The Bootstrap: What the Government Auditor Should Know

Alan H. Kvanli
College of Business Administration
University of North Texas
Denton, TX 76203
940-565-3115

Robert Schauer
Computer Audit Specialist
Washington State Department of Revenue
8008 NE Fourth Plain Blvd, Suite 320
Vancouver, WA 98665-1648
360-260-6167

Introduction

Many government auditors routinely use random samples to estimate some unknown amount from a population of interest. These populations frequently have a low occurrence rate of these items: such as incorrectly booked amounts, over-deductions, underpaid tax amounts, or overpaid claims. In these cases, the samples may have only a few occurrences of the audit item of interest and many examples of correctly handled transactions. In other words, errors typically will be rare and the sample will contain many zeros (correct items) and a few nonzeros (incorrect items).

In addition to discovering errors, typically auditors will also want to estimate the total error amount. This can be done using traditional statistical techniques that make certain assumptions. The auditor, with a given confidence level, can compute a *confidence interval* around the projected total. This confidence interval measures the precision of the projected total. When the confidence interval is too wide, indicating that the projected total is tenuous, and increasing sample size is not an option, the auditor frequently will base his adjustment on the *lower limit* of the confidence interval (interval adjustment). It is common for government agencies to base their adjustments on the lower limit rather than the projected total (the projected total is referred to as the *point estimate*). For example, if a confidence interval is determined to be $80,000 ± $20,000, the point estimate would be $80,000 and the lower limit would be $60,000.

It has been well known for many years that the traditional statistical techniques typically do not provide reliable confidence intervals when applied to typical audit populations encountered in government audits.[1] For sample sizes normally used by government auditors, these techniques usually provide a lower limit that is too low, or in other words, too conservative. Since no other practical alternatives have emerged up until now, the accepted procedure has been to use statistical packages that compute a confidence interval based on traditional statistical methods. The excessively conservative nature of this method has simply been accepted.

This article is intended to introduce government auditors to an alternate method in computing the lower limit that is more appropriate provided two ordinary conditions are met. First, the samples contain only a few occurrences of the item of interest (less than thirty nonzeros in the sample). This statement applies to any sample size. And second, the number of items sampled as a percentage of the population, or *sampling percentage*, is low, say less than 10%. This alternate method is called the *bootstrap*. The bootstrap procedure will provide more realistic or reliable, that is not overly conservative, lower limits.[2] We have some real life examples and computer simulations that illustrate how the bootstrap provides a higher lower limit when compared to traditional statistical computations. Use of the bootstrap has become a viable methodology only through the emergence of the computer, given its intensive computational requirements, as we will demonstrate.

Traditional Statistical Appraisals

Traditional statistical methods can be used to compute precision. These computations, easily done by a computer, are too complex to fully explain here, so we will only give a brief explanation. For example, in evaluating an audit sample, say we are 90% confident that the true unknown population amount is within $20,000 of the point estimate of $80,000. There are various methods common in the auditing profession used to compute the point estimate.[3] The lower and upper limits of the confidence interval here are $80,000 ± $20,000; that is $60,000 and $100,000, respectively. It can also be said, using the same calculations, that we are 95% confident that the true population amount is at least $60,000 (the lower limit of the 90% confidence interval).[4] If the confidence interval procedure is reliable, we can expect that 5% of the time this statement will be incorrect if sample after sample is drawn.

But we know that in samples where the sampling percentage is under 10% and fewer than thirty errors are observed, traditional computations of the lower limit will be too conservative. In other words, using the statistical tables to determine a 90% confidence interval, the lower limit statement will be incorrect far fewer than 5% of the time. These tables are based on some assumptions that possibly cannot be applied to audit populations and sample sizes typically found in the government auditing profession. Mindful of this, government agencies (such as the Internal Revenue Service, the Department of Health & Human Services, and many state Revenue Departments) have established policies about the minimum number of occurrences (nonzeros) in the sample that will allow the auditor to make an adjustment based on a projected amount from the sample. These minimums (denoted as NZ) vary, but numbers such as NZ equal to three, five, six, or ten can be found to be the specified minimums. We will also address the need for these minimums. We have found that auditors across the country are interested in what number should be used as a minimum and we have some interesting insights on this topic as a sidelight to the issue of how to appraise audit samples.

