Irish Accounting Review (1997), Vol. 4, No. 2, 1–23

THE BOOTSTRAP – USEFUL FOR ACCOUNTING RESEARCHERS ?

Tony Brabazon

University College Dublin

ABSTRACT

Over the past 15 years, there has been tremendous interest in the area of computer-intensive statistical methods. One specific technique whose use has become increasingly widespread is the "bootstrap". Using the computational power of computers, it is possible to obtain estimates of standard error and bias and to construct confidence intervals for estimators without having to make assumptions about the sample distribution of the estimator. To date, little use of the bootstrap has been made in accounting research primarily because few accounting researchers were aware of its existence. This article provides an intuitive overview of the bootstrap, demonstrates the technique using three examples and provides information regarding the software which is available to implement bootstrapping techniques.

INTRODUCTION

The primary objective in statistics is to draw inferences about the (unknown) population parameters by examining a sample which has been drawn randomly from the population. Much statistical theory has evolved to aid in this task. However, prior to the widespread introduction of computers, it was difficult to perform lengthy computations. This led to concentration on statistical models which had concise analytical form and which were not expensive from a computational point of view. This imposed a limitation on the usefulness of statistical inference in certain cases. The availability of low cost computing power is beginning to remove these constraints.

One of the major advantages of increased availability of computational power is that statistical inference need no longer be bound by classical theory. Inferences can be made for more complex estimators of population parameters than was possible in the past.

Additionally, the analysis of data need no longer rely on assumptions of gaussian or other well-known distributions. The researcher will be increasingly freed from the need to understand the detailed mathematical underpinnings of the statistical techniques used. Indeed, the use of the new techniques may prevent some mis-applications of classical techniques and thus lead to better data analysis.

AN OVERVIEW OF THE BOOTSTRAP

Suppose a random sample is drawn from a large population and the values obtained are 2.0, 2.5, 2.25, 2.12, 10.0. A point estimate of the population parameter of interest may be made from this sample information. How can the researcher estimate the standard error or bias or construct a confidence interval for this point estimate? Classical statistics relied on mathematical theory to calculate these quantities. The bootstrap aims to estimate these without overt reliance on complex theory but rather by using vast amounts of computing power. The bootstrap belongs to a class of techniques known as resampling techniques. The underlying premise is that the sample contains more information than has typically been used by classical statistical methods. Bootstrap techniques move the traditional sampling analogy one step further. By resampling (with replacement) many times from the original sample, it is possible to construct estimates of standard error, bias and confidence intervals for the chosen estimator. Bootstrapping techniques permit such estimations for many statistics even where no obvious mathematical theory exists as to the sample distribution of that statistic. In the case of the above sample, resampling with replacement could yield a new (or bootstrap) sample such as 2.5, 2.5, 2.12, 2.0, 10.0. Indeed, a bootstrap sample such as 10.0, 10.0, 10.0, 10.0, 10.0 could also be obtained.

The bootstrap samples mirror the original sample in some way, which in turn mirrors the original population in some (possibly unknown) way.

For each bootstrap sample that is drawn from the original sample, it is possible to calculate the relevant statistic of interest. If 1,000 bootstrap samples are drawn, 1,000 bootstrap estimates of the statistic of interest are calculated. From these bootstrap estimates, inferences can be made concerning the relevant population parameter. The above explanation can be stated in a more precise mathematical fashion.

Consider the following general problem. Let R(X,F) be a random variable where $X=(x_1, ..., x_n)$ is a vector of data drawn as a sample from the unknown probability distribution F. We are interested in estimating some feature of the distribution of R, for example E(R), Var(R) or SD(R). Let $\hat{F}$ denote the empirical cumulative distribution function (cdf) of the sample X, $\hat{F}(x) = \frac{\#(X_i \leq x)}{n}$, and define a bootstrap sample as a random sample of size n drawn from $\hat{F}$. Hence, the bootstrap sample (denoted X^*) is simply a sample drawn with replacement from the original data. Define R^* as $R(X^*, \hat{F})$, that is, the version of R computed using the bootstrap sample X^* (the resample) rather than using the original sample. If this procedure is repeated B times, we obtain $R^*_1, ..., R^*_B$. The histogram of the R^* values is often called the bootstrap distribution of R and can be shown in many cases to approximate the true distribution of R.

