

Difference Estimation versus Mean per Unit Methods for Skewed Populations: A Simulation Study

John Chantis, Office of Inspector General, Department of Defense, Arlington, VA
Kandasamy Selvavel, Office of Inspector General, Department of Defense, Arlington, VA

ABSTRACT

In most of the financial statements audits, Different Estimation and Mean per Unit (MPU) methods are often used to estimate the error or the audit value of the population. In this research we compare Mean per Unit to Different Estimation methods and study the characteristics of the estimates. In each of these methods we focus our study on Simple Random Sample (SRS) and Stratified Sample designs. Most of the financial data are strongly positive skewed, and our primary interests are only on these types of data. For each case, our simulations consisted of 1,000 replications.

For a fixed sample size, we estimate the total audit value of the population using Difference Estimation and MPU methods. We then calculate the coverage probability and precision for each of these cases. The coverage probability is the probability that the true value is within the confidence interval. We ideally want the coverage probability to be closer to the stated $(1 - \alpha)$ level. The estimator with larger precision is not useful for decision makers or to book the adjustment in financial statements. In addition to the coverage probability and precision we also examine the convergence rate of the estimator to the true value in our simulation study.

Based on the simulation results the coverage probability for the difference estimation method is very low. The same pattern holds for the stratified design. We notice that both the precision and the coverage probability significantly improved in the stratified design. The MPU method provides better coverage probability than the difference method. However the difference method is better than the MPU method in terms of relative precision. These results are consistent for SRS and stratified designs. Differences in coverage probability and relative precision between the two methods narrows for stratified sampling designs.

INTRODUCTION

In financial audits Difference Estimation and Mean Per Unit (MPU) methods are commonly used to estimate the audit value. In the difference estimation method audit value is defined as book value plus error. In the MPU method total audited value is directly estimated from sample audited value.

We get better result in sampling when the population data is normally distributed. The normally distributed data is in the form of the traditional bell shaped curve where the mean and median are equal. But, in practical, this is not the case most of the time. We know empirically that financial data is highly positively skewed. That is, few very large dollar transactions relative to mid to small transactions. The statistical sampling in financial audit requires certain level of confidence. But when the data is highly skewed we are not assured that we achieve that confidence level.

We randomly simulate audit errors based on various error rates for a skewed population. We then randomly select samples from the population for SRS and stratified designs and calculate point estimates, confidence intervals and relative precisions of the estimated audited values for difference and MPU methods. We also calculate and compare the coverage probability with the confidence level to see if we achieve the target value.

PRECISION ANALYSIS

We generated a positively skewed population of size 1500. A 10 percent error rate was assigned to the population and the audited values are calculated by adding the book value to the error. The total error in the population is four percent of the book value. For difference estimation method the total error in the population was calculated and then added to the book value to get the audited value of the population. For mean per unit method the audited value of the population was directly calculated from the sample. We also calculate 90% percent confidence level and the relative precision of the estimator.

Simple Random Sample

Table 1: Precision (Population Size = 1,500, SRS Sample Size = 120, Error Rate = 10% and Total Error = 4% of Book Value)

Sample Size	Precision for Difference Estimation Method (Audit Value = Book Value + Error)	Precision for Mean Per Unit Method (Estimate Audit Value Directly)
120	10%	91%
180	8%	83%
240	7%	77%

We now increase the error rate to 40 percent. Total error is eight percent of the book value. Below is the simulation result of the precision for different sample sizes.

Table 2: Precision (Population Size = 1,500, SRS Sample Size = 120, Error Rate = 40% and Total Error = 8% of Book Value)

Sample Size	Precision for Difference Estimation Method (Audit Value = Book Value + Error)	Precision for Mean Per Unit Method (Estimate Audit Value Directly)
120	145%	92%
180	37%	85%
240	28%	78%

The precision improves for both methods when the sample size increases. The precision for difference method is better than the precision for the mean per unit method for large samples.