A closer look at audit populations reveals that we really have a *mixture* of two populations. We have many non-errors (zeros) and a few errors (nonzeros). If the population error rate is say 10% or under, samples of 200 or less can be expected to have less than thirty errors in the sample. We can assert that audit populations frequently have error rates of 10% or less, and that auditors may be reluctant or unable (due to limited audit resources) to use very large sample sizes.[5] Yet traditional evaluation techniques are routinely applied to these samples when the bootstrap procedure may be a more reliable choice.

The Bootstrap

Traditional statistical techniques make assumptions that are generally valid, but in the case of low error rate populations, may not be appropriate. The traditional approach (a *parametric* approach) assumes that the distribution of sample means is approximately *normal*. In low error rate populations, the distribution may be *skewed* unless a very large sample is taken. The bootstrap typically is not affected by these assumptions.[6] More importantly, as noted above, in nearly all cases the bootstrap lower limit will be much less conservative and more reliable. As the following discussion will make clear, when about thirty or more errors are observed in the sample, the bootstrap and traditional methods will generally provide lower confidence limits that are nearly the same.

Bootstrap procedures fall into the category of *resampling methods*. That is, the sample results are resampled many times over. The fundamental assumption when using bootstrapping is that the sample is representative of the underlying population.

The bootstrap procedure begins in the usual way by obtaining a *simple random sample* from the population, say of size 200 items. An error value (zero or nonzero) is established for each sample item. To derive a bootstrap confidence interval, the following steps are carried out:

1) Obtain say, 1000 new samples by sampling the original sample of 200 errors *with replacement*. In sampling *with replacement*, the sampling units are returned to the population (in this case the original sample), and may be selected into the sample again. Traditionally, auditors will select the initial sample *without replacement*. The number of resamples can vary, but the more resamples, the better. The size of each resample should be the same as the initial sample.

2) For each of the 1000 resamples, derive the *sample mean* (average error dollar amount). The sample mean is multiplied by the population count to arrive at an estimate of the population total.

3) Arrange these 1000 population total error estimates in order, from smallest to largest. One procedure for deriving a 90% confidence interval is to find the 5th and 95th percentiles of these 1000 estimates; that is, the 50th estimate (the lower confidence limit) and the 950th estimate (the upper confidence limit). Here, 50 is (.05)(number of resamples) = (.05)(1000) and 950 is (.95)(number of resamples) = (.95)(1000).

Clearly, to be practical, performing 1000 or more resamples will require a computer.[7]

The procedure described in step three is referred to as the bootstrap confidence interval using the *empirical percentiles*. An improvement over this procedure calculates the *bias-corrected and adjusted* (*BCa*) percentiles.[8] For example, the *BCa* procedure for a particular sample may be to use the 11.0th and the 98.4th percentiles, that is, the 110th bootstrap estimate (the lower confidence limit) and 984th bootstrap estimate (the upper confidence limit) when all the resample estimates are arranged smallest to largest. The *BCa* percentiles vary from sample to sample, require more computation, and generally are much more accurate. For either procedure, at least 1000 resamples are recommended.

An Example

A simple random sample of 200 items ($n = 200$) is obtained from a population of 5000. Note that the sample size is 4% of the population size. Through normal audit procedures, the auditor determined that the sample contained 191 zero values and nine nonzero errors with the following amounts: $104.23, $236.71, $250.56, $309.82, $324.15, $401.33, $653.58, $1114.60, and $1824.92.

From the example, we know that the point estimate of the total error is $130,500.[9] Also, we can safely assume that the population has a low error rate since the sample error rate is 9/200 = 4.5%. Using traditional appraisal techniques, the 90% confidence interval has a lower and upper limit of $36,600 and $224,400, respectively.[10]

A bootstrap confidence interval is obtained by sampling the initial sample say, 1000 times with replacement and deriving the estimate of the population total error for each sample.[11] To illustrate, the first resample might contain 192 zero error values and the following eight (0+0+1+0+2+1+1+2+1) nonzero error values selected from the original sample:

# of Times Selected →	0	0	1	0	2	1	1	2	1
Value →	140.23	236.71	250.56	309.82	324.15	401.33	653.58	1114.60	1824.92

The corresponding bootstrap estimate of the population total error amount would be $150,200.[12] If we want 1000 resamples, then this procedure would be repeated 999 additional times and the resulting 1000 bootstrap estimates would be arranged from smallest to largest. Confidence intervals can then be derived using both the Empirical and *BCa* methods.[13]

Using Empirical Percentiles

After arranging the 1000 bootstrap estimates from smallest to largest, the 50th and 950th estimates were $47,754 and $236,954. Consequently, the 90% confidence interval for the total population error amount is from $47,754 to $236,954.