Estimation of Standard Errors

Once a statistic has been calculated for a set of data it is necessary to determine how accurate an estimate it is of the true (unknown) population parameter. Standard errors are one way of assessing this. If we take a simple example, we can see how the bootstrap can be of use. The researcher is often interested in the 'central tendency' of a set of data. Common measures are the mean, mode and the median. Due to the straightforward analytical properties of means, they are often used as a measure of central tendency. The central limit theorem enables the researcher to calculate a confidence interval for a sample mean when a large sample is available. However, outliers can have a distorting affect on the sample mean and it may sometimes be preferable to use a trimmed mean or some other measure of central tendency.

However, it is not a straightforward procedure to obtain a measure for the standard error of this statistic. The bootstrap algorithm comes to the rescue. By generating many bootstrap samples from the original sample and generating the relevant statistic for each bootstrap sample, the standard error may be estimated. Thus, standard errors may be estimated for many statistics where no obvious mathematical formulae exist. This enables the researcher to use the most appropriate statistic for the task at hand rather than a statistic for which distributional theory exists.

The bootstrap algorithm for calculation of estimated standard errors is as follows:

(Algorithm 1)

Bootstrap Algorithm for Estimating Standard Errors

Data : $x_1, \ldots\ldots, x_n$; Statistic of interest is $R(X,F)$

↓ B Times

Resample: Draw $x^*_1, \ldots, x^*_n$ at random with replacement from $x_1, \ldots, x_n$

Compute $R^*_b = R_b(X^*, \hat{F})$ based on $b=1, \ldots., B$

↓

Estimate the standard error $s.e._F(R)$ by the sample standard deviation of the B replications

↓

$$s.e._{\hat{F}}(R^*) = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} (R^*_b - \overline{R}^*)^2 \right\}^{1/2}$$

$$\text{where } \overline{R}^* = \frac{1}{B} \sum_{b=1}^{B} R^*_b$$

Notice that the bootstrap estimate based on B resamples is a Monte Carlo approximation to the nonparametric maximum likelihood estimate (MLE) of the quantity of interest, the approximation arising from the fact that B is finite rather than infinite.

From the examples given so far, it may appear that bootstrap procedures only apply where the sample data consists of individual data points. This is not so. The individual data items $x_1, \ldots., x_n$ can be much more complex. They could be vectors or matrices. Consequently, it is possible to extend bootstrapping procedures to more complex settings. Regression models are one such example.

Regression Models

The common OLS linear regression model depends on several assumptions. The model is assumed to be linear in the parameters ($\beta_{0,}$ β_1 etc.), the error terms (ε) are assumed to be drawn from an error distribution (Z) where $E_Z(\varepsilon) = 0$ and the distribution is assumed to be normal. In the simple regression case, it can be shown that the bootstrap estimates of standard error are asymptotically identical to those of classical theory (Hjorth, 1994). However, the same bootstrap procedure used to produce estimates of standard error in the simple OLS case can be applied to much more complex regression models for which no simple mathematical theory exists. Examples of this would be situations where the regression function is non-linear in the parameters and/or where the fitting criterion is not least squares. This allows the researcher the freedom to use the most appropriate model rather than restricting the analysis to modeling techniques that are amenable to simple mathematical analysis. If used wisely, this increased power can only strengthen the researcher's analysis.

To demonstrate the use of bootstrapping techniques, we will look at the simplest case.

Suppose we have a standard linear regression model $Y_i = x_i \beta + \varepsilon_i$ ($i = 1, ..., n$) where Y_i is the ith response variable, β is the vector of regression coefficients, x_i is the ith row of predictor variables and ε_i is the ith error term.

Two simple bootstrapping methods may be used (Efron and Tibshirani, 1993). Resampling can be applied either to the residuals or to the data vectors. The second case will be examined here.

If we consider that all the original data observations (the Y_is and the x_is) form the rows of a data matrix with say 'n' rows, it is possible to randomly select 'n' rows with replacement from the original data matrix and construct a bootstrapped (resampled) data matrix. Performing OLS on this matrix will generate bootstrapped values for the predictor variables (the βs). Having generated B (a large number) of these, the bootstrap algorithm for calculation of the standard error is applied to each regression coefficient in turn and the values for the standard errors of each coefficient are obtained.

Given that two methods of bootstrapping are possible in the regression case, which is preferable? It depends on the nature of the underlying system being modeled. Bootstrapping of residuals does assume that the distribution of error terms is the same for all values of the dependent variables. This is not always so. Consequently, bootstrapping data vectors may sometimes be preferable. If resampling of the data vectors were to be performed, the following algorithm would be applied.