Stratified Sample Design

Table 3: Precision (Population Size = 1,500 [5, 48, 255, 535, 657], Stratified Sample Size = 120, Error Rate = 10% and Total Error = 4% of Book Value)

Sample Size	Precision for Difference Estimation Method (Audit Value = Book Value + Error)	Precision for Mean Per Unit Method (Estimate Audit Value Directly)
120 (5, 15, 40, 35, 25)	6%	15%
180 (5, 20, 55, 55, 45)	5%	12%
240 (5, 25, 70, 75, 65)	4%	10%

We now increase the error rate to 40 percent. Total error is eight percent of the book value. Below is the simulation result of precision for various sample sizes.

Table 4: Precision (Population Size = 1,500 [5, 48, 255, 535, 657], Error Rate = 40% and Total Error = 8% of Book Value)

Sample Size	Precision for Difference Estimation Method (Audit Value = Book Value + Error)	Precision for Mean Per Unit Method (Estimate Audit Value Directly)
120 (5, 15, 40, 35, 25)	7%	18%
180 (5, 20, 55, 55, 45)	5%	15%
240 (5, 25, 70, 75, 65)	4%	12%

The estimator with larger precision is not useful for decision makers or to book the adjustment in financial statements. The precision is inversely proportional to the sample size for both methods. The precision for difference method is better than the precision for the mean per unit method. These results are consistent with the results for the SRS method for large samples.

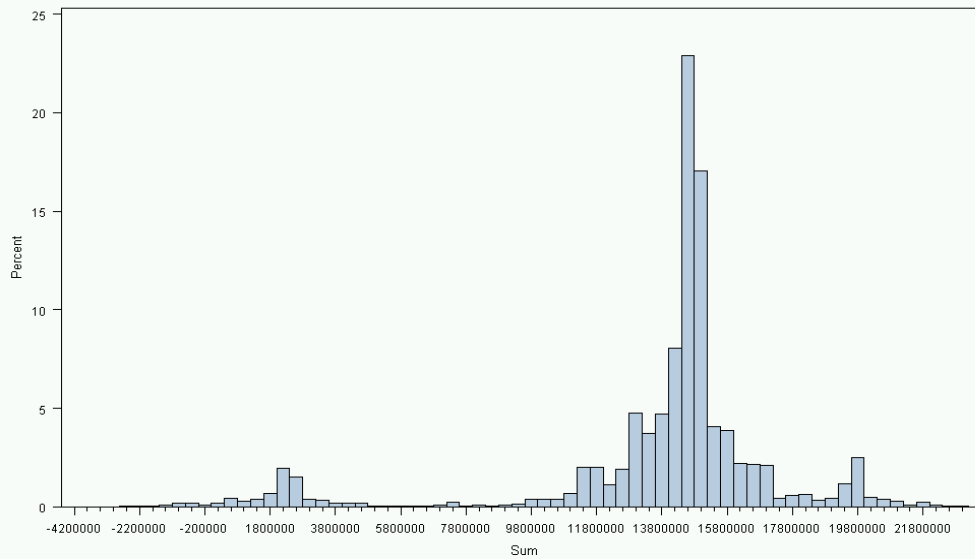
THE SAMPLING DISTRIBUTIONS AND THE CONVERGENCE RATE OF ESTIMATORS

In this section we study the sampling distributions and the convergence rates of MPU and difference estimators for the second population (POP2). We simulate SRS and stratified samples of size 120 and calculate the MPU and the difference estimators. We then draw the histograms for each case to study the distribution. We also calculate the bias of the estimators from the simulation runs and plot the bias along the y-axis and the number replications in the x-axis. The results given herein are for the higher error rate (40%) population.