Using *BCa* Percentiles

The *BCa* percentiles for this sample were 13.256% and 99.068%. The corresponding confidence interval is found by finding the (.13256)(1000) = 133rd bootstrap estimate and the (.99068)(1000) = 991st bootstrap estimate (when all estimates are arranged smallest to largest). These were $67,838 (lower limit) and $288,780 (upper limit).

Observations

1. Both bootstrap methods produced higher confidence limits than those obtained using the traditional method.
2. The lower limit produced by the bootstrap *BCa* method is 85% higher than that produced by the traditional approach ($36,600 versus $67,838). This is a typical result when using this bootstrap procedure on audit samples.

Bootstrap Simulations

A number of computer simulations were generated to gain insight as to the effects of audit sample sizes and error rates on the performance of the bootstrap confidence interval procedure versus the traditional method. These simulations show that bootstrap confidence intervals are more reliable.[14]

The Error Populations

Error rates of 2%, 5%, and 10% were considered. Three populations of error values, each having 5000 units, were created using the three error rates. Population #1 contained 4900 zero values and 100 nonzero values. The mean error value is $16 and the total error amount is $80,000. Similarly, error population #2 contains 4750 zero values, 250 nonzero values, with a mean error value of $40 and a total value of $200,000. Error population #3 consists of 4500 zero values, 500 nonzero values, has a mean error value of $80 and a total value of $400,000.[15]

The Sample Sizes

Sample sizes of 200, 300, 400, 500, and 600 were considered. The expected number of nonzero errors in the sample (*EE*) ranged from four (2% x 200), to sixty (10% x 600). The results show that the bootstrap procedure performed quite well even when a very small number of nonzero errors is expected in the sample; in particular, when EE is less than10.

Minimum Number of Nonzero Errors in Each Sample

In order to project a sample result to the entire population, government auditors may require a minimum number of nonzero errors. For example, federal auditors within the Department of Health and Human Services require at least six nonzero errors. Otherwise the sample results are not projected. For this study, the minimum number of nonzeros (*NZ*) was set at three, six, and ten.

The Simulation Runs

For each sample size within each population, 5000 simple random samples were created (for a total of 75,000 initial samples). For each generated sample, 1000 bootstrap resamples were created. Subsequently, for each of the 5000 created samples, a bootstrap confidence interval using the *BCa* bootstrap procedure was generated.

For simulations on each population, the following information was captured:

- A The percentage of the samples that contained the minimum number of nonzero errors (NZ). Any samples that did not were discarded and ignored.
- B A 90% confidence interval using traditional procedures was derived for each sample if the sample contained the minimum number of nonzero errors (*NZ*). The percentage of computed lower limits that were more than the actual population error is noted (this should be about 5% of the time).
- C A 90% confidence interval using the *BCa* bootstrap procedure was derived for each sample if the sample contained the minimum number of nonzero errors. The percentage of computed lower limits that were more than the actual population mean is reported (again, the closer this number is to 5%, the more reliable the procedure).
- D The percentage of time the traditional 90% confidence interval contained the population mean (ideally, this should be 90% of the time).
- E The percentage of the 90% *BCa* bootstrap confidence intervals containing the population mean (again, the closer this number is to 90%, the more reliable the procedure).
- F The average lower limit using the traditional approach.
- G The average lower limit using the *BCa* bootstrap approach.

Examining the Results

The simulation results for Populations #1, #2, and #3 are summarized in Tables 1, 2, and 3 respectively. The columns of the three tables report the information noted above (A – G). If a sample fails to have at least the minimum number of nonzero errors (NZ) it was discarded and confidence intervals were not derived. If more than 50% of the 5000 generated samples failed to have the minimum number of nonzero errors (summarized in column A), the results are not reported (any conclusions drawn from the remaining samples would be highly suspect due to the limited number of usable samples).

Column B. These values clearly demonstrate how conservative the lower limit is using the traditional confidence interval, because these values should ideally be around 5%.