(Algorithm 2)

Bootstrap Algorithm for Resampling Data Vectors

Create a new data matrix by resampling 'n' rows with replacement from the original data matrix

↓

Calculate the OLS regression coefficients for the bootstrapped data matrix

↓

Repeat the process B times

↓

Calculate the bootstrap estimate of standard error for each regression coefficient based on the B bootstrapped values of that coefficient using algorithm 1 above.

The choice of bootstrapping methods available in the regression case indicates that there may not be a unique way to use bootstrap concepts in attacking a statistical inference problem. The final choice of method may depend on the computational complexity of each.

Construction of Confidence Intervals

Whilst a point estimate of a population parameter provides some information, an estimate of its variability provides additional useful information. An estimate (bootstrap or otherwise) of the standard error of a statistic is of limited use if we have no idea as to the sample distribution of that statistic. Commonly, assumptions are made (and not always well tested) that a statistic has some well-known sample distribution such as the gaussian distribution. Clearly the bootstrap methodology will have to provide some means to construct confidence intervals if it is to enjoy widespread use.

Several bootstrap methods for the construction of confidence intervals are in common use. Two methods are briefly mentioned here. The reader seeking more information is referred to Hjorth (1994).

The simplest way to construct a bootstrap confidence interval is to use the bootstrap percentile method interval. It is one of the most commonly-used methods due to its simplicity but it can suffer from poor coverage of the interval (normally undercoverage, i.e. an interval appearing to give 95% coverage may actually only be giving 89% coverage) although it does provide intervals with stable length. How does the method work?

The basic premise of the method is that the bootstrap distribution of θ^* (denoting the bootstrap version of the statistic) conditional on the data $x_1, ..., x_n,$ is similar to the unknown, unconditional distribution of θ (the statistic of interest). A very large number (say B) θ^*s are calculated and are then ordered from lowest to highest to obtain

$$\theta^*_{(1)} \leq .. \leq \theta^*_{(B)}\,.$$

A two sided 100(1-α)% interval for θ is

$$(\theta^*_{((\frac{\alpha}{2}B)+1)}, (\theta^*_{(((1-\frac{\alpha}{2})B)+1)})$$

where $(\frac{\alpha}{2}B)$ denotes the integer part of $\frac{\alpha}{2}B$ etc.

These values are merely the relevant sample quantiles of the bootstrap histogram of the $\hat{\theta}^*$ values.

A second type of percentile method is the percentile - t method. This is based on the asymptotically pivotal quantity $(\hat{\theta} - \theta)/\ s.e.(\hat{\theta})$ where $\hat{\theta} = \theta(x_{1,} \ldots, x_n)$ (the statistic based on the initial sample) and $s.e.(\hat{\theta}) = s.e.\ \theta(x_{1,} \ldots, x_n)$. Consequently, to use this method it is necessary to estimate the standard error of $\hat{\theta}^*$. The bootstrap can be used to do this but, as shall be seen, this introduces a 'double bootstrap' element into the calculations. For each bootstrap sample drawn from the original sample, it is necessary to calculate an estimate of standard error for the bootstrap estimator calculated from that specific resample. This involves a second bootstrap resampling which takes place by resampling from that bootstrap sample.

The basic premise here is that the bootstrap distribution of $(\hat{\theta}^* - \hat{\theta})/s.e.(\hat{\theta}^*)$ (where $s.e.(\hat{\theta}^*)$ is the bootstrap estimate of the standard error of $\hat{\theta}^*$), conditional on the data, should closely approximate the unconditional distribution of $(\hat{\theta} - \theta)/\ s.e.(\hat{\theta})$.

The need for a double bootstrap greatly increases the computational burden, generally by a factor of 100 or more. Better methods of constructing confidence intervals exist and more information can be found in the references.

The use of bootstrap techniques to construct confidence intervals gives the researcher a powerful new tool. Whatever the mathematical complexity of the sampling distribution of the statistic/regression coefficient of interest, a point estimate and a confidence interval for that estimate can be constructed. This frees the researcher from the need to restrict his/her analysis to the small subset of statistics for which distribution theory has simple mathematical form. It opens up the opportunity for the researcher to develop new statistics of interest for the research question at hand. This enables the researcher to precisely target the analytical tools used on the research question rather than using general purpose tools of analysis which may not be fully appropriate for the question under review.

