Minimum Sample Sizes with Rare Events in Stratified Designs

Eric Falk, Joomi Kim and Wendy Rotz, Ernst and Young

Abstract

There are many statistical issues in using stratified sampling for rare events. They include designing to ensure enough of the rare events are found in each stratum, achieving adequate accuracy of the estimates and their variance, and meeting the normality assumptions for confidence interval construction. When sampling to estimate sales and use tax, a large population is sampled to find a rare event (taxable amount) and a small ratio is calculated. Many state taxing authorities that use statistical sampling in sales and use tax studies require minimum stratum sample sizes of anywhere between 100 to 350 records. In this paper we study when this minimum requirement is too much, too little, or just right.

Simulation was used to assess the adequacy of stratum sample sizes. We examined the accuracy of estimation, relative precision, frequency of under/over estimation, and calculation of confidence intervals. The results of the simulations show that in some situations samples could be reduced if more information about estimated values were known in advance. This information could be obtained through historical information or through a pilot study.

Background

The impetus for this paper is the minimum sample size requirements of several state taxing authorities. For example, the states of Texas, Ohio and New York require minimum sample sizes of 250 per stratum. California requires 300 per stratum. If this sample size is larger than needed, the additional sample adds unnecessary cost and time and is likely to increase the measurement error. If the sample size is too small, then the estimated values may not be reliable.

Depending on the state, businesses are charged a sales tax for the purchase of equipment and supplies. An example of a taxable purchase would be office supplies such as furniture. Items for resale, such as clothing, are nontaxable.

Sales tax data tends to be large in population and highly skewed to the right. That is, there are many purchases between $0 to $10,000 and a relatively small percentage of purchases that are greater than $10,000.

Only five to twenty-five percent of purchases are taxable. Therefore, we are estimating a somewhat rare event within a highly skewed population of financial data. Typically, a purchase is either zero or one hundred percent taxable.

The total tax collected by all states last year was $233 billion, and $100 billion was from sales tax paid by businesses. Forty-five states have a sales tax and five states do not (Oregon, Delaware, Alaska, Montana and New Hampshire).

Thirty-six states use some form of statistical sampling to estimate sales and use tax. Eight states use non-statistical sampling or block sampling (Arizona, Arkansas, Hawaii, Maine, Nevada, New Jersey, Oklahoma, and Rhode Island) and one state (New Mexico) is undecided. Individual statistical sampling requirements among the 36 states vary from state to state [3]. The Federation of Tax Administrators released a report on the use of statistical sampling for Sales and Use tax studies in a December 2002 report. The appendix for the state-by-state sampling methodologies was updated in January 2004.

The states that allow statistical sampling generally require confidence levels between 75% and 95%, relative precision between 5% and 25%, stratified or simple random sample designs, and minimum sample sizes between 100 and 350 per stratum. There are statistical reasons for larger sample sizes, such as rarity of the taxable items and normality assumptions in the construction of confidence intervals.

Probabilities

It is intuitive that a certain number of taxable items must be found in a sample in order to achieve a reasonably accurate estimate of the taxable amount. Figure 1 shows probabilities of finding at least a minimum number of taxable purchases in every stratum of a five-stratum sample for different taxable rates and different sample sizes.

Based on these probabilities, stratum sample sizes of 50 appear inadequate when the taxable rate is only 5%, yet this might be a reasonable sample size when 15% or more is expected. For the purpose of finding a minimum number of taxable purchases in every stratum, a sample size of 200 may be unnecessarily large for 5% taxable and 100 is adequate for 10% taxable.

Figure 1, however, only addresses the probabilities of achieving a minimum number of taxable purchases in each stratum. This does not assess whether that minimum number is sufficient to achieve a reasonable estimate of the total taxable amount. That assessment is done via the simulation described below.