Column C. These values are the percentage of samples for which the lower limit exceeded the population mean using the *BCa* bootstrap procedure and (as for column B) should be about 5%. The values in the three tables are roughly 3.5%

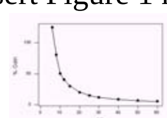
to 4%. The lower limit for the bootstrap interval is still conservative but much closer to 5% than that produced by the traditional procedure.

Columns D and E. These two columns contain the percentage of the usable samples that contained the population mean, and should ideally be about 90%. The actual percentage of time that the intervals contained the mean is referred to as the <u>coverage</u> provided by this procedure. Generally, the bootstrap procedure provided slightly better coverage than the traditional method.

Columns F and G. Columns F and G contain the average of the lower limits (in thousands of dollars) for each of the two confidence interval procedures. The far-right column contains the increase in the lower limit using the *BCa* bootstrap interval. As expected, on average, the bootstrap always provided for higher lower limits. In many instances, the increase was dramatic.

Notice that the percent gain in the lower limit values appears to be driven by the expected number of nonzero errors (*EE*). The gain in the bootstrap lower limit over the traditional lower limit appears to be highly dependent upon *EE* as Figure 1 clearly shows. Figure 1 is a plot of *EE* versus the percent gain and shows that the expected gain is related to the expected number of errors and after about 30 errors, the curve (or benefit to using the bootstrap) flattens out.

Insert Figure 1 here



What is Important. We feel the auditor should know that a dramatic increase in the lower limit would be realized when using the *BCa* bootstrap procedure in low error rate situations, as shown in the “*% Incr*.” column of the tables. Further, this increased amount is more reliable than a lower limit derived using the traditional approach.

<u>A Look at the Minimum Number of Nonzero Errors</u>

It appears that the minimum number of errors required to project a sample result is not a critical concern. The results obtained by requiring three nonzero errors were just as reliable as those obtained when requiring six or ten nonzero errors. This conclusion also applies to the highly conservative intervals derived using the traditional approach. Requiring ten nonzero errors would not be advisable since (1) there is a much higher chance of obtaining a sample not containing the minimum number of nonzero errors and (2) the confidence interval coverage (in column E) is often too large.

Bootstraps on Actual Samples

We have taken random samples found in actual audit situations and applied the bootstrap procedure. These are summarized in Table 4. Similar to the simulations, the bootstrap *BCa* lower limits are significantly larger when compared to the traditional approach. However, the percent gain in the lower limit is not as dependent on the expected number of nonzero errors (*EE*) as in the simulations (Figure 1), but still appears exponentially related to the number of nonzero errors found in the sample.

Summary and Further Research

The primary result to emerge from this study is that the auditing agency can expect a much larger amount when requesting an adjustment based on the lower limit of a bootstrap confidence interval. Further, the confidence intervals are more reliable when compared to traditional methods.

Statisticians have long been aware of the problems associated with appraising low error rate populations with traditional methods. The simulations and actual data have demonstrated the dramatic improvement offered by using a bootstrap procedure. Of the two bootstrap procedures introduced, the *BCa* method will generally provide more reliable results and is the procedure recommended for appraising low error rate samples. The lower limit provided by the *BCa* methodology is still on the conservative side but far less conservative when compared to the traditional approach.

The data, research and examples used in this article have utilized difference estimation, but we believe the same results may apply to other estimation methodologies such as ratio and regression estimation – but this will take further research.

Results for 2% Error Rate (Population #1)

Sample Size (*n*)	*NZ*	A	B	C	D	E	F	G	% Incr.
***n* = 200** (*EE* = 4)	3	23.6	.3	3.7	88.6	90.1	4.9	30.9	**530.6**
	6	79.0	X	X	X	X	X	X	X
	10	99.3	X	X	X	X	X	X	X
***n* = 300** (*EE* = 6)	3	5.9	.6	3.9	83.9	86.1	13.4	33.1	**147.6**
	6	44.2	.5	3.6	96.4	95.1	21.9	44.2	**101.8**
	10	92.7	X	X	X	X	X	X	X
***n* = 400** (*EE* = 8)	3	1.2	.6	3.4	82.3	86.0	19.1	35.4	**85.6**
	6	18.4	.7	3.9	91.9	91.8	23.4	40.6	**73.9**
	10	72.7	X	X	X	X	X	X	X
***n* = 500** (*EE* = 10)	3	.2	.5	3.2	84.5	88.6	24.6	38.6	**57.2**
	6	5.6	.7	3.6	87.9	90.6	26.9	41.3	**53.3**
	10	45.0	.7	3.7	97.5	95.3	35.2	50.4	**43.0**
***n* = 600** (*EE* = 12)	3	0	.8	3.5	85.9	89.3	29.2	41.7	**42.6**
	6	1.6	.7	3.6	86.9	89.9	29.6	42.1	**42.3**
	10	21.7	.7	3.5	94.6	94.7	33.9	46.6	**37.3**