PRACTICAL APPLICATIONS OF THE BOOTSTRAP

Three examples of the bootstrap are developed to demonstrate how the technique can be applied in a wide variety of settings. The first example demonstrates how the bootstrap can be used to develop estimates of standard error and confidence intervals for the correlation coefficient. The second demonstrates the estimation of standard error and confidence intervals for the median. Traditional statistical methods cannot calculate confidence intervals for these items without making restrictive assumptions. The use of bootstrap techniques enables the researcher to extend his/her analysis beyond the limits imposed by traditional statistical methods. The final example shows how the standard errors of regression coefficients may be estimated.

The three examples are constructed for the purpose of demonstration of the bootstrap technique. The interpretation of the statistical results in each case is not obvious as no attempt has been made to clearly define the relevant population from which the sample data was drawn. The basic assumption in each example is that the sample data used in the bootstrap procedures has been randomly drawn from some population. The lack of precision in defining the relevant population does not impact on the mechanical procedures in applying bootstrap procedures to the sample data.

The Correlation Coefficient

Correlation coefficients are often quoted in research papers and, commonly, little effort is expended in showing how accurate these point estimates are. This is similar to calculating a mean and drawing conclusions from it without estimating its standard error or confidence interval. In the following example, grade data for undergraduate accounting students is examined to demonstrate how a standard error and a confidence interval may be estimated for a correlation coefficient.

The grade data was drawn from the first and second year accounting grades of 105 college students (the population). The true population correlation coefficient was calculated for the students and was found to be 0.679.

To demonstrate how the bootstrap could be used had only a sample of student grades been available, a random sample of 15 students was drawn from the original 105. One hundred bootstrap resamples were then drawn from this random sample and the bootstrap estimate of standard error for the random sample was calculated using algorithm 1. To assess the stability of the estimate, the entire procedure was repeated 10 times to obtain 10 different estimates.

The bootstrap calculations were performed using a spreadsheet. The resampling procedure was performed by randomly generating a series of numbers (0-1) and combining these with a series of conditional IF statements to determine which students' grades were selected to construct the initial sample of 15 students. A similar procedure was repeated to generate 100 bootstrap resamples of 15 students from the initial 15 students selected. The built-in spreadsheet function for calculation of a correlation coefficient was used to calculate the bootstrapped correlation coefficient for each of these 100 resamples. The results are as follows:

Table 1

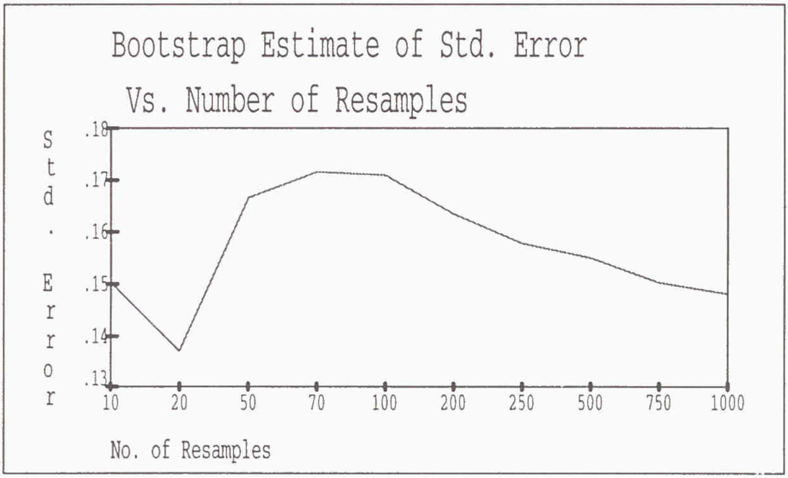
Trial	Mean Bootstrap Estimate of Correlation	Maximum	Minimum	Range	Bootstrap Estimate of Standard Error
1	0.7055	0.9393	0.2678	0.6715	0.1408
2	0.5549	0.9248	-0.3790	1.3038	0.2267
3	0.4540	0.9003	-0.4449	1.3452	0.2300
4	0.4917	0.8932	0.0230	0.8702	0.1770
5	0.5369	0.8787	0.0859	0.7928	0.1640
6	0.7616	0.9628	0.4676	0.4952	0.0980
7	0.7337	0.9217	0.0488	0.8729	0.1418
8	0.7987	0.9569	0.4997	0.4572	0.0811
9	0.7111	0.9470	0.2501	0.6969	0.1170
10	0.5348	0.8703	-0.1003	0.9706	0.1868

