac{\alpha}{2}}(df = n-1)\sqrt{\widehat{v}(\overline{y}_{estimate})} \tag{8}$$

where

$$\overline{y}_{estimate} = \frac{1}{\Theta}\sum_{ss=1}^{\Theta}\overline{y}_{ss} = \frac{1}{\Theta n^2}\sum_{saddle=1}^{\Theta}\sum_{i\in s_{saddle}}\left(\frac{y_i}{p_i}\right) \tag{9}$$

and

$$\widehat{v}(\overline{y}_{estimate}) = \frac{1}{\Theta}\sum_{ss=1}^{\Theta}\widehat{v}(\overline{y}_{ss}) = \frac{1}{\Theta}\sum_{saddle=1}^{\Theta}\{\widehat{v}(\overline{y}_{ss})\}_{saddle}. \tag{10}$$

In the Appendix we shall provide the FORTRAN code which we used to draw histograms by exporting their outputs in MINITAB for different situations which could occur between a study variable y and an auxiliary variable x. Also, the output of the program gives us $(1-\alpha)100\,\%$ confidence interval estimates for both techniques saddlestrapping and bootstrapping and also prints sample correlation between y and x in the first original sample.

3 Graphical and numerical comparisons

In this section, first we show graphical comparisons between bootstrapping and saddlestrapping sample means and sample variances by considering the possibility of different magnitudes of correlation between the study variable y and the auxiliary variable x in the given original sample. Following Singh et al. [26], we generated n independent pairs of random numbers y_i^* and x_i^*, (say), $i = 1, 2, \ldots, n$from the standard normal distribution by using the IMSL subroutine RNNOR. For fixed values of $\sigma_y = 140$, $\sigma_x = 125$, $\mu_y = 415$, and $\mu_x = 335$, we generated transformed variables y_i and x_i as:

$$y_i = \mu_y + \sqrt{\sigma_y^2(1-\rho_{xy}^2)}y_i^* + \rho_{xy}\sigma_y x_i^* \tag{11}$$

and

$$x_i = \mu_x + \sigma_x x_i^* \tag{12}$$

for different values of the population correlation coefficient ρ_{xy} between the study variable Y and the auxiliary variable X. We consider three different cases: (i) reasonable amount of population correlation coefficient ρ_{xy} between x and y such as 0.95, 0.90, 0.80, 0.70, 0.60, and 0.50; (ii) extreme cases of correlation such as 1.00 and 0.00, and (iii) a situation with unknown distribution of Y and X but with large outliers and reasonable amount of correlation such as 0.85. In each case we decided to draw $\Theta = 50,000$ random samples, because we can not handle more than 50,000 data points with the software packages available at present. Later on, we also provide 95 % confidence interval estimates of the population mean using both the bootstrapping and the saddlestrapping methods for various values of correlation between the study and auxiliary variable along with situations of outliers. Here in the entire simulation study we have kept $n = 10$, and 100, because the results are not much sensitive to the sample size. Rather these results are sensitive towards the value of correlation between the study and auxiliary variables. We tried these simulations for $n = 10, 20, 30, 50$ and 1000, and similar results were observed, and one can also verify by running the FORTRAN code given in the Appendix.

3.1 Graphical comparisons

Case 1. Population correlation coefficient between y and x is 0.95. We generated $n = 10$ data points, the values of the study variable y and the auxiliary variable x, from (11) and (12) by setting the value of $\rho_{xy} = 0.95$, as shown in Table 1. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ is 0.95886, which is quite high.

Table 1.

y	172.99	487.79	523.55	662.27	213.75	316.53	318.53	400.94	503.01	531.27
x	186.16	355.15	416.86	575.85	212.24	201.96	261.53	297.62	444.49	422.42

The histograms for sample means shows that spread of saddlestrapping means is much less than that due to bootstrapping sample means. In other words, the presence of correlation between the study variable and auxiliary variable shows impact on the distribution of the bootstrapping sample means which we call the saddlestrapping means. See the histograms in Fig. 1(a) for means and Fig. 1(b) for variances based on both bootstrapping and saddlestrapping methods with small sample size $n = 10$.

Now we generated $n = 100$ data points from (11) and (12) by keeping $\rho_{xy} = 0.95$. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ of 100 units has been observed as 0.96185, and the results are shown in Fig. 1(c) and Fig. 1(d).

As the sample size increases from 10 to 100, the distributions of bootstrapping mean and saddlestrapping means in Fig. 1(a) and Fig. 1(c) show similar behavior. Although the peak point of saddlestrapping means reduces from 6500 to 2800 and the peak point of bootstraping means reduces from 2000 to 1200. The distribution of bootstrapping variance for $n = 10$ and $n = 100$ remains the same as shown in Fig. 1(b) and Fig. 1(d), but the distribution of saddlestrapping variance shows several peaks as the sample size increases. Interestingly, the distribution of bootstrapping variance is almost symmetric, but the distribution of saddlestraping variance remains skewed to the right distributed. As expected the variance of sample variance should be approximately chi-square distribution. Thus, in case of higher correlation between the study and auxiliary variables the distribution of saddlestrapping variance remains close to chi-square distribution as expected.

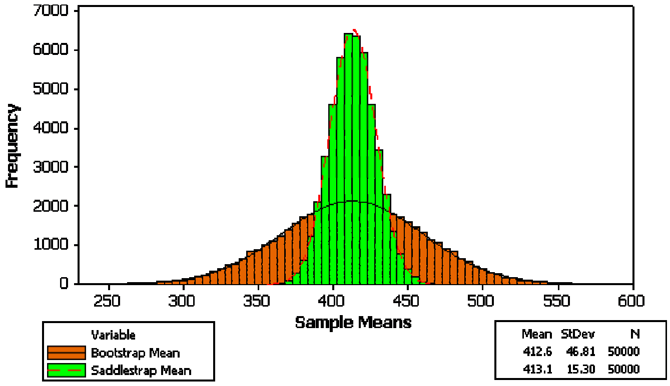

Fig. 1(a). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.95886, sample size 10.

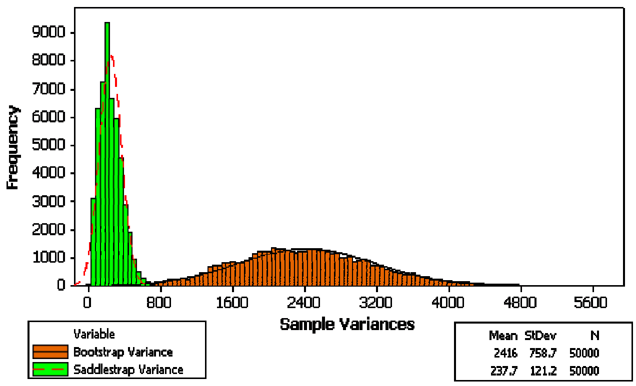

Fig. 1(b). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.95886, sample size 10.

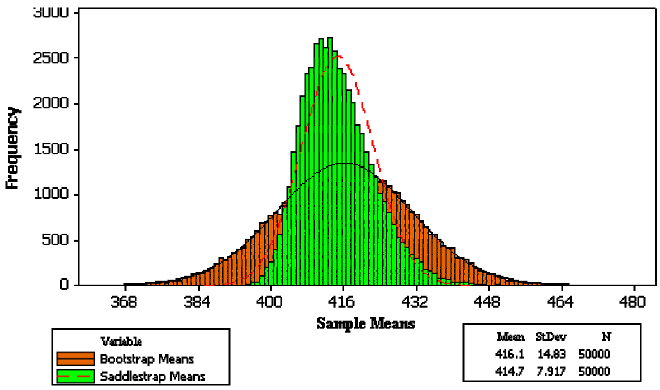

Fig. 1(c). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.96185, sample size 100.

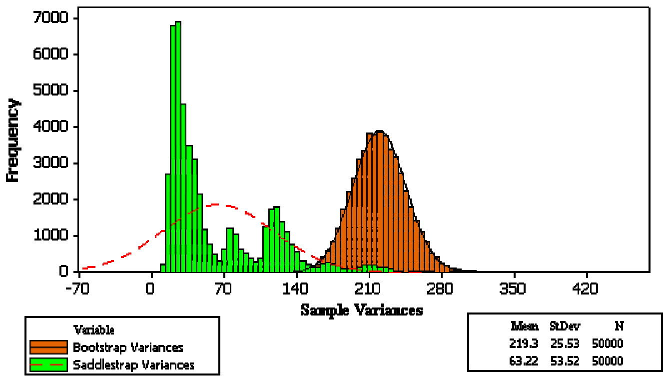

Fig. 1(d). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.96185, sample size 100.