Figure 1. Probability of Finding At Least a Minimum Number of Taxable Cases in Every Stratum of a Five-Stratum (Non-Certainty) Sample

Tax Rate	Min. #	Stratum Sample Size					
		50	100	150	200	250	300
5%	1	68	97	100	100	100	100
	2	20	84	98	100	100	100
	3	2	55	92	99	100	100
	4	0	23	77	96	100	100
	5	0	6	53	89	98	100
10%	1	98	100	100	100	100	100
	2	85	100	100	100	100	100
	3	56	99	100	100	100	100
	4	24	97	100	100	100	100
	5	6	90	100	100	100	100
15%	1	100	100	100	100	100	100
	2	99	100	100	100	100	100
	3	93	100	100	100	100	100
	4	80	100	100	100	100	100
	5	56	100	100	100	100	100
20%	1	100	100	100	100	100	100
	2	100	100	100	100	100	100
	3	99	100	100	100	100	100
	4	97	100	100	100	100	100
	5	91	100	100	100	100	100
25%	1	100	100	100	100	100	100
	2	100	100	100	100	100	100
	3	100	100	100	100	100	100
	4	100	100	100	100	100	100
	5	99	100	100	100	100	100

Figure 1 shows the probability (rounded to the nearest percent) of finding at least the number of taxable cases shown in every stratum of five (non-certainty) strata for a given stratum sample size and actual population percent of taxable purchases. While the figures are displayed for a stratum with 2,500 purchases, the results are similar for almost any stratum size of 1,000 or more. If the number of strata is increased, the probabilities would decrease, if the number is decreased, the probabilities would increase.

Simulation

We evaluated sample size requirements using computer simulations. The simulated population data consisted of approximately 90,000 records with an associated purchase amount of about $215 million. We then simulated five estimation variables associated with each purchase: taxable 5%, taxable 10%, taxable 15%, taxable 20% and taxable 25%, where the percentage indicates the percent of taxable purchases in the hypothetical population. The taxable purchases in the population were randomly assigned using the random uniform distribution function within SAS (RANUNI).

The design structure for the sample is a one-stage stratified random sample, where the equal NhSh method was used to form stratum boundaries [1, pp. 127-131]. An advantage of this method is optimal allocation assigns equal sample sizes to the non-certainty strata, thus simplifying our testing of different stratum sample sizes. Figure 2 shows the sample design that was used in the simulations.

Using the sample design shown in Figure 2, we assigned stratum sample sizes of 50 to 700 (incrementally by 50) to all of the non-certainty strata (strata 1 through 5). Figure 3 provides a summary of the sample size scenarios tested. For each scenario, we drew 1,000 simulated samples and estimated the taxable amount of each of the five hypothetical variables using the combined ratio estimator.

Figure 2. Sample Design

Stratum	Stratum Definition	Population	Population Amount ($Millions)
1	$0 to $1.15K	69,500	$16.30
2	$1.15K to $4.47K	12,500	$28.00
3	$4.47K to $11.96K	4,750	$34.00
4	$11.96K to $27.47K	2,000	$35.10
5	$27.47K to $290K	1,200	$70.50
6	$290K and over	50	$31.10
Total		**90,000**	**$215.00**

Figure 2 shows the sample design used in all simulations.

Results

In evaluating the stratum sample sizes, we compared the distributions of estimated values, confidence interval coverage, and relative precision of the estimated values.

Figure 3. Sample Size Summary

Scenario	Sample Size per Stratum	Total Sample Size
1	50	300
2	100	550
3	150	800
4	200	1,050
5	250	1,300
6	300	1,550
7	350	1,800
8	400	2,050
9	450	2,300
10	500	2,550
11	550	2,800
12	600	3,050
13	650	3,300
14	700	3,550

Figure 3 shows the sample size per non-certainty stratum and the total sample size used in each stratum.

First we compared the distribution of the estimated values using box plots. However, for ease of comparison, rather than graphing the distribution of the estimates, we plotted the ratio of the estimated amount to the actual value. See Figure 4 for the 5 and 25 percent taxable distributions. A ratio near one represents an estimate close to its true value.

As expected, as the sample size increases, the ratios decrease. Notice that for both estimation variables, the ratios decrease dramatically for minimum sample size of 50 to 100. However, a point of diminishing returns is reached rather quickly as the distributions markedly level off. For the 5% taxable variable, the ratios stabilize around a stratum size of 300 to 350 (near some of the state guidelines for minimum stratum sample sizes). However, the 25% taxable variable stabilizes much sooner, around 100 to 150, much lower than many state minimum requirements.

For both estimation variables, the majority of the ratios are centered on 1.0, but the range is quite a bit smaller in the 25% taxable sampling distribution. One would expect less variability with the higher occurrence rates, but it is interesting to note how much effect the occurrence rate has on the stability of the estimates. For example, with stratum sample sizes of 50, several simulations had amounts that were 2 times the actual for the 5% taxable variable, yet for the 25% taxable variable, the ratio is at most 1.5.