As expected, the 10 different random samples produced differing estimates of the standard error. It is noticeable that the trials that had a mean bootstrap estimate of the correlation which was some distance from the true correlation (0.679) also produced the larger estimates of standard error. This seems intuitive, as the fact that the mean is far from the true correlation would tend to indicate that the sample contains some outlying student grade observations. This increases the variability between the correlation figures produced by the bootstrap resampling procedures within that trial and consequently would inflate the bootstrap estimate of standard error. The range of the bootstrap correlation coefficients in each trial further demonstrates this. Those trials with the greatest spread between the highest and lowest bootstrap correlation coefficients tended to have the larger estimates of standard error.

Another consideration in using the bootstrap to estimate standard error is the number of bootstrap resamples that should be taken from the initial sample to ensure the bootstrap estimate of standard error is stable. Using a spreadsheet, a random sample of 15 student grades were drawn from the population and 10, 20, 50, 70, 100, 200, 250, 500, 750 and 1,000 resamples respectively were drawn from this sample. For each of these resample sizes, the bootstrap estimate of standard error was calculated and the results have been graphed below (**Diagram 1**). The graph below shows the effect on the bootstrap estimate of standard error as the number of resamples taken from the initial sample increases. Clearly, when the number of resamples is small, the bootstrap estimate of standard error is quite variable. As the number of resamples increases, the bootstrap estimate of standard improves in stability. The graph shows that the bootstrap estimate of standard error is quite stable after only a few hundred resamples.

The number of bootstrap resamples used when estimating the standard error can often be limited to approximately 100.

Diagram 1



Confidence Intervals for the Correlation Coefficient

Two methods for the calculation of confidence intervals have already been outlined. The simplest, the percentile method, will be used to demonstrate how a confidence interval may be constructed. Initially, a confidence interval will be constructed using the sample from which the 1000 resamples above were drawn.

The 5th, 10th, 90th and 95th percentiles of the histogram of the 1,000 bootstrap estimates of the correlation coefficient can be obtained either by using spreadsheet functions to sort the bootstrapped correlation coefficients or by importing the bootstrapped correlation coefficients into a statistical package such as SPSS and using the built-in functions to determine the relevant percentiles. The percentiles form approximate confidence intervals. Confidence intervals are also calculated for cases where the number of resamples drawn from the original sample are 100 and 500 respectively. The results are:

Table 2

Number of Resamples	5th Percentile	10th Percentile	90th Percentile	95th Percentile
100	0.180	0.306	0.731	0.788
500	0.266	0.347	0.724	0.768
1000	0.276	0.355	0.720	0.758

It is noteworthy that the above intervals all contain the true, known value of the population correlation coefficient. Returning to the initial 10 trials in which a different random sample of 15 students was drawn on each occasion and on which 100 bootstrap resamples were then performed, we get the following estimated confidence intervals:

Table 3

Trial Number	5th Percentile	10th Percentile	90th Percentile	95th Percentile
1	0.430	0.512	0.860	0.911
2	0.165	0.307	0.803	0.884
3	0.030	0.178	0.708	0.791
4	0.190	0.271	0.706	0.805
5	0.255	0.285	0.745	0.804
6	0.619	0.633	0.888	0.892
7	0.493	0.558	0.867	0.880
8	0.659	0.704	0.900	0.921
9	0.493	0.546	0.844	0.866
10	0.140	0.264	0.751	0.823

All the 5th–95th percentile intervals contain the true value of the correlation coefficient. Nine out of the ten 10th–90th percentile intervals contain the true value.

The Median

Several measures of the central tendency within a data set exist. Most analysis of central tendency focuses on the mean due to the relative ease with which standard errors and confidence intervals can be constructed. However, outliers can dramatically affect the mean and so it may be preferable to use another measure of central tendency. One alternative is the median but this raises the problem of how to determine standard errors and confidence intervals for this measure. This example shows how the bootstrap can be used to calculate a standard error and a confidence interval for the median.