Case 2. Population correlation coefficient between y and x is 0.90. We generated $n = 10$ data points, the values of the study variable y and the auxiliary variable x, from (11) and (12) by setting the value of $\rho_{xy} = 0.90$, as shown in Table 2. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ is 0.89488, which is again quite high.

Table 2.

y	236.22	339.54	567.80	267.99	542.23	364.62	458.49	335.03	322.31	566.64
x	155.99	280.23	470.08	203.25	426.80	373.85	299.63	197.03	300.53	490.98

Again the histogram for sample means shows that the spread of saddlestrapping means is much less than that due to bootstrapping sample means, but remains wider than the case when correlation was 0.95886. Such situations are shown in Fig. 2(a) and Fig. 2(b), respectively.

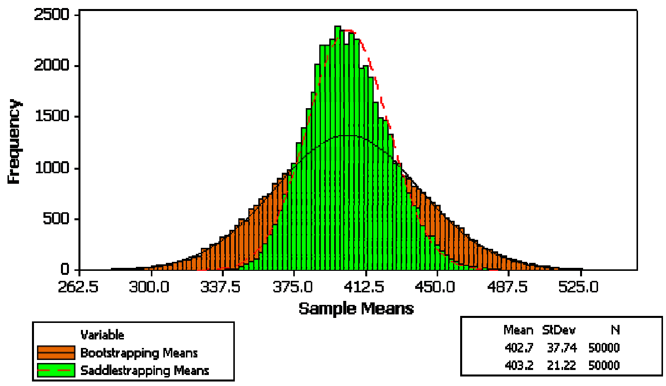

Fig. 2(a). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.95886, sample size 10.

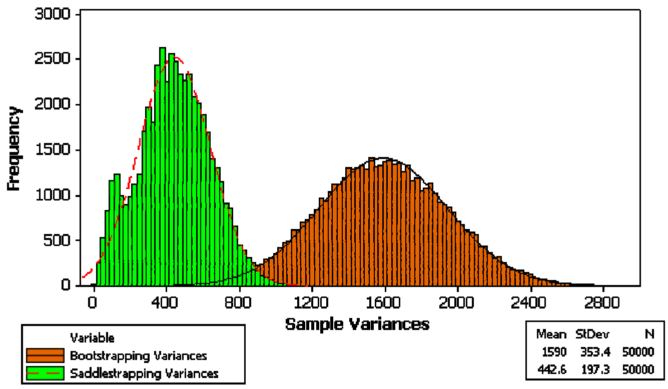

Fig. 2(b). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.95886, sample size 10.

As before now we generated $n = 100$ data points from (11) and (12) by keeping $\rho_{xy} = 0.90$. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ has been observed as 0.91340, and the results are shown in Fig. 2(c) and Fig. 2(d).

As before as the sample size increases from 10 to 100, the distributions of bootstrapping means and saddlestrapping means and sample variances show similar behavior. The behavior of saddlestrapping variance in Fig. 2(d) is more smooth than in case of Fig. 1(d). The reason becomes clear from Case 7 where an extreme situation of a perfect correlation between the study and auxiliary variable has been considered. Too high correlation between the study and the auxiliary variable may lead to single point distribution in case of saddlestrapping.

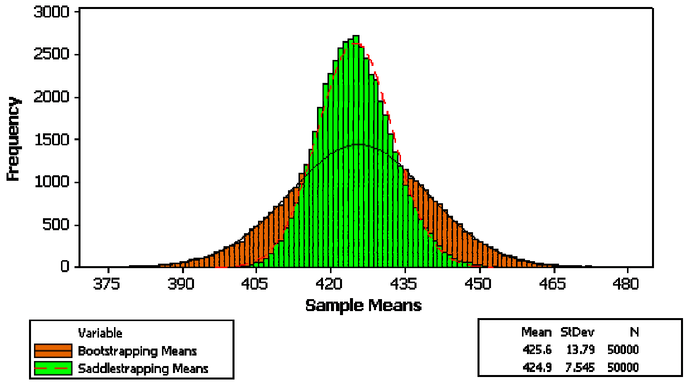

Fig. 2(c). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.91340, sample size 100.

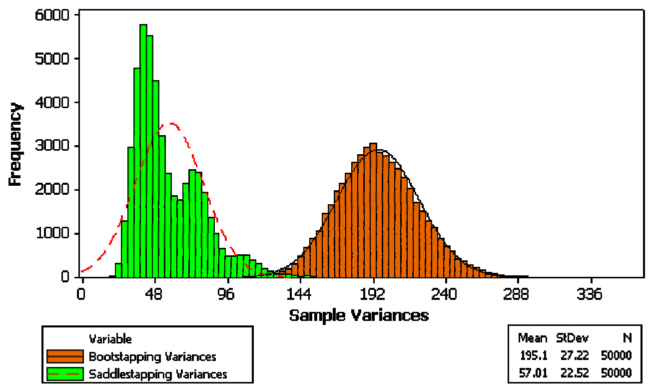

Fig. 2(d). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.91340, sample size 100.

Case 3. Population correlation between y and x is 0.80. We generated $n = 10$ data points, the values of the study variable y and the auxiliary variable x,from (11) and (12) by setting the value of $\rho_{xy} = 0.80$, as shown in Table 3. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ is 0.8011, which is still quite high, but seems practical.

Table 3.

y	235.41	631.68	450.77	344.45	396.21	331.83	459.71	368.44	613.56	587.28
x	73.77	460.22	402.09	347.30	199.07	290.93	494.72	190.33	433.76	442.08

Now the histograms for sample means shows that spread of saddlestrapping means is now comparable to that of bootstrapping sample means, but remains wider than the case when correlation was 0.89488. Such situations are shown in Fig. 3(a) and Fig. 3(b), respectively. The distribution of sample variances in case of saddlestrapping shows higher peak close to zero, which shows that saddlestrapping variance remains smaller than bootstraping variance in most of the cases as shown in Fig. 3(b).

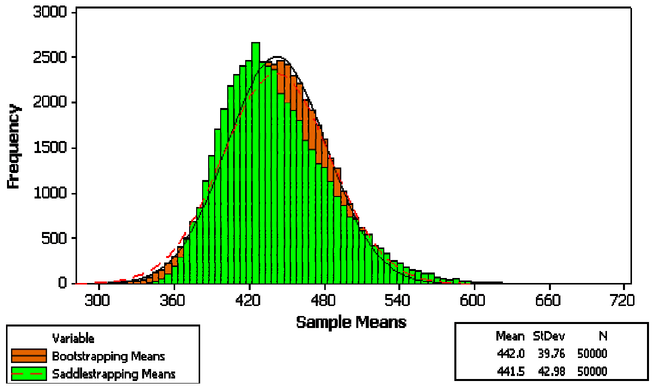

Fig. 3(a). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.89488, sample size 10.

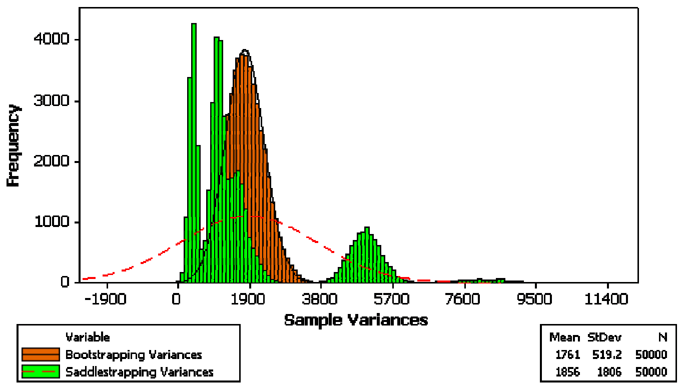

Fig. 3(b). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.89488, sample size 10.

Now we generated $n = 100$ data points from (11) and (12) by keeping $\rho_{xy} = 0.80$. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ has been observed as 0.80121; the results are shown in Fig. 3(c) and Fig. 3(d).

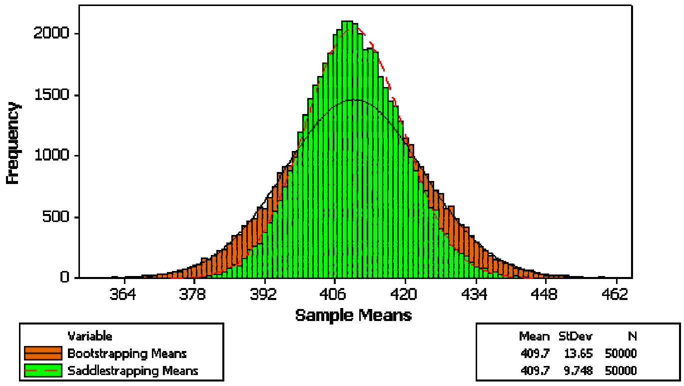

Fig. 3(c). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.80121, sample size 100.