Table 1

Notes to Tables 1 – 3:

- *EE* is the expected number of errors equal to the error rate times the sample size.
- *NZ* is the minimum number of nonzero errors required to project the results. If the minimum is not met, the sample is discarded.
- A: Percentage of samples discarded because *NZ* was not met. If more than 50% of the simulations have been discarded, the information in columns B-G is not reported (labeled “X”).
- B: Percentage of traditional confidence intervals where the lower limit is greater than the total population error amount. This should be at or near 5%.
- C: Percentage of *BCa* bootstrap confidence intervals where the lower limit greater than the total population error amount. This should be at or near 5%.
- D: Percentage of traditional confidence intervals containing the total population error amount. This should be at or near 90%.
- E: Percentage of *BCa* bootstrap confidence intervals containing the total population error amount. This should be at or near 90%.
- F: Average lower limit for total error amount using traditional confidence intervals (x $1,000).
- G: Average lower limit for total error amount using BCa bootstrap confidence intervals (x $1,000).

Results for 5% Error Rate (Population #2)

Sample Size (*n*)	*NZ*	A	B	C	D	E	F	G	% Incr.
***n* = 200** (*EE* = 10)	3	.2	.8	4.2	84.2	87.2	63.0	98.1	**55.7**
	6	5.8	.9	4.2	87.8	89.3	67.0	102.6	**53.1**
	10	44.9	.9	4.1	97.4	94.9	89.6	126.9	**41.6**
***n* = 300** (*EE* = 15)	3	0	.8	3.8	86.5	88.9	85.6	111.4	**30.1**
	6	.3	1.2	4.2	87.1	89.2	86.4	112.3	**30.0**
	10	5.6	1.2	3.8	89.9	91.4	90.0	116.1	**29.0**
***n* = 400** (*EE* = 20)	3	0	1.5	4.4	87.4	89.5	100.6	121.0	**20.3**
	6	0	1.2	4.1	87.5	89.3	100.4	120.9	**20.4**
	10	.4	1.4	3.7	87.5	89.6	100.1	120.7	**20.6**
***n* = 500** (*EE* = 25)	3	0	1.7	4.1	88.6	90.4	111.1	128.0	**15.2**
	6	0	1.2	3.9	88.0	90.0	109.4	126.2	**15.4**
	10	0	1.4	3.4	88.2	90.3	109.4	126.4	**15.6**
***n* = 600** (*EE* = 30)	3	0	1.6	3.8	89.3	90.6	117.9	132.3	**12.2**
	6	0	1.4	3.5	90.7	91.6	118.0	132.7	**12.5**
	10	0	1.6	3.9	90.0	91.1	117.6	132.1	**12.3**

Table 2

Results for 10% Error Rate (Population #3)

Sample Size (*n*)	*NZ*	A	B	C	D	E	F	G	% Incr.
***n* = 200** (*EE* = 20)	3	0	1.5	4.4	87.6	89.2	204.2	244.9	**20.0**
	6	0	1.6	4.4	86.8	88.7	203.8	244.7	**20.1**
	10	.3	1.3	4.2	87.0	88.4	202.6	243.3	**20.1**
***n* = 300** (*EE* = 30)	3	0	1.6	4.4	88.1	89.1	237.2	266.1	**12.2**
	6	0	1.5	4.5	89.6	90.0	240.1	269.2	**12.1**
	10	0	1.8	4.0	88.6	90.3	238.4	267.2	**12.1**
***n* = 400** (*EE* = 40)	3	0	2.2	4.5	89.0	89.7	259.7	282.4	**8.7**
	6	0	1.8	4.0	88.9	89.9	257.7	280.3	**8.8**
	10	0	1.7	3.9	89.4	90.4	257.7	280.4	**8.8**
***n* = 500** (*EE* = 50)	3	0	2.1	4.4	90.0	90.3	273.0	291.6	**6.8**
	6	0	1.9	3.7	90.5	91.3	273.4	292.0	**6.8**
	10	0	1.8	4.2	91.0	91.1	273.6	292.3	**6.8**
***n* = 600** (*EE* = 60)	3	0	2.1	4.1	90.9	91.8	284.0	299.8	**5.6**
	6	0	2.0	3.7	90.9	91.2	283.2	298.9	**5.5**
	10	0	2.3	4.1	91.0	91.4	285.4	301.2	**5.5**