Also note, for the 5% taxable population, the median estimate is slightly below the actual taxable amount for the first few sample sizes. This occurred because there were too few taxable purchases in some strata, causing the sample to under-estimate the true value more often than not with the smaller sample sizes.

Figure 4. Ratio of Estimated to Actual Taxable Amounts

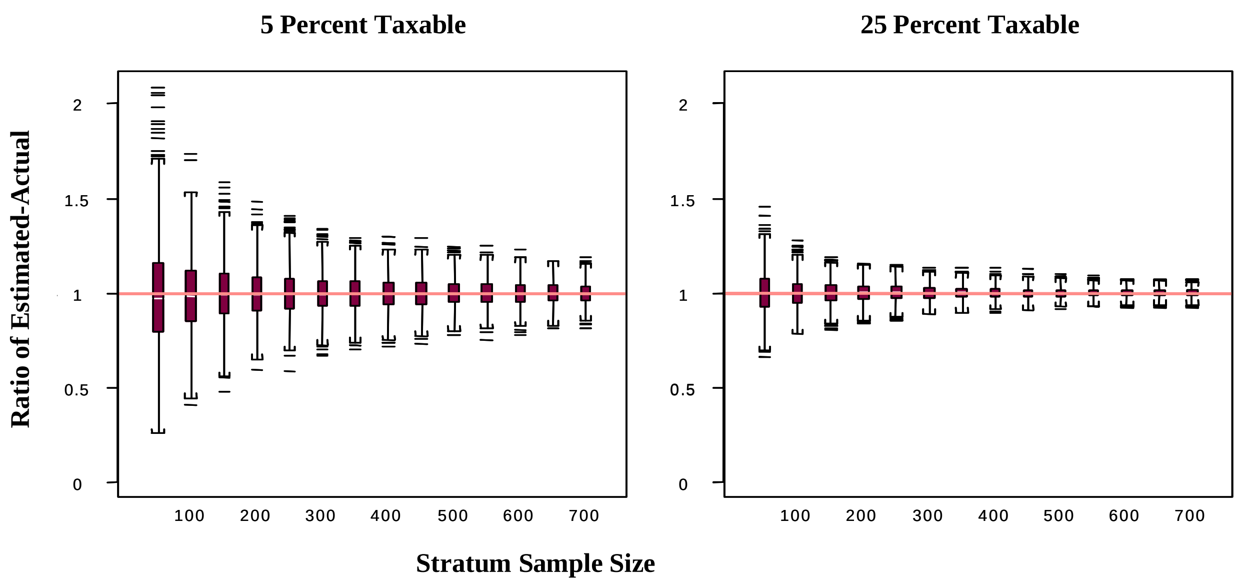


Figure 4 displays box plots showing the distribution of the ratio of the estimated values to their actual value for the 5 and 25 percent taxable populations. A ratio near 1 indicates an estimate near its true value. The figure shows the estimate from the 25 percent taxable population as much less variable as from the 5 percent taxable population. However, there is a point of diminishing returns for both populations, where increases to the sample have only minimal impact on the estimate's variability.

Figure 5. Percent of Simulations that Under Estimating the True Value

Scenari	Stratum Sample Size	Estimation Variable				
		5% Taxable	10% Taxable	15% Taxable	20% Taxab	25% Taxab
1	50	53.9%	51.4%	51.9%	50.5%	49.8%
2	100	54.3%	52.6%	51.3%	49.6%	49.5%
3	150	51.9%	49.7%	51.5%	49.6%	50.0%
4	200	52.3%	51.2%	50.0%	48.1%	48.2%
5	250	51.8%	51.8%	50.9%	47.9%	47.3%
6	300	52.8%	51.9%	49.1%	46.4%	48.3%
7	350	52.0%	51.5%	49.9%	48.4%	48.0%
8	400	51.9%	51.9%	50.0%	49.0%	50.5%
9	450	51.5%	51.2%	49.2%	48.4%	49.8%
10	500	52.1%	49.8%	47.6%	50.5%	50.0%
11	550	50.3%	51.9%	49.1%	51.3%	49.6%
12	600	50.8%	52.5%	49.5%	51.9%	50.0%
13	650	51.0%	52.5%	50.8%	51.6%	50.8%
14	700	50.1%	51.0%	50.7%	52.7%	50.1%