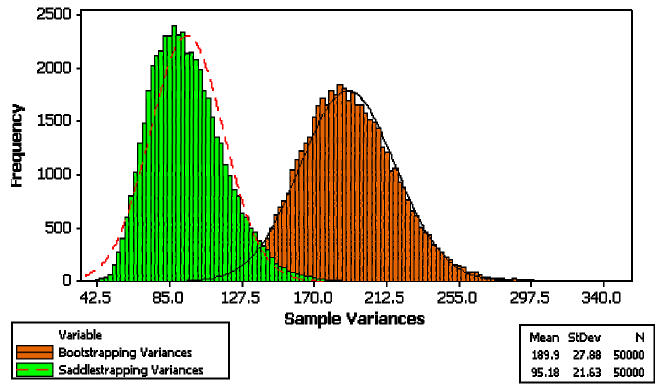

Fig. 3(d). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.80121, sample size 100.

For large sample $n = 100$, the distribution of saddlestrapping means remains narrower than that of bootstrapping means as shown in Fig. 3(c). The values of saddlestraping variances remains smaller than those obtained from bootstrapping variances as shown in Fig. 3(d). Clearly, for $n = 100$, a better inference about the population mean is expected in case of sadlestrapping than in case of bootstrapping with a practicable value of correlation coefficient between the study and the auxiliary variable.

Case 4. Population correlation coefficient between y and x is 0.70. We generated $n = 10$ data points, the values of the study variable y and the auxiliary variable x, from (11) and (12) by setting the value of $\rho_{xy} = 0.70$, as shown in Table 4. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ is 0.72432, which looks very much practical.

Table 4.

y	466.82	198.10	355.36	393.54	252.01	265.36	431.28	470.56	511.32	194.36
x	273.34	267.16	414.20	513.99	244.08	305.24	516.59	376.34	573.73	186.51

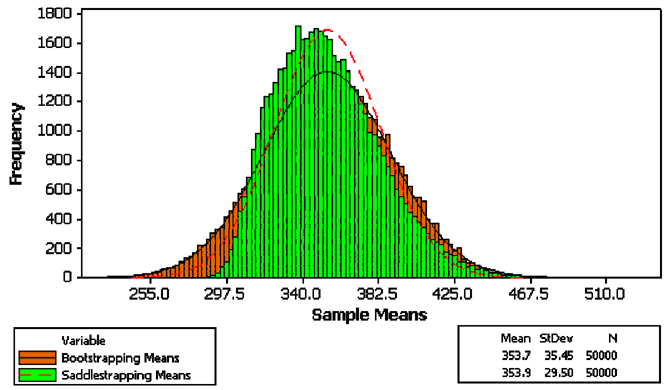

Fig. 4(a). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.8011, sample size 10.

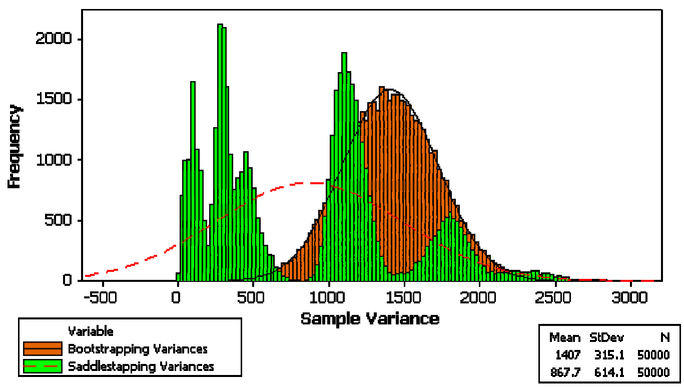

Fig. 4(b). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.8011, sample size 10.

Again for small sample of size $n = 10$, the histograms for sample means in Fig. 4(a) shows that spread of saddlestrapping means is compareable to that of bootstrapping means, but remains wider than the case when correlation was 0.8011. In Fig. 4(b), the saddlestraping variances shows four different peaks with overall distribution skewed to the right. The first two peaks show that the saddlestrapping variances remain smaller than bootstrapping variances in several cases.

Now again we generated $n = 100$ data points from (11) and (12) by keeping $\rho_{xy} = 0.70$. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ has been observed as 0.72313. For a large sample case, the saddlestraping means show less variation than bootstrapping means as shown in Fig. 4(c). In Fig. 4(d) the distribution of saddlestraping variances remains slightly skewed to the right, but most of the values remain smaller than the bootstrapping variances.

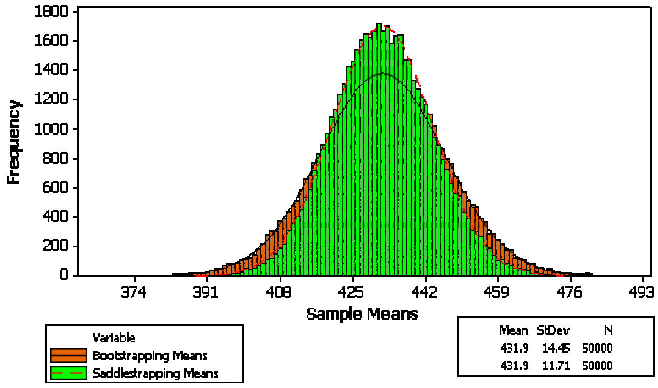

Fig. 4(c). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.72313, sample size 100.

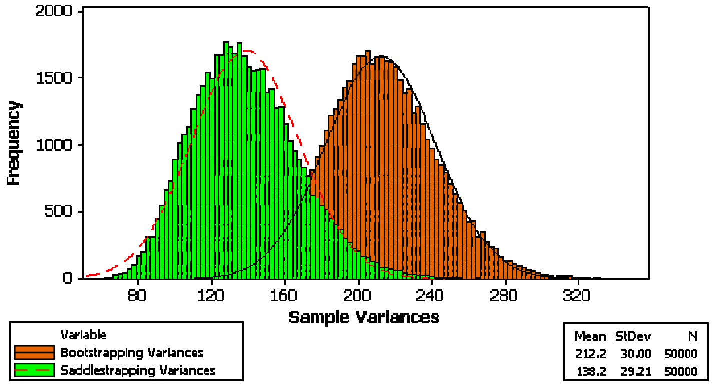

Fig. 4(d). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.72313, sample size 100.

Case 5. Population correlation coefficient between y and x is 0.60. We generated $n = 10$ data points, the values of the study variable y and the auxiliary variable x, from (11) and (12) by setting the value of $\rho_{xy} = 0.60$, as shown in Table 5. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ is 0.58821, which is neither very low nor very high value.

Table 5.

y	409.38	589.87	487.06	493.45	424.96	411.08	241.64	453.19	606.99	610.46
x	250.95	441.37	334.63	392.15	405.83	503.04	193.95	396.62	420.19	381.53

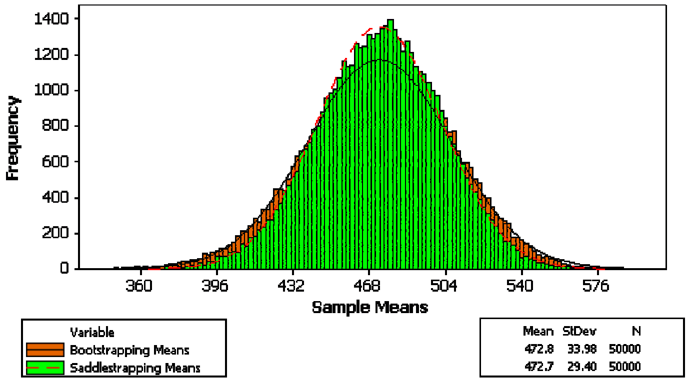

Fig. 5(a). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.6, sample size 10.

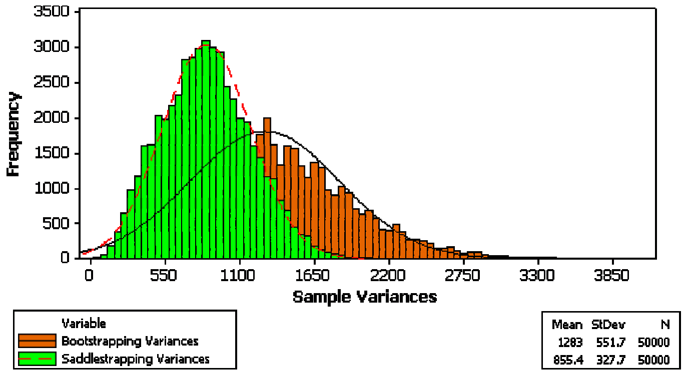

Fig. 5(b). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.6, sample size 10.