Table 3

Results from Actual Samples

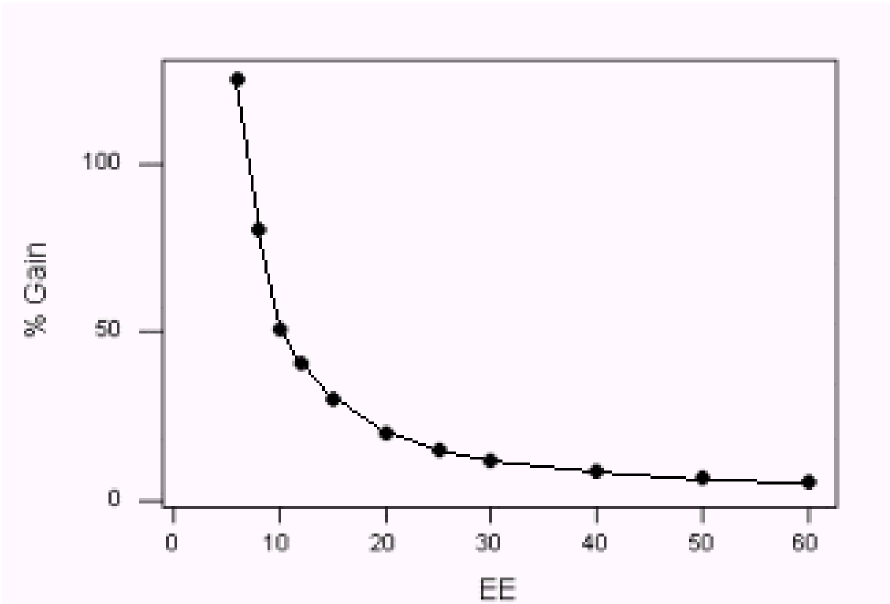
H Sample	I Sample Size (n)	J Nonzero Errors	K Traditional Lower Limit 95% Conf.	L Bootstrap *BCa* Lower Limit 95% Conf.	% Incr.
1	365	3	-1,273,609	104,791	**108.2**
2	300	6	180,935	277,856	**53.6**
3	365	8	1,611,222	6,178,412	**283.5**
4	365	10	1,085,248	7,007,027	**547.7**
5	247	10	16,923	19,976	**18.0**
6	300	11	725,080	875,401	**20.7**
7	120	12	11,687	13,359	**14.3**
8	365	13	59,614	75,371	**26.4**
9	300	13	167,383	182,820	**8.9**
10	300	15	187,831	208,231	**10.9**
11	300	18	664,749	774,963	**16.6**
12	300	30	411,956	425,947	**3.4**

Table 4

Notes to Table 4:

- All samples have come from simple random samples from various sales and use tax audits done by the Washington State Department of Revenue within the last year.
- The lower limits are 95% one-sided using a difference point estimator.
- In each case, the *sampling percentage* is less than 10%.
- For sample 1, the percent increase was computed using |(L - K) / K| x 100. For samples 2 – 12 the increase was computed in the same manner as used for tables 1-3, that is (L - K) / K x 100.

Benefit Using the Bootstrap



Percent Gain in the Lower Limit as a Function of the Expected Number of Nonzero Errors (*EE*)

Figure 1

Endnotes

[1] Panel on Nonstandard Mixture of Populations, *Statistical Models and Analysis in Auditing: A Study of Statistical Models and Methods for Analyzing Nonstandard Mixtures of Distributions in Auditing*, Washington, D.C. National Academy press, 1988.

[2] Upper limits, using traditional techniques are also too low (not conservative enough). Although upper limits are not dealt with in this article, the bootstrap also provides for more reliable upper limits.

[3] Some of the methods include *difference estimation*, *ratio estimation* or *regression estimation*.