The data used for this example consists of a subset of financial statements filed by companies on the Irish Stock Exchange and by statutory bodies for the year 1991 (O'Brien and McCallig, 1995). Banks, government bodies and exploration companies were removed from the sets of financial statements examined. This left a population of 70 companies from which smaller samples would be drawn in the bootstrap calculations. The current ratios of these companies were computed and the object was to calculate an estimate of standard error and a confidence interval for the median current ratio. When looking at such data, the mean may not be an appropriate measure of central tendency as very high ratios will have undue influence (the distribution of ratios is truncated at zero). The removal of outliers (ie. the use of a trimmed mean) will potentially lead to the loss of some information. Additionally, trimming the mean will result in the use of an estimator of central tendency which does not have an obvious sampling distribution. Under traditional statistical methods, there is no simple way to calculate a confidence interval for such an estimator. Use of the median enables the researcher to use all of the sample data (avoiding potential information loss) and the application of bootstrap techniques enables the researcher to construct a confidence interval for the median.

The following population parameters were calculated from the entire 70 companies:

Mean	1.49
Median	1.43
Std. Error of the Mean	0.085

To demonstrate the bootstrap, a random sample of size 15 was drawn from the total population of 70 companies. This process was repeated 10 times. Ten bootstrap estimates of standard error were calculated for each of the random samples of size 15.

The resampling and the calculation of the median current ratio for each bootstrap sample was performed using a spreadsheet. The results were as follows:

Table 4

Sample size: 15

Trial Number	Mean of Bootstrap Medians	Minimum	Maximum	Range	Bootstrap Estimate of Standard Error
1	1.3667	1.153	1.692	0.539	0.0976
2	1.4223	0.982	1.708	0.726	0.1079
3	1.3710	1.188	1.692	0.504	0.1042
4	1.4310	1.258	1.50	0.242	0.0477
5	1.4516	1.258	1.706	0.448	0.1041
6	1.4898	1.369	1.531	0.162	0.0352
7	1.4509	1.126	1.525	0.399	0.0623
8	1.3084	1.022	1.511	0.489	0.1014
9	1.5098	0.857	1.839	0.982	0.1351
10	1.3936	0.887	2.239	1.352	0.2890

The standard errors are not large compared with the mean of the medians in each trial. Examination of the range of bootstrap estimates of the median in each trial shows that those trials with the greatest range tended to have the larger bootstrap estimates of standard error. This is intuitive and in keeping with the results in the correlation example above.

Confidence Intervals for the Median

Simple percentile confidence intervals can be calculated for the median in the same manner as those calculated for the correlation coefficient above. The results are as follows:

Table 5

Sample Size: 15

Trial Number	5th Percentile	10th Percentile	90th Percentile	95th Percentile
1	1.217	1.252	1.478	1.531
2	1.229	1.326	1.567	1.573
3	1.223	1.254	1.511	1.683
4	1.373	1.373	1.499	1.499
5	1.258	1.352	1.635	1.703
6	1.373	1.450	1.525	1.531
7	1.352	1.447	1.500	1.524
8	1.126	1.217	1.373	1.510
9	1.305	1.352	1.706	1.706
10	1.126	1.126	2.102	2.102

The true median falls in all the 5th–95th percentile intervals and in seven of the 10th–90th percentile intervals.

Regression and the Bootstrap

This example demonstrates the use of a bootstrap procedure for estimating the standard errors of regression coefficients. In the example, the OLS fitting criteria are used but the same bootstrap principles could be applied to cases where the fitting criteria are more complex.

Suppose the following data on movements in petroleum prices and transportation companies share prices is obtained and the researcher wishes to fit a simple linear regression model.

Petroleum Price Movements	Transportation Share Price Movements
-9.89%	20.74%
19.27%	14.14%
13.89%	3.26%
8.37%	20.08%
11.14%	19.64%
-5.41%	14.18%
20.58%	6.35%
8.21%	-1.73%
22.59%	-8.03%
10.26%	11.19%
12.64%	9.99%

It seems reasonable that an increase in petrol prices might tend to decrease the share price of transport companies. The linear model is assumed for the purposes of this example.

The bootstrap is carried out in this case by resampling rows from the above data matrix of price changes to generate the new bootstrapped data matrix of price changes. The OLS regression coefficients for this bootstrapped (resampled) matrix are calculated. Repeating this step a large number of times produces a large number of bootstrap estimates for the constant and slope regression coefficients. From these, the standard errors for the regression coefficients can be estimated using algorithm 1. The entire bootstrap procedure was performed using a spreadsheet. A macro was written which would read in the raw price/share price information and which would automatically resample from this information to create 100 bootstrapped data matrices. The macro then calculated the regression coefficients for each of these bootstrapped data matrices and from this information calculated the bootstrap estimates of standard error.