We consider a value of correlation between the sampled y and x values as 0.60. The histograms in Fig. 5(a) show that spread of “saddlestrapping” means has almost spread as that of bootstrapping means, and remains wider than the case when correlation was 0.70. Fig.‘5(b) shows that the saddlestrapping variance remains smaller than bootstraping variance in majority of the cases. Thus in case of small sample $n = 10$ with a value of correlation around 0.6, some gain is expected due to the use of saddlestrapping.

Again we generated $n = 100$ data points from (11) and (12) by keeping $\rho_{xy} = 0.60$. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ has been observed as 0.63065, and the results are shown in Fig. 5(c) and Fig 5. 5(d).

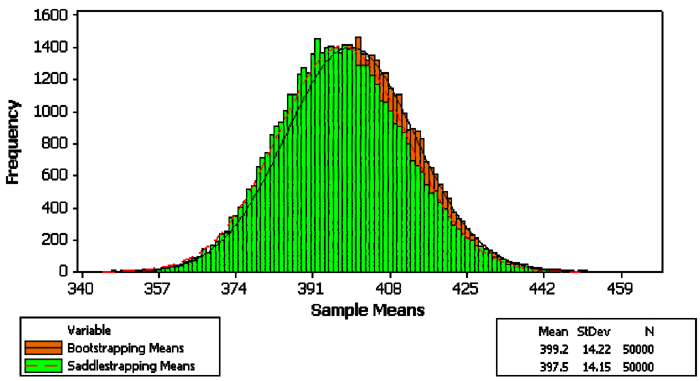

Fig. 5(c). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.63065, sample size 100.

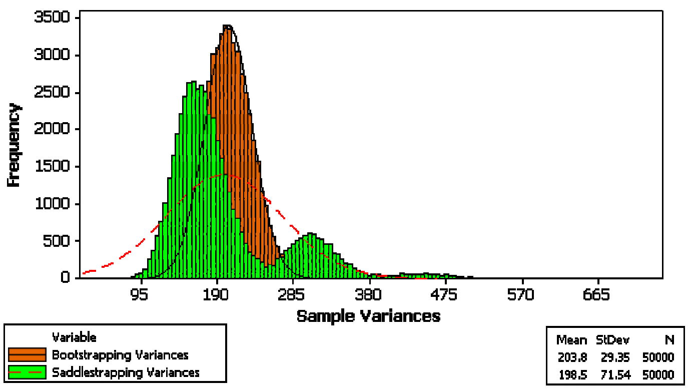

Fig. 5(d). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.63065, sample size 100.

Fig. 5(c) shows saddlestrapping means have almost complete overlap with the bootstraping means; thus, not much gain is expected over the bootstrapping means. Obviously, as we are decreasing the value of correlation coefficient between the study and the auxiliary variables, we are converging to a "saddlepoint", and beyond that point the saddlestrapping will not be useful.

Case 6. Population correlation coefficient between y and x is 0.50. Now we generated $n = 10$ data points, the values of the study variable y and the auxiliary variable x, from (11) and (12) by setting a low value of $\rho_{xy} = 0.50$, as shown in Table 6. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ is 0.48345, which we consider a low correlation. Note that for the ratio estimator (see Cochran [27]) to perform better than the sample mean estimator, the value of correlation should be more than 0.5 by assuming both study and auxiliary variables have equal values of the coefficient of variations.

Table 6.

y	423.17	487.15	290.60	508.19	391.42	455.96	236.20	446.81	412.09	405.16
x	200.62	414.67	295.24	316.85	362.12	486.62	204.31	357.54	426.68	487.04

It is interesting to see that if the correlation between y and x becomes as low as 0.50, then the spread of saddlestrapping means becomes wider than that of bootstrapping means as shown in Fig. 6(a). Fig. 6(b) demonstrates that the variation in saddlestrapping variances remains wider than that of bootstrapping variances. Thus, use of high positive correlation between the study and auxiliary variables is suggested in saddlestrapping, and this concept of high correlation also comes from Hansen and Hurwitz [24].

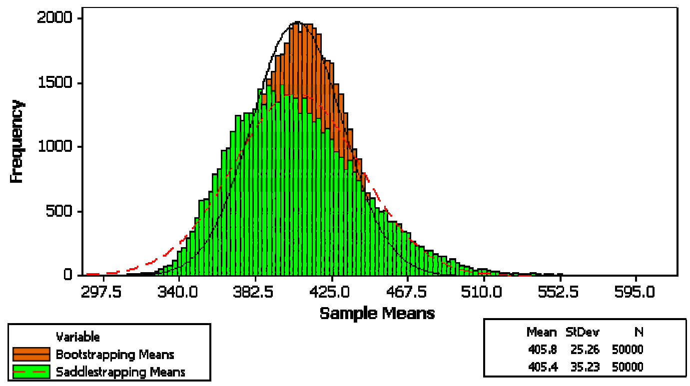

Fig. 6(a). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.50, sample size 10.

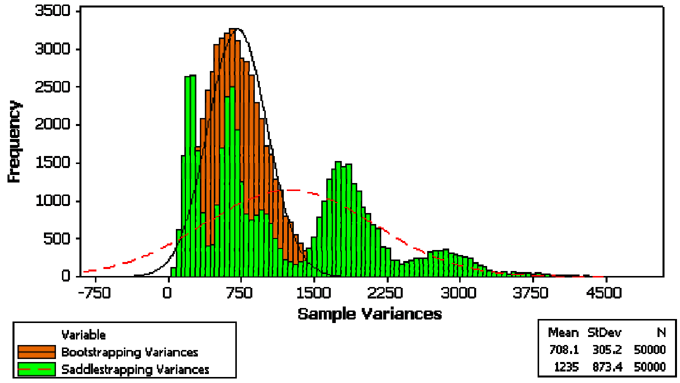

Fig. 6(b). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.50, sample size 10.

To see the effect of sample size, we generated $n = 100$ data points from (11) and (12) by keeping $\rho_{xy} = 0.50$. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ of 100 units has been observed as 0.46443.

For large sample size $n = 100$, and for a low value of correlation coefficient, the saddlestrapping means in Fig. 6(c) show more spread than bootstrapping means. Also Fig. 6(d) shows the saddlestrapping variances have much more variation than bootstrapping variances. Thus, the use of saddlestrapping or weighted bootstrapping in case of low correlation between the study and auxiliary variable is strictly prohibited.

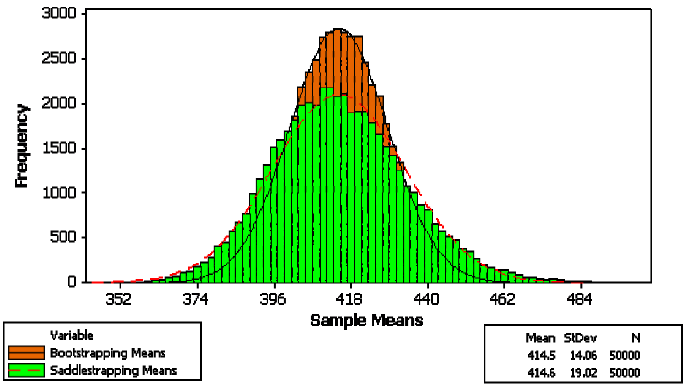

Fig. 6(c). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.46443, sample size 100.

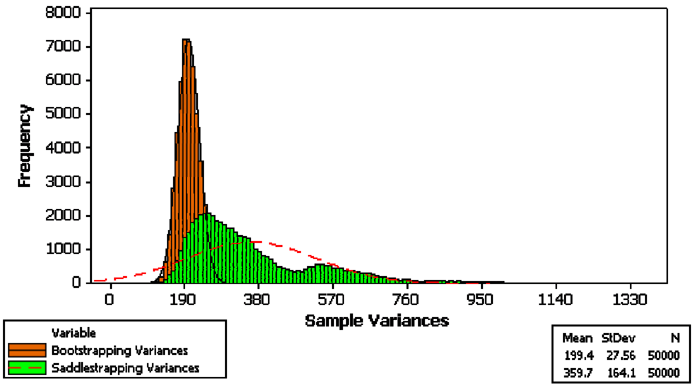

Fig. 6(d). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.46443, sample size 100.