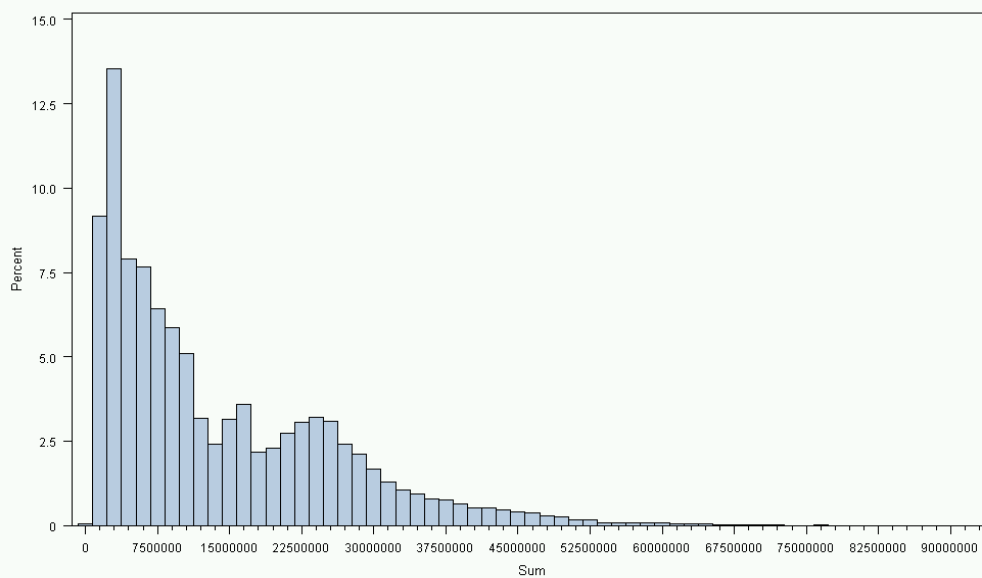
Sampling Distribution

Below are the graphs of sampling distribution and the convergence behavior of the estimators.