In this example, we can analytically calculate the values for the standard errors if we make the usual OLS assumptions. This enables the bootstrapped values to be compared with the theoretical values to determine whether the bootstrap values seem accurate. (Of course, the accuracy of the supposed 'true' values is dependent on whether the OLS assumptions are met in this case).

The procedure was performed for resample sizes of 10, 20, 50 and 100. For each of these resample sizes, several bootstrap trials of differing length were performed. This enables a number of comparisons to be made. First, what is the effect of varying the number of resamples in bootstrap trials of the same length? Second, what is the effect of varying the number of bootstrap trials while holding the number of resamples chosen constant? Finally, how close are the bootstrap estimates of standard error to the true standard errors for the regression coefficients in each case?

The results of the average of the first 10 bootstrap trials are:

Table 6

Number of Resamples	**Bootstrap Estimate of Constant**	**Bootstrap Estimate of Slope**	**Bootstrap Estimate of SE of Constant**	**Bootstrap Estimate of SE of Slope**
10	0.157584	-0.51868	0.033696	0.229753
20	0.155388	-0.51013	0.038963	0.275954
50	0.155411	-0.52127	0.039005	0.270142
100	0.15466	-0.51398	0.038448	0.266354
OLS values	0.15121	-0.50623	0.03597	0.257179

The results of the average of the first 100 bootstrap trials are:

Table 7

Number of Resamples	**Bootstrap Estimate of Constant**	**Bootstrap Estimate of Slope**	**Bootstrap Estimate of SE of Constant**	**Bootstrap Estimate of SE of Slope**
10	0.156614	-0.5345	0.03806	0.26839
20	0.154687	-0.51175	0.03902	0.27962
50	0.154274	-0.51257	0.03847	0.27426
100	0.15434	-0.51352	0.03824	0.27234
OLS values	0.15121	-0.50623	0.03597	0.25717

Examining the two tables produces some interesting results. The results are variable for the cases where the number of resamples is small (10 or 20) but the general trend within each individual table is clear once the number of resamples gets larger.

Increasing the number of resamples seems to improve the accuracy of the bootstrap estimates of standard error. Intuitively, this is not surprising, given the dilution of the weight attached to an unrepresentative resample as the number of resamples increases. Examining the effect of increasing the number of bootstrap trials over which the results are averaged gives less clear results. Increasing the number of trials the results are averaged over from 10 to 100 would be expected to improve the accuracy of the bootstrap estimate. This has not clearly happened in this case.

However, it should be pointed out that the results obtained above indicate reasonably rapid convergence to the true standard errors. **Table 8** below highlights the percentage difference between the bootstrap estimates in each case and the true values.

Table 8

Number of Bootstrap Trials Averaged Over	Percentage Difference Between Bootstrap Estimate of SE of Constant and OLS Value	Percentage Difference Between Bootstrap Estimate of SE of Slope and OLS Value
10	7.40%	3.57%
100	6.31%	5.90%

This table highlights the performance of the bootstrap. Even in the worst case, its estimate of standard error was within 7.40% of the OLS value despite the fact that it was not based on any of the assumptions underlying the OLS regression model.

CONCLUSIONS

The above examples highlight some of the advantages and disadvantages of using bootstrap techniques. The primary advantage is the lessening of the researcher's dependence on traditional statistical techniques which are sometimes used due to their mathematical tractability rather than their exact fit to the task at hand.

The ultimate goal of bootstrap techniques is to free the researcher from the need to use overtly complex mathematical theory along with its attendant assumptions when making inferences and to encourage the researcher to use the most appropriate statistical tool rather than the statistical tool which has a simple mathematical form. The substitution of increasing amounts of computation for mathematical theory is becoming more cost-efficient as the cost of computing power falls and the cost of developing complex mathematical theory increases.

Bootstrapping is not without its limitations. It only addresses the analysis of the data and cannot compensate for poor data collection or poor interpretation of the results of the analysis. Perhaps a more fundamental drawback is the lack of developed statistical theory supporting the technique in all possible applications. The bootstrap does rely on mathematical underpinnings but these are hidden from all but an in-depth analysis of the technique. Much work has been done on verification of the technique and it has been shown to work in a wide number of settings (Hjorth,1994).

Apart from this issue, there are still unresolved issues in the practice of bootstrapping. How many resamples should be taken from a sample to construct an estimate of standard error, bias or a confidence interval? It appears there is no simple answer as it depends on the characteristics of the estimator being used and the data set at hand. What is the effect of taking differing sample sizes on the accuracy of the bootstrap estimates? It appears intuitive that no data analysis technique can compensate for a lack of data to analyse.