Case 7. (Extreme situation) Population correlation coefficient between y and x is 1.00. We generated $n = 10$ data points, the values of the study variable y and the auxiliary variable x, from (11) and (12) by setting the value of $\rho_{xy} = 1.00$, as shown in Table 7. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ is also 1.00, which is clearly an extreme case.

Table 7.

y	215.99	771.80	237.69	389.56	483.29	550.34	600.82	316.49	411.33	440.37
x	157.31	653.57	176.68	312.28	395.97	455.84	500.91	247.04	331.72	357.65

Such situations for means and variances are shown in Fig. 7(a) and Fig. 7(b) respectively.

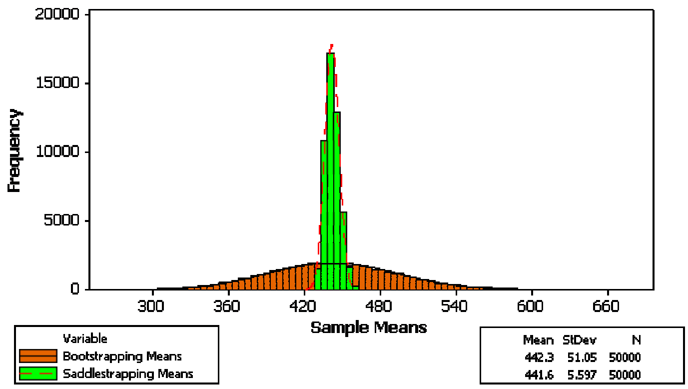

Fig. 7(a). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 1.000, sample size 10.

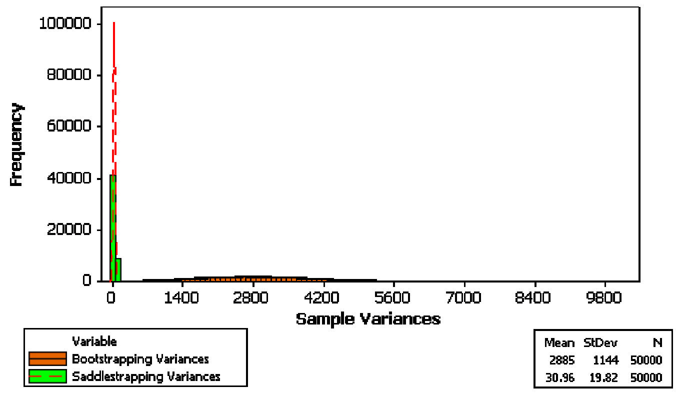

Fig. 7(b). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 1.000, sample size 10.

Fig. 7(a) shows that the saddlestrapping means are converging to a single point distribution, but the distribution of bootstrapping means remains as wider as before because it is not effected by the correlation. The peak of bootstraping means became low because of higher or single value distribution of saddlestrapping means. Similar behavior is shown by saddlestrapping variances and bootstrapping variances in Fig. 7(b).

The following histogram in Fig. 7(c) gives another look at the distribution of the saddlestrapping and bootstrapping means in Fig. 7(a) and saddlestrapping and bootstrapping variances in Fig. 7(b) as:

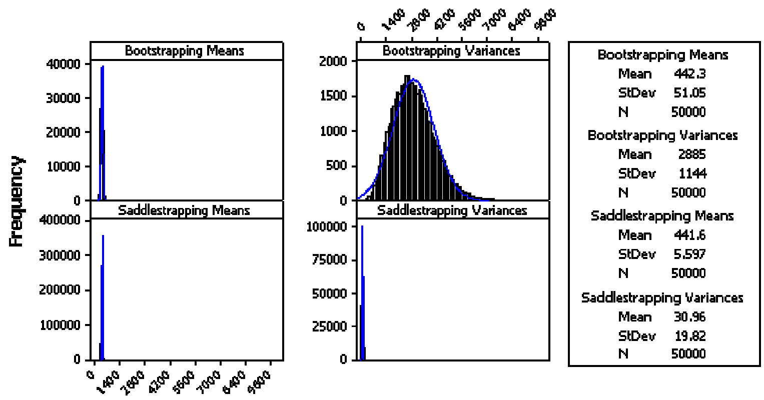

Fig. 7(c). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 1.000, sample size 10.

Note that in Fig. 7(c) bootstrapping means show only one vertical line because these four panel graphs are drawn on the same scale; otherwise, bootstrapping means will show wider bell-shaped distributioon. Thus, do not misinterpret it.

Further note that as soon as the value of the sample correlation coefficient becomes one $(r_{xy} = 1)$, the distribution of saddlestrapping mean $\overline{y}_{ss}$ becomes single point distribution equal to $\overline{y}_{original}$ with zero variance. Also note that the distributions of bootstrapping means and variances remain unaffected. In other words, if there is an auxiliary variable which is positive and perfectly correlated with the study variable, then it is always possible to find y if x is known; no need of saddlestrapping or any other estimation strategy. An analogous of Fig. 7(c) for $n = 100$ is given in Fig. 7(d), where one can see better spread of bootstrapping means on the same scale in a panel graph.

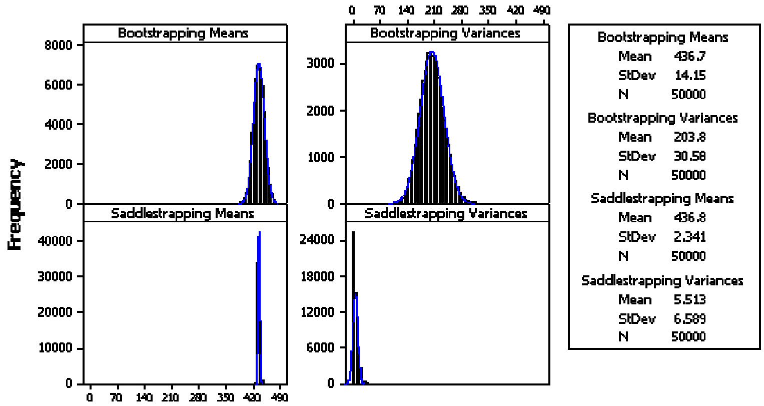

Fig. 7(d). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 1.000, sample size 100.

Case 8. (Extreme situation) Population correlation coefficient between y and x is 0.00. We generated $n = 10$ data points, the values of the study variable y and the auxiliary variable x, from (11) and (12) by setting the value of $\rho_{xy} = 0.00$, as shown in Table 8. The value of sample correlation coefficient r_{xy} between the sampled y and x values in the original sample $s_{original}$ is also 0.0615, which again looks like an extreme case. Here $r_{xy} = 0.0615$ does not mean that x and y are independent, but the value of the correlation coefficient between them is close to zero.

Table 8.

y	306.41	330.72	170.14	478.66	437.12	353.87	338.78	428.14	520.20	336.68
x	236.60	199.27	339.57	391.84	421.92	472.60	469.41	33.31	421.06	445.74

We observed that if the correlation between y and x is 0.0615, then bootstrapping works better than saddlestrapping.

To have another look at Fig. 8(a) and Fig. 8(b), we have Fig. 8(c) and Fig. 8(d).

We generated $n = 100$ data points, the values of the study variable y and the auxiliary variable x, from (11) and (12) by setting the value of $\rho_{xy} = 0.018$. An analogous of Fig. 8(c) and Fig. 8(d) for $n = 100$ are shown in Fig 8(e) and Fig. 8(f).

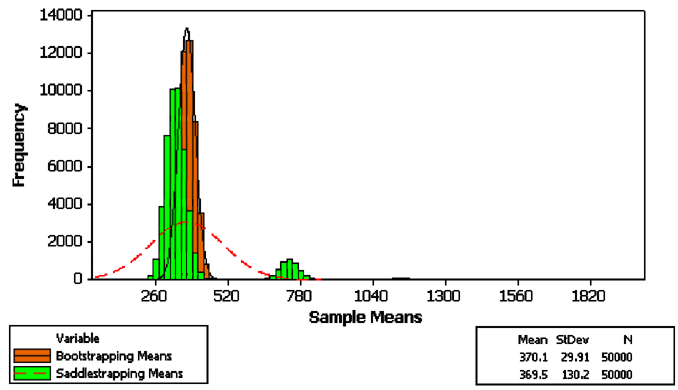

Fig. 8(a). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.0615, sample size 10.