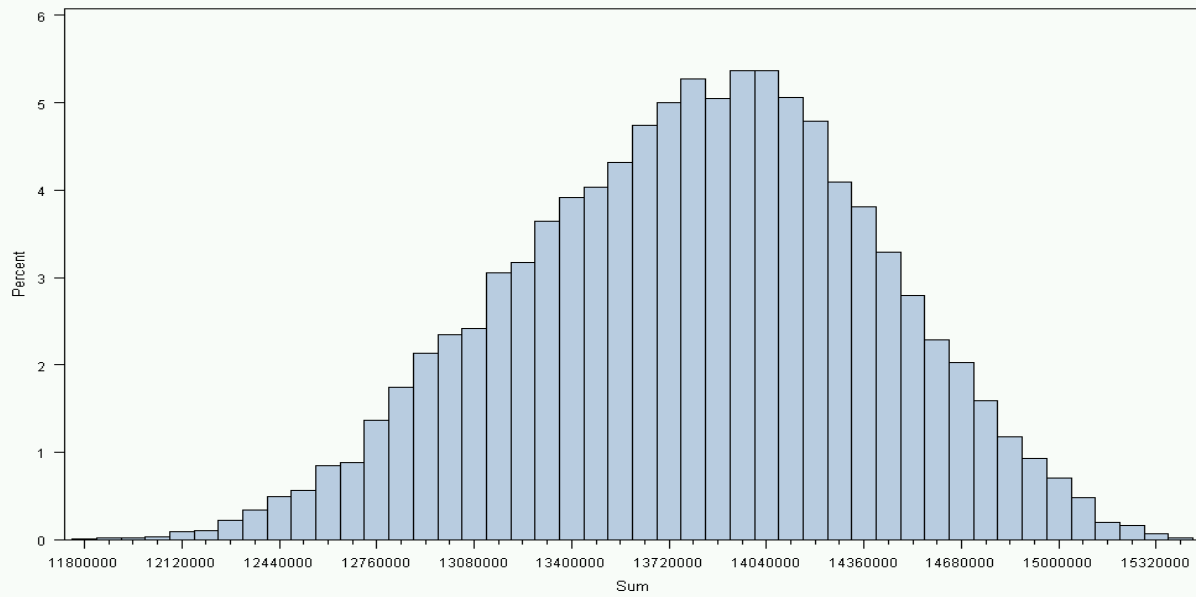
SRS n=120 Diff method, POP2



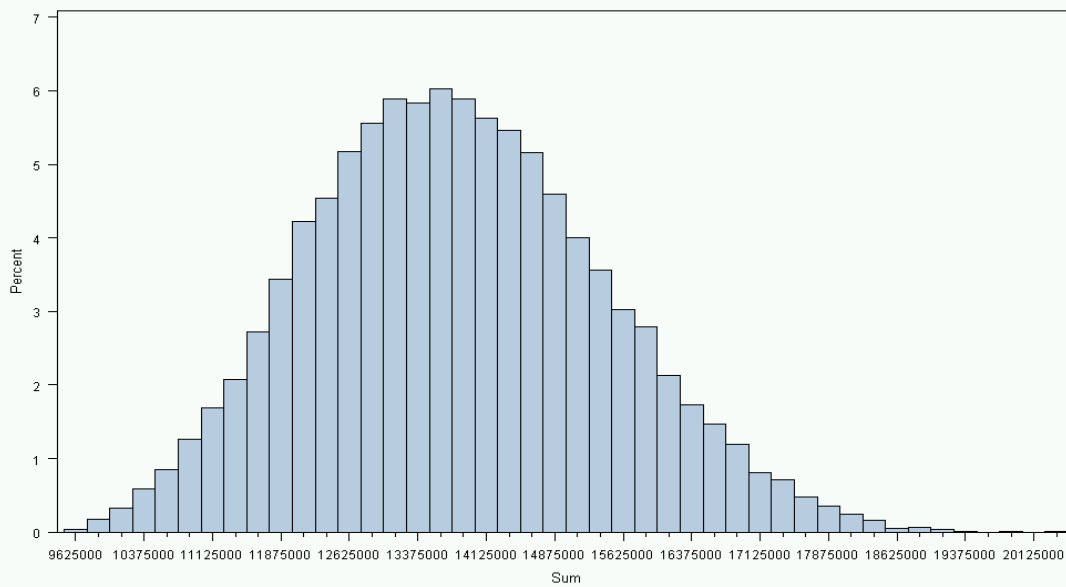
SRS n=120 MPU method, POP2



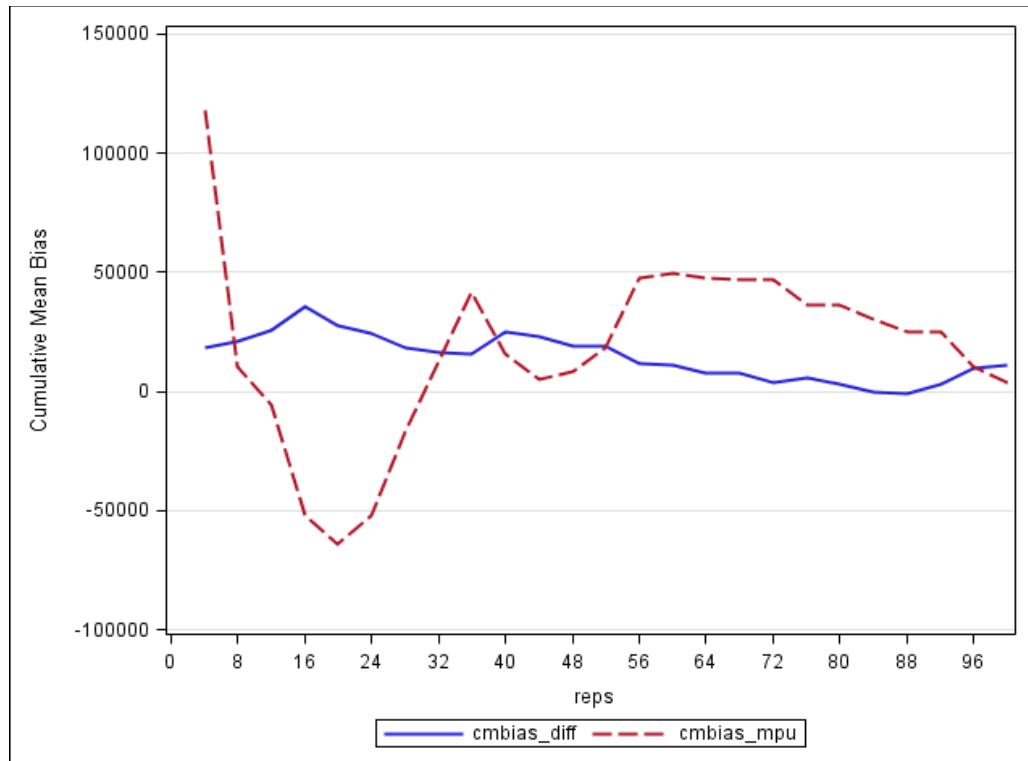
Strat n=120 Diff method, POP2



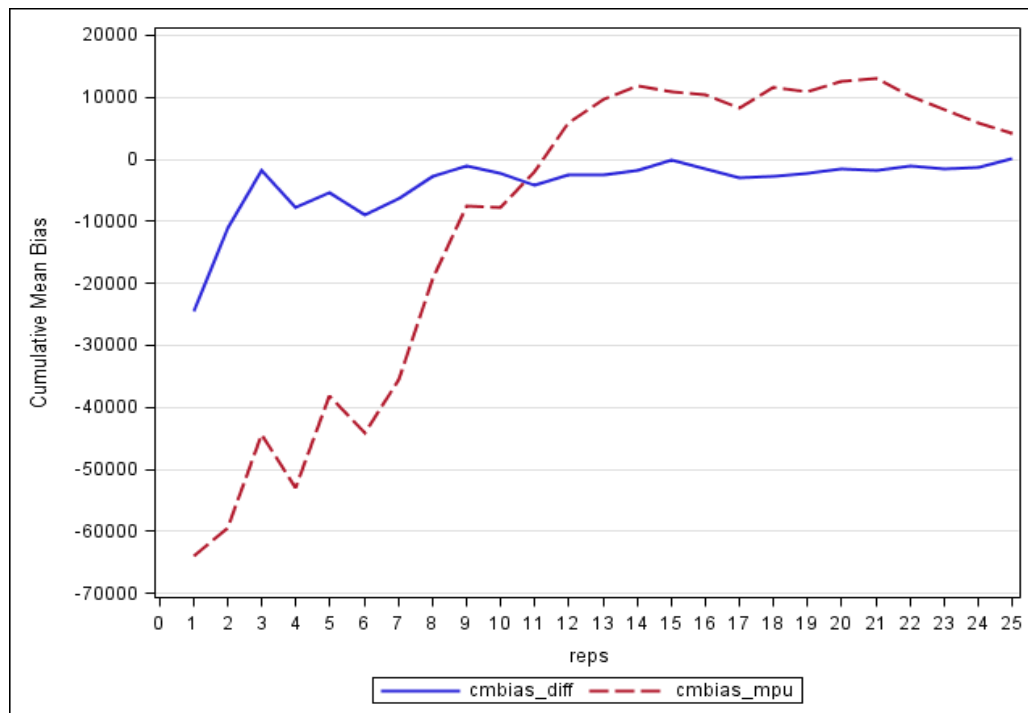
Strat n=120 MPU method, POP2



Bias of the Total: MPU and Difference Estimation Method (SRS, n = 120)



Bias of the Total: MPU and Difference Estimation Method (Stratified, n = 120)



It is clear from the graph that the convergence behavior of the bias for the difference method is more stable.

COVERAGE PROBABILITY

The coverage probability is the probability that the true value is within the confidence interval. We ideally want the coverage probability to be closer to $1 - \alpha$. We calculated the coverage probability and compared with the confidence level for both the Difference Estimation and the MPU methods. Below are the simulation results of coverage probability for SRS and stratified designs with various error rates.