The techniques do provide an interesting contrast to classical methods and do appear to indicate that there is information within the sample that classical techniques ignore.

It would seen reasonable that they are a useful tool for the researcher to use when classical methods make assumptions that do not seem justified by the data under analysis.

Bootstrapping techniques are already in widespread use in many statistical applications. As yet, they have not been extensively applied in accounting research. Clearly, research quality can only be enhanced by use of the best research tools.

SOFTWARE RESOURCES FOR IMPLEMENTATION OF BOOTSTRAP PROCEDURES

The three examples above were implemented using an ordinary spreadsheet. However, this approach is not practical when dealing with very large data sets or when the researcher wishes to use statistics for which the spreadsheet does not have built-in functions.

Two basic approaches can be taken in the implementation of bootstrap techniques when the limits of spreadsheets are exceeded. Bootstrap procedures can be applied using specialist software packages designed for this purpose. Resampling Stats is one of the better known of these packages and is available from Resampling Stats, 612 N. Jackson Street, Arlington, VA 22201, USA. A trial version can be downloaded from the company's web-site (http://www.statistics.com/software.html) for test purposes. This package provides good facilities for resampling from a data set and enables the user to calculate the more common statistics of interest. However, the package will not suffice for the researcher who wishes to bootstrap complex, non-standard statistics.

More complex implementations of bootstrapping techniques are generally specifically designed for the task at hand. The bootstrapping algorithms outlined in the paper can be easily implemented in any computer language. The researcher needs to ensure only that the language/compiler uses a good quality random number generator, as flaws in this will impact on the randomness of the resampling procedure. Using a computer language with built-in mathematical and statistical functions such as Matlab, Stata, Mathematica, S or S-Plus will greatly simplify the programming requirements and the implementation of bootstrap procedures will involve little more than the writing of an iterative routine to resample with replacement from the data and a calculation of the statistic of interest. The latest versions of Stata and Matlab (Statistics Toolbox) contain some built-in bootstrap functions. An extensive set of routines to carry out bootstrapping procedures in S and S-Plus are available from the statistics archive at Carnegie-Mellon University and can be downloaded by FTP access from lib.stat.cmu.edu using the login id ‘statlib’ (file name – bootstrap.funs).

SUGGESTED FURTHER READING

Efron and Tibshirani (1993) provide a very readable account of bootstrapping concepts and provide many examples using real data. It is probably the best introductory book on the subject. The bootstrap was popularised in Efron (1979) but this article drew on earlier work by Hartigan (1971). The development of the asymptotic theory of the bootstrap (explaining why it works) commenced in Bickel and Freedman (1981). Extensions of the theory to regression cases are discussed in Freedman (1981) and Holm (1993). Currently, much work is being performed concerning the applicability of bootstrap concepts in differing areas of statistics and many other references exist. Lepage and Billard (1992) provides an excellent source of such material. Hjorth (1994) provides a good source of up-to-date references for the bootstrap and explains the mathematical underpinnings of the techniques in some detail.

REFERENCES

Bickel, Peter J. and Freedman, David A. (1981). 'Some Asymptotic Theory for the Bootstrap', *Annals of Statistics*, 9, pp. 1196–1217.

Efron, Bradley (1979). 'Bootstrap Methods: Another Look at the Jackknife', *Annals of Statistics*, 7, pp. 1–26.

Efron, Bradley and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall.

Freedman, David A. (1981). 'Bootstrapping Regression Models', *Annals of Statistics*, 9, pp. 1218–1228.

Hartigan, John A. (1971). 'Error Analysis by Replaced Samples, *Jnl. Royal Statistical Society, Series B*, Vol. 33, pp. 98–101.

Hjorth, J. S. Urban (1994). *Computer Intensive Statistical Methods*, Chapman and Hall.

Holm, Sture (1993). 'Abstract Bootstrap Confidence Intervals in Linear Models', *Scandinavian Jnl. of Statistics*, pp. 157–170.

Lepage, Raoul and Billard, Lynne (1992). *Exploring the Limits of the Bootstrap*, John Wiley & Sons, Inc.

O'Brien, Frank J. and McCallig, John (1995). 'Asset revaluations by Irish Companies', *Accountancy Ireland*, April, 27(2), pp. 10–12.