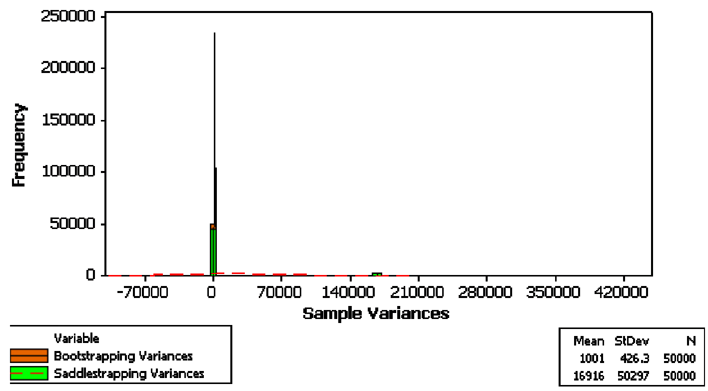

Fig. 8(b). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.0615, sample size 10.

Caution! If bootstrapping weights chosen with the methods of Lee and Young [19] have low correlation with the study variable, then their methods fails at a saddlepoint and lead to the name saddlestrapping instead of weighted-bootstrapping. Thus, it is most important to check if the choice of weights by following Lee and Young [19] have high positive correlation with study variable or not, and, due to the Hansen and Hurwitz [24] contribution, it is a most important factor.

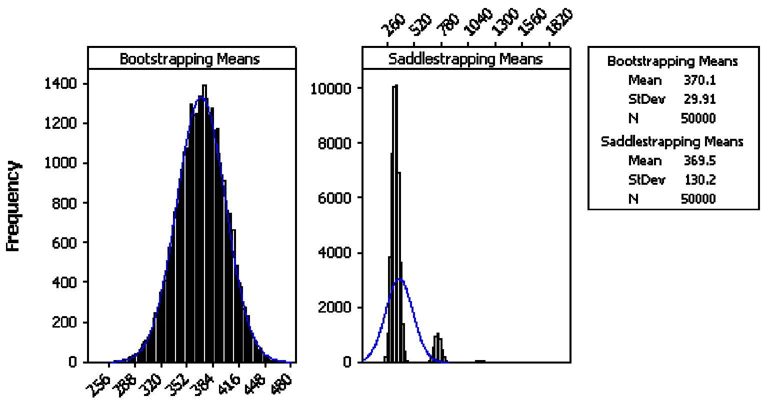

Fig. 8(c). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.0615, sample size 10.

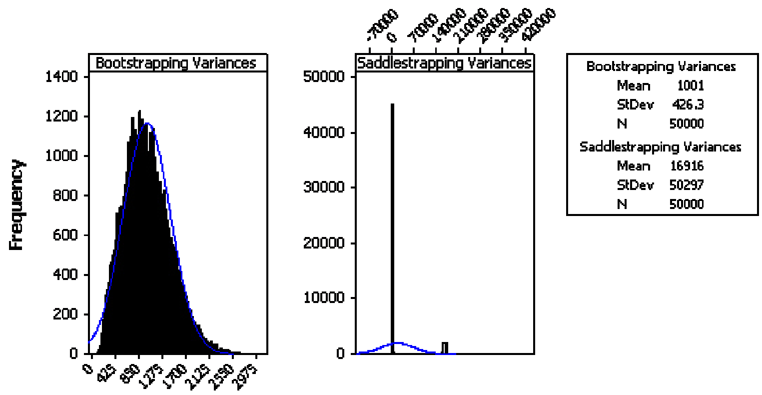

Fig. 8(d). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.0615, sample size 10.

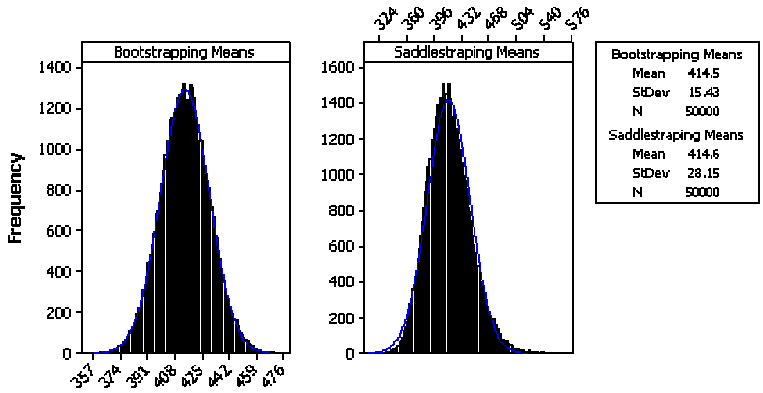

Fig. 8(e). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.018, sample size 100.

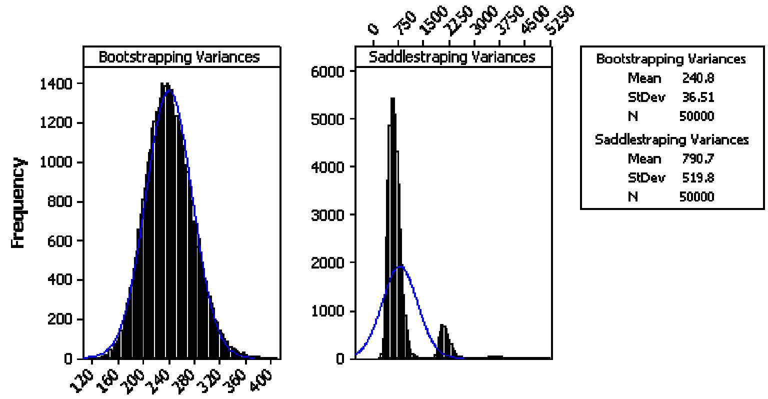

Fig. 8(f). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.018, sample size 100.

Case 9. (Large outliers) Population correlation coefficient between y and x is 0.85. We generated $n = 10$ data points, the values of the study variable y and the auxiliary variable x, from (11) and (12) by setting the value of $\rho_{xy} = 0.85$, and later replaced the last two values with two large outliers as shown in Table 9. Now we do not know the distributions of y and x due to these outliers.

To see the effect of large outliers on the bootstrapping and the saddlestrapping methods, we have the outputs as shown in Fig. 9(a) and Fig. 9(b) for means and variances, respectively. It is true that large outliers have large chance of selection in saddlestrapping and spread in the distributions of mean and variance are quite stable, and do not fluctuate like in the case of the bootstrapping method.

Table 9.

y	235.41	631.68	450.77	344.45	396.21	331.83	459.71	368.44	**9613.56**	**3587.28**
x	73.77	460.22	402.09	347.30	199.07	290.93	494.72	190.33	**7433.76**	**8442.08**

We generated $n = 100$ data points, the values of the study variable y and the auxiliary variable x, from (11) and (12) by setting the value of $\rho_{xy} = 0.85$, and later replaced the last two values with the same outliers as shown in Table 9. To see the effect of large outliers on the bootstrapping and the saddlestrapping methods, we have the outputs as shown in Fig. 9(c) and Fig. 9(d).

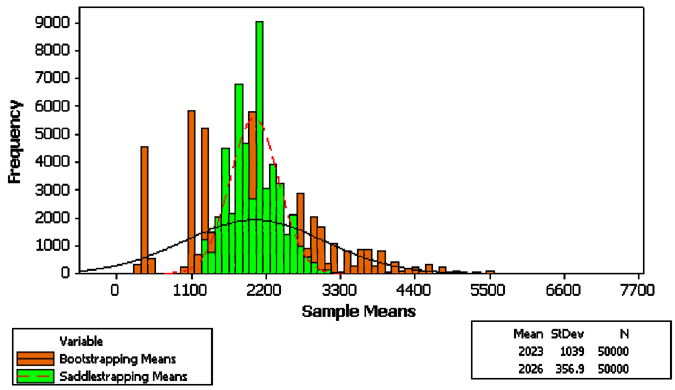

Fig. 9(a). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.82916, sample size 10.

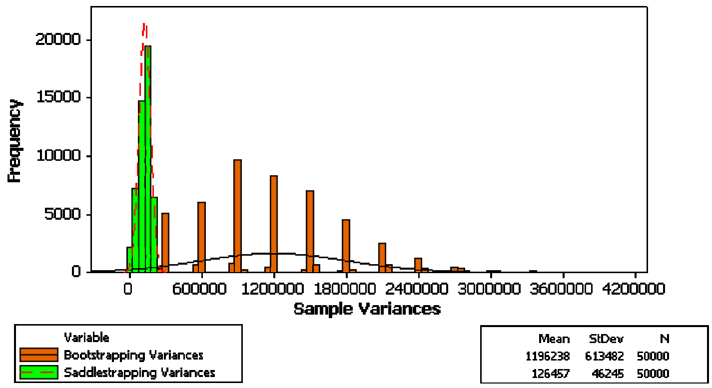

Fig. 9(b). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.82916, sample size 10.