Figure 5 shows the percentage of under-estimated values from each set of 1,000 simulations. Pink indicate more frequent under-estimation and blue indicates more frequent over-estimation (where under-estimation is less than 50%). The lighter shading indicates percentages within 2% of 50%. Note that the majority of under-estimations occur with the lower taxable percentages, especially with the smaller sample sizes.

the sample to under-estimate the true value more often than not with the smaller sample sizes. Under-estimating taxes would be a state cause for concern. However, sample sizes in excess of 300 do not have this problem for the 5% taxable population. The 25% taxable population does not have this problem at all, not even with stratum samples of only 50 purchases.

Figure 5 shows the percent of simulations that under estimated the true value for each scenario. Pink cells indicate more than 50% of the simulations under-estimated. Blue indicates more than 50% over-estimated. All sample size scenarios for the 5% taxable variable under-estimated more frequently than over estimated. However, by 15% taxable, there is a more even mixture of under and over estimation.

Next, we assessed coverage of 90% confidence intervals, illustrated in Figure 6-10 below. The striped orange area in the center represents the portion of simulations where the confidence interval contained the actual value. The blue portion to the left represents the proportion of the simulations where the entire confidence interval fell below the actual value and the green portion to the right represents the proportion of confidence intervals that were above the true value.

Note that regardless of the sample size, the 25% taxable population has fairly symmetrical coverage and good coverage near 90%. However the 5% taxable population is asymmetrical for all levels and is often less than 90%.

Figure 6. Over/Under Estimating the Taxable Amounts - 5%

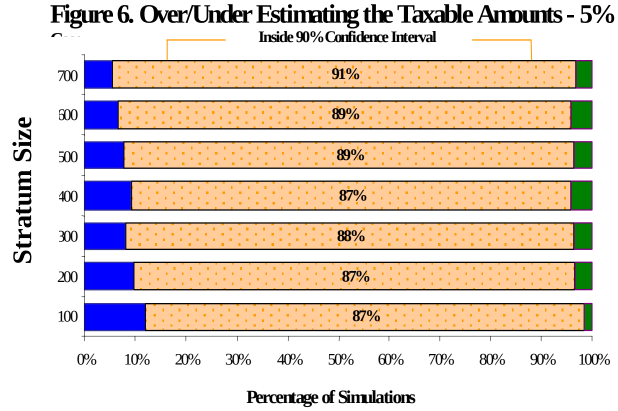


Figure 7. Over/Under Estimating the Taxable Amounts - 10%

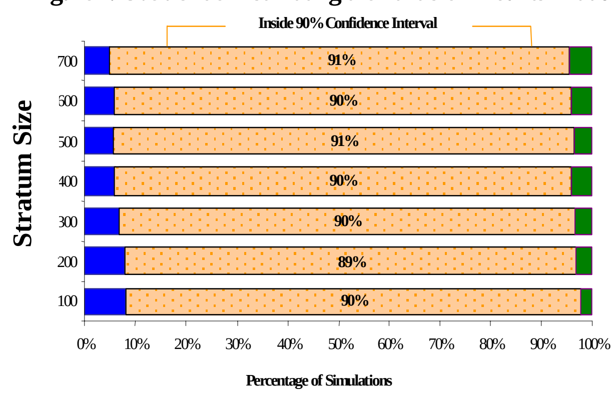


Figure 8. Over/Under Estimating the Taxable Amounts - 15%

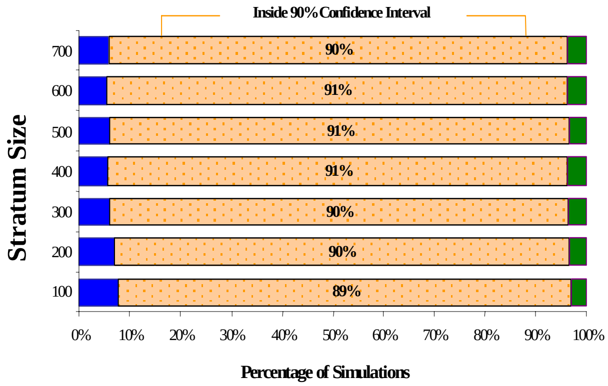


Figure 9. Over/Under Estimating the Taxable Amounts - 20%