Simple Random Sample Design

Table 5: Coverage Probability (Population Size = 1,500, Error Rate = 10% and Total Error = 4% of Book Value)

Sample Size	Coverage probability for Difference Estimation Method (Audit Value = Book Value + Error)	Coverage probability for Mean Per Unit Method (Estimate Audit Value Directly)
120	0.31	0.62
180	0.35	0.63
240	0.35	0.69

We now increase the error rate to 40 percent. Total error is eight percent of the book value. Below is the simulation result of coverage probability for various sample sizes.

Table 6: Coverage Probability (Population Size = 1,500, Error Rate = 40% and Total Error = 8% of Book Value)

Sample Size	Coverage Probability for Difference Estimation Method (Audit Value = Book Value + Error)	Coverage Probability for Mean Per Unit Method (Estimate Audit Value Directly)
120	0.55	0.62
180	0.60	0.68
240	0.66	0.70

It is clear from the simulation results that SRS design produces poor coverage probabilities for both methods and the coverage probability increases as the error rate and sample size increases. The mean per unit method is marginally better than the difference estimation method.

Stratified Sample Design

Table 7: Coverage Probability (Population Size = 1,500 [5, 48, 255, 535, 657], Error Rate = 10% and Total Error = 4% of Book Value)

Sample Size	Coverage probability for Difference Estimation Method (Audit Value = Book Value + Error)	Coverage probability for Mean Per Unit Method (Estimate Audit Value Directly)
120 (5, 15, 40, 35, 25)	0.55	0.86
180 (5, 20, 55, 55, 45)	0.66	0.88
240 (5, 25, 70, 75, 65)	0.77	0.88

We now increase the error rate to 40 percent. Total error is eight percent of the book value. Below is the simulation result of coverage probability for various sample sizes.

Table 8: Coverage Probability (Population Size = 1,500 [5, 48, 255, 535, 657], Error Rate = 40% and Total Error = 8% of Book Value)

Sample Size	Coverage Probability for Difference Estimation Method (Audit Value = Book Value + Error)	Coverage probability for Mean Per Unit Method (Estimate Audit Value Directly)
120 (5, 15, 40, 35, 25)	0.82	0.83
180 (5, 20, 55, 55, 45)	0.85	0.86
240 (5, 25, 70, 75, 65)	0.83	0.87

The results of the stratified design are consistent with the results of the SRS. The mean per unit method outperforms the difference estimation method.

CONCLUSION

The estimator with larger precision is not useful for decision makers or to book the adjustment in financial statements. In general, the precision for difference method is better than the precision for the mean per unit method for SRS and stratified designs. The precision improves for both methods when the sample size increases. In many cases, the difference estimation method outperforms the mean per unit method.

For SRS it is clear from the simulation results that difference estimation method produces poor coverage probabilities. The coverage probability increases as the sample size increases. The mean per unit method outperforms the difference estimation method.

From the graph we see that the difference estimator converges smoothly to the true value. The mean per unit estimator oscillates and converges to the true value.

For skewed population SRS is not reliable since it produces large precision and poor coverage probability. Since the difference estimation method produces poor coverage probabilities the mean per unit estimator is preferable for skewed populations.

We plan to consider various positive skewed populations with different error rates and study the characteristic of the difference and mean per unit estimators. As a special case we will also consider the population with almost equal -over and -under statement errors.

REFERENCES

- Cochran, William, 1977, Sampling Techniques, New York, John Wiley & Sons.
- Brown, Lawrence, etc., 2001, Confidence Intervals for a Binomial proportion and Asymptotic Expansions, The Annals of Statistics, 160-201.

ACKNOWLEDGMENTS

We would like to thank Mr. Jim Hartman, Technical Director, Quantitative Methods Division, for encouraging and providing support for this research work.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: John Chantis
Address: Office of Inspector General, Department of Defense
Suite 13F25
4800 Mark Center Drive
Arlington, VA 22350
Work Phone: (703) 604 8925
E-mail: john.chantis@dodig.mil

Name: Kandasamy Selvavel
Suite 13F25
4800 Mark Center Drive
Arlington, VA 22350
Work Phone: (703) 604 8919
E-mail: kandasamy.selvavel@dodig.mil

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.