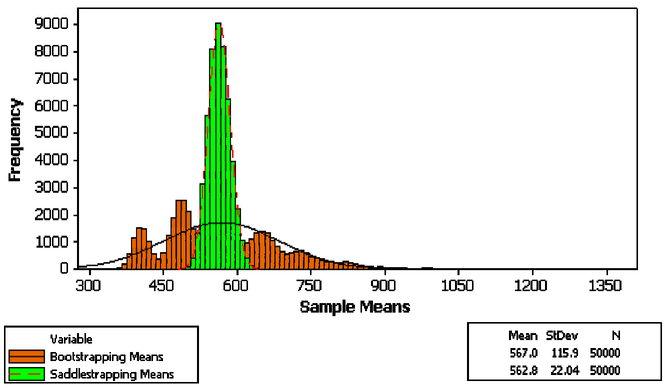

Fig. 9(c). Histogram of bootstrapping and saddlestrapping means; sample correlation coefficient 0.85693, sample size 100.

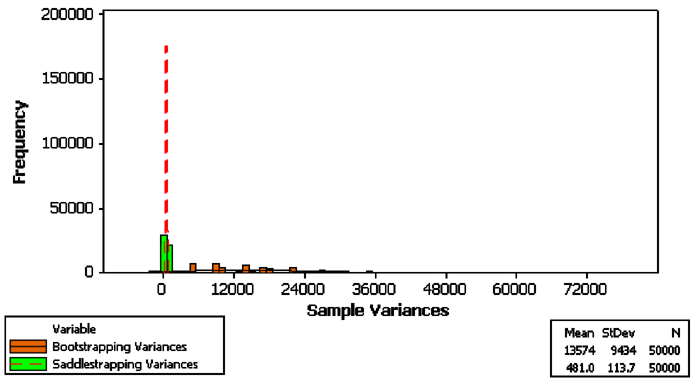

Fig. 9(d). Histogram of bootstrapping and saddlestrapping variances; sample correlation coefficient 0.85693, sample size 100.

From Fig. 9(c) and Fig. 9(d), we can see that even in cases of large sample sizes like $n = 100$, only two outlies are enough to bring the distribution of saddlestrapping means close to single point distribution. It happens because everytime those outliers have very high chance of selection in the saddlestrapping sample. Thus, the use of saddlestrapping or weighted bootstrapping in case of outliers may spoil the inference bout the parameters.

3.2 Numerical comparisons

Table 10 gives the 95 % confidence interval estimates of the population mean obtained by using the method of “bootstrapping” and “saddlestrapping” under different situations discussed before for $\Theta = 50,000$ saddlestraps.

Table 10. A comparison of bootstrap and saddlestrap CI estimates

	r_{xy}	n	95 % CI estimates of the Mean	
			Bootstrapping	Saddlestrapping
Reasonable Cases				
	0.95886	10	$(301.36, 523.74)$	$(378.20, 447.95)$
	0.96185	100	$(386.72, 445.49)$	$(398.92, 430.47)$
	0.89488	10	$(312.52, 492.94)$	$(355.59, 450.76)$
	0.91340	100	$(397.84, 453.27)$	$(409.93, 439.89)$
	0.80110	10	$(347.08, 536.91)$	$(344.07, 539.00)$
	0.80121	100	$(382.35, 437.03)$	$(390.35, 429.06)$
	0.72432	10	$(268.86, 438.58)$	$(287.26, 420.53)$
	0.72313	100	$(402.97, 460.78)$	$(408.55, 455.21)$
	0.58821	10	$(391.73, 553.80)$	$(406.58, 538.90)$
	0.63065	100	$(370.96, 427.56)$	$(369.50, 425.41)$
	0.48345	10	$(345.57, 465.96)$	$(325.94, 484.96)$
	0.46443	100	$(386.53, 442.57)$	$(376.94, 452.20)$
Extreme Cases				
	1.0000	10	$(320.75, 563.75)$	$(429.05, 454.23)$
	1.0000	100	$(408.48, 465.14)$	$(432.07, 441.39)$
	0.0615	10	$(298.52, 441.66)$	$(75.30, 663.75)$
	0.0180	100	$(383.73, 445.32)$	$(358.81, 470.40)$
Large Outliers				
	0.82916	10	$(-451.45, 4496.92)$	$(1221.23, 2830.14)$
	0.85693	100	$(335.86, 798.21)$	$(519.28, 606.31)$

Remember that we generated all populations with true population mean $\mu_y = 415$. Interestingly, for reasonable cases of the value of correlation coefficient considered in Table 10, the true population mean $\mu_y = 415$ lies in both the bootstrapping confidence interval estimates as well as in the saddlestrapping confidence interval estimates. Note that for a low value of correlation coefficient around 0.5, the length of saddlestrapping confidence interval becomes more than that of bootstrapping confidence interval estimates. For a reasonably high value of correlation coefficient, the length of saddlestrapping confidence interval estimates remains smaller than that of bootstrapping confidence interval estimates. It is also interesting to note that for the extreme case of $\rho = 1.0$, the saddlestraping confidence interval estimates do not include the true population mean 415. Thus, it seems that too much high correlation between the study and auxiliary variable in case of saddlestrapping may also provide too narrow confidence interval estimate which may not include true parameters of interest in it. Based on our simulation, a reasonably high value of correlation coefficient (say, 0.8) is recommended in case of saddlestrapping or weighted bootstrapping. In case of zero correlation between the study and auxiliary variable, the confidence interval estimates based on saddlestrapping remain much wider than in case of bootstrapping. The confience interval estimates based on saddlestrapping in the presence of outliers show unusally higher estimates because these outliers have a very high chance of selection in saddlestrapping than in case of bootstrapping. Thus, outliers need to be identified before using saddlestrapping or weighted bootstrapping in real practice.

4 Further study

The idea of saddlestrapping is easily extendable in all directions where bootstrapping has so for been applied by ignoring the availability of auxiliary variable. Obvious examples are: saddlestrapping the distribution of correlation coefficient, ratio estimator, regression estimator, product estimator, median, mode, quantiles and percentiles, etc. under different sampling schemes. It is not possible to tabulate or list all possibilities of studies of saddlestrapping. Researchers could explore or investigate the proposed saddlestrapping in all situations where it is possible. It should also to be seen if a method like saddlestrapping could be made if the correlation between the study variable and auxiliary variable in the first original sample is negative.