[4] For this illustration, \$60,000 can be interpreted as a 95% lower limit (bound) of a one-sided confidence interval (we are 95% confident that the true population amount is at least \$60,000) or as the lower limit of a 90% two-sided confidence interval (we are 90% confident that the true population amount is between \$60,000 and \$100,000). In general, the lower limit of an A% two-sided interval can be interpreted as a

$(50+\frac{A}{2})\%$ lower bound.

[5] The question of how large of a sample is required using traditional evaluation approaches has been addressed in several studies:

- John Neter and James Loebbecke, *Behavior of Major Statistical Estimations in Sampling Accounting Populations: An Empirical Study*, New York, American Institute of Certified Public Accountants, 1975.
- Alan H. Kvanli, Y. K. Shen and L.Y. Deng, *A Construction of Confidence Intervals for the Mean of a Population Containing a Large Number of Zero Values*, Journal of Business & Economics, Vol. 16, No. 3, 1998, pp. 362-8.

[6] In the bootstrap it is not required to assume that the distribution of sample means is normal. Since this procedure makes no assumptions regarding the distribution of sample means, it is called a *nonparametric* method of deriving a confidence interval. A recent research article recognized this advantage, and recommended the bootstrap to audit populations:

- Gary C. Biddle, Carol M. Bruton, and Andrew F. Siegel, *A Computer-Intensive Methods in Auditing: Bootstrap Difference and Ratio Estimation,* Auditing: A Journal of Practice and Theory, Vol. 9, No. 3, Fall 1990, pp. 92 - 114

[7] For more information on bootstrap procedures, refer to:

- *An Introduction to the Bootstrap* by Bradley Efron and Robert J. Tibshirani (New York: Chapman & Hall/CRC, 1998)
- *Bootstrap Methods and their Application* by A.C. Davison and D.V. Hinkley (New York: Cambridge University press, 1998)

[8] In the textbook by Efron and Tibshirani (mentioned in the previous endnote), they discovered that in actual practice, the percentile confidence interval had to be corrected for two factors, bias and acceleration. As a result, they derived the (generally more reliable) bias-corrected accelerated bootstrap confidence interval, or BCa interval. Refer to this textbook for a more detailed explanation of the BCa procedure.

[9] From the information provided, we can compute the sample mean and the *standard deviation* to be $26.10 and $164.05, respectively. The point estimate is calculated as follows:

$$(5000)(26.10) = \$130{,}500$$

[10] A sample of 200 using 90% confidence interval will have a *t-value* of 1.6525 (from statistical tables). The upper and lower limits are calculated:

$$(5000)(1.6525)\left(\frac{164.05}{\sqrt{200}}\sqrt{\frac{5000-200}{5000}}\right) \approx \$93{,}900$$

$$\text{upper limit} = \$130{,}500 + \$93{,}900 = \$224{,}400$$

$$\text{lower limit} = \$130{,}500 - \$93{,}900 = \$36{,}600$$

[11] Note that one must sample the original sample with replacement to avoid repeatedly obtaining the original sample for each resampling. Bootstrap procedures always involve sampling the original sample with replacement.

[12] The estimate from this resample can be computed:

$$(5000)\left(\frac{(192)(0)+(0)(140.23)+(0)(236.71)+(1)(250.56)+(0)(309.82)+(2)(324.15)+(1)(401.33)+(1)(653.58)+(2)(1114.60)+(1)(1824.92)}{200}\right) = \$150{,}200$$

[13] The BCa confidence intervals in this study were derived using S-Plus. Other statistical packages, such as SAS, SPSS, and MINITAB, to our knowledge, are not able to compute BCa confidence intervals. The authors have an Excel-based template (available upon request) for computing BCa confidence intervals when using a simple random sample.

[14] In a computer simulation, an error population is created with a known mean error value and a known total error amount. Samples are randomly selected from this population and a confidence interval for the total error amount is derived for each sample. If the confidence level is set at 90%, then a *reliable* confidence interval procedure is one for which (approximately) 90% of the generated samples produce a confidence interval that contains the (known) population total error amount. A similar definition applies to a procedure that produces a reliable 95% lower bound, whereby (approximately) 95% of the generated samples produce a lower limit that is less than or equal to the (known) population total error amount.

[15] The nonzero errors are *exponentially* distributed and have a mean and standard deviation of $800.