Appendix: FORTRAN codes

```
! FORTRAN CODE USED FOR BOOTSTRAPPING AND SADDLESTRAPPING
      USE NUMERICAL_LIBRARIES
      IMPLICIT NONE
      INTEGER I, NS, J, IR(1), IHX(1), ICTX, JJ, II, III, IIII, NHORSES
      REAL Y(60), BY(60), HY(60), SBY, SHY, BYM, HYM, ANS,
     1 VBY, VBH, SY, YM, VY
      REAL X(60), CTX(60), HX(60), P(60),VARP, VX, CXY, RXY, SX, XM
      REAL DF1, T95, ALLHM, AULHM, ALLBM, AULBM
      REAL SUMVBY, SUMVARP, SUMBYM, SUMHYM
      DATA NS /10/
      DATA Y /250, 320, 438, 521, 230, 340, 478, 350, 631, 579/
      DATA X /200, 278, 367, 376, 123, 250, 420, 331, 545, 456/
      CHARACTER*20 OUT_FILE
      WRITE(*, '(A)') 'NAME OF THE OUTPUT FILE'
      READ(*, '(A20)') OUT_FILE
      OPEN(42, FILE = OUT_FILE, STATUS = 'UNKNOWN')
      ANS = NS
      SY = 0.0
      SX = 0.0
      DO 1 I = 1, NS
         SY = SY + Y(I)
 1    SX = SX + X(I)
      YM = SY/ANS
      XM = SX/ANS
      VY = 0.0
```

```
        VX = 0.0
        CXY = 0.0
        DO 2 I = 1, NS
          VY = VY + (Y(I) – YM)**2
          VX = VX + (X(I) – XM)**2
2       CXY = CXY + (X(I) – XM)*(Y(I) – YM)
        VY = VY/(ANS – 1)
        VX = VX/(ANS – 1)
        CXY = CXY/(ANS – 1)
        RXY = CXY/SQRT(VX*VY)
        WRITE(42, 109)RXY, YM, SQRT(VY), XM, SQRT(VX)
109     FORMAT(2X, F9.5, 2X, F9.2, 2X, F9.2, 2X, F9.2, 2X, F9.2)
        SUMVBY = 0.0
        SUMVARP = 0.0
        NHORSES = 50000
        DO 8888 IIII = 1, NHORSES
          WRITE(*,*)IIII
          DO 11 I = 1, NS
            CALL RNUND(1, NS, IR)
            J = IR(1)
            BY(I) = Y(J)
11        CONTINUE
          CTX(0) = 0
          CTX(1) = X(1)
          DO 14 I = 2, NS
14        CTX(I) = CTX(I – 1) + X(I)
          ICTX = CTX(NS)
          III = 0
          DO 26 II = 1, NS
            CALL RNUND(1, ICTX, IHX)
            JJ = IHX(1)
            DO 15 I = 1, NS
              IF((JJ.GE.CTX(I – 1)).AND.(JJ.LE.CTX(I)))THEN
                III = III + 1
                HX(III) = CTX(I) – CTX(I – 1)
                HY(III) = Y(I)
              ELSE
                CONTINUE
              ENDIF
     15     CONTINUE
     26   CONTINUE
          DO 18 I = 1, NS
     18   P(I) = HX(I)/DBLE(ICTX)
          SBY = 0.0
```

```
          SHY = 0.0
          DO 16 I = 1, NS
             SBY = SBY + BY(I)
      16  SHY = SHY + HY(I)/(ANS*P(I))
          BYM = SBY/ANS
          HYM = SHY/ANS
          VBY = 0.0
          DO 17 I = 1, NS
      17  VBY = VBY + (BY(I) – YM)**2
          VBY = VBY/(ANS*(ANS – 1))
          VARP = 0.0
          DO 19 I=1, NS
      19  VARP = VARP + (HY(I)/P(I))**2
          VARP = (VARP – ANS**3*HYM**2)/(ANS**3*(ANS – 1))
          WRITE(42,103)NS, YM, BYM, VBY, HYM, VARP
     103  FORMAT(2X, I4, 9(F14.2,2X))
          SUMBYM = SUMBYM + BYM
          SUMHYM = SUMHYM + HYM
          SUMVBY = SUMVBY + VBY
          SUMVARP = SUMVARP + VARP
 8888  CONTINUE
       SUMBYM = SUMBYM/DBLE(NHORSES)
       SUMHYM = SUMHYM/DBLE(NHORSES)
       SUMVBY = SUMVBY/DBLE(NHORSES)
       SUMVARP = SUMVARP/DBLE(NHORSES)
       DF1 = ANS – 1
       T95 = TIN(0.975, DF1)
       ALLHM = SUMHYM – T95 * SQRT(SUMVARP)
       AULHM = SUMHYM + T95 * SQRT(SUMVARP)
       ALLBM = SUMBYM – T95 * SQRT(SUMVBY)
       AULBM = SUMBYM + T95 * SQRT(SUMVBY)
       WRITE(42, 104)YM, ALLBM, AULBM, ALLHM, AULHM
 104   FORMAT(2X,5(F14.2,2X))
       STOP
       END

!FORTRAN CODE FOR GENERATING DIFFERENT DATA SETS
       USE NUMERICAL_LIBRARIES
       IMPLICIT NONE
       INTEGER I,NS
       REAL YS(2000),XS(2000),Y(2000),X(2000)
       REAL RHO, SY, SX, YMEAN, XMEAN
       CHARACTER*20 OUT_FILE
       WRITE(*,'(A)') 'NAME OF THE OUTPUT FILE'
```

```
      READ(*,'(A20)') OUT_FILE
      OPEN(42, FILE=OUT_FILE, STATUS = 'UNKNOWN')
      NS = 10
      CALL RNNOR(NS, YS)
      CALL RNNOR(NS, XS)
      RHO = 0.80
      SY = 140
      SX = 125
      YMEAN = 415
      XMEAN = 335
      DO 10 I =1, NS
         Y(I) = YMEAN + SY*SQRT(1 – RHO**2)*YS(I) + RHO*SY*XS(I)
10    X(I) = XMEAN + XS(I)*SX
      DO 11 I = 1, NS
11    WRITE(42,107)Y(I), X(I)
107   FORMAT(2X, F9.2, 2X, F9.2)
      STOP
      END
```

Acknowledgements

The author is thanful to the Journal Secretary Dr. Romas Baronas and two learned referees to bring the original manuscript in the present form. Help from a Prof. Kalidas Jana, University of Texas at Brownsville, in reading the paper and a very careful editing of the revised manuscript by Mr. Raymond Garcia III, Department of Language and Literature, Texas A&M University-Kingsville, has also been duly acknowledged.

References

1. B. Efron, R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, 1993.
2. M. R. Chernick, *Bootstrap Methods. A Practitioner's Guide*. Wiley, NY, 1999.
3. G. Casella, Introduction to the silver anniversary of the Bootstrap, *Stat. Sci.*, **18**(2), pp. 133–134, 2003.
4. B. Efron, Second thoughts on the bootstrap. *Stat. Sci.*, **18**(2), pp. 135–140, 2003.
5. A. C. Davison, D. V. Hinkley, G. A. Young, Recent developments in bootstrap methodology, *Stat. Sci.*, **18**(2), pp. 141–157, 2003.
6. R. Beran, The impact of the bootstrap on statistical algorithms and theory, *Stat. Sci.*, **18**(2), pp. 175–184, 2003.
7. S. R. Lele, Impact of bootstrap on the estimating functions, *Stat. Sci.*, **18**(2), pp. 185–190, 2003.

8. J. Shao, Impact of the bootstrap on sample, surveys, *Stat. Sci.*, **18**(2), pp. 191–198, 2003.
9. P. Lahiri, On the impact of bootstrap in survey sampling and small-area estimation, *Stat. Sci.*, **18**(2), pp. 199–210, 2003.
10. D. N. Politis, The impact of bootstrap methods on time series analysis, *Stat. Sci.*, **18**(2), pp. 219–230, 2003.
11. M. D. Ernst, A. D. Hutson, Utilizing a quantile function approach to obtain exact bootstrap solutions, *Stat. Sci.*, **18**(2), pp. 231–240, 2003.
12. M. Rueda, A. Arcos, E. Artes, Quantile interval estimation in a finite population using a multivariate ratio estimator, *Metrika*, **47**, pp. 203–213, 1998.
13. M. Rueda, Martinez-Miranda, A. Arcos, Bootstrap confidence intervals for finite population quantiles in the presence of auxiliary information, *Model Assisted Statistics and Applications*, **1**, pp. 279–290, 2005, 2006.
14. S. Holmes, Bootstrapping phylogenetic trees: Theory and methods, *Stat. Sci.*, **18**(2), pp. 241–255, 2003.
15. P. S. Soltis, D. E. Soltis, Applying the bootstrap in phylogeny reconstruction, *Stat. Sci.*, **18**(2), pp. 256–267. 2003.
16. S. Holmes, C. Morris, R. Tibshirani, Bradley Efron: A Conversation with good friends, *Stat. Sci.*, **18**(2), pp. 268–281, 2003.
17. B. Efron, Bootstrap methods: Another look at the Jackknife, *The Annals of Statistics*, **7**(1), pp. 1–26, 1979.
18. P. Barbe, P. Bertail, *The Weighted Bootstrap*, Lecture notes in Statistics, Vol. 98, Springer-Verlag, 1995.
19. S. M. S. Lee, G. A. Young, Prepivoting by weighted bootstrap iteration, *Biometrika*, **90**(2), pp. 393–410, 2003.
20. R. W. Johnson, An introduction to the bootstrap. Teaching Statistics, *Journal of the Royal Statistical Society D*, **23**(2), pp. 49–54, 2001.
21. T. Hesterberg, *Bootstrap methods and permutation tests*, A course at the conference of Statisticians at San Antonio, Texas during March 2008, 2008.
22. K. R. W. Brewer, M. Hanif, *Sampling with Unequal Probabilities*, Springer, 1983.
23. D. B. Lahiri, A method for sample selection providing unbiased ratio estimates, *Bull. Ins. Statist. Inst.*, **33**(2), pp. 133–140, 1951.
24. M. H. Hansen, V. N. Hurwitz, On the theory of sampling from finite populations, *Ann. Math. Stat.*, **14**, pp. 333–362, 1943.
25. S. Singh, *Advanced Sampling Theory with Applications: How Michael Selected Amy*, Kluwer, The Netherlands, 2003.
26. S. Singh, S. Horn, F. Yu, Estimation of variance of general regression estimator: Higher level calibration approach, *Survey Methodology*, **24**(1), pp. 41–50, 1998.
27. W. G. Cochran, *SamplingTechniques*, 3rd ed., Wiley, 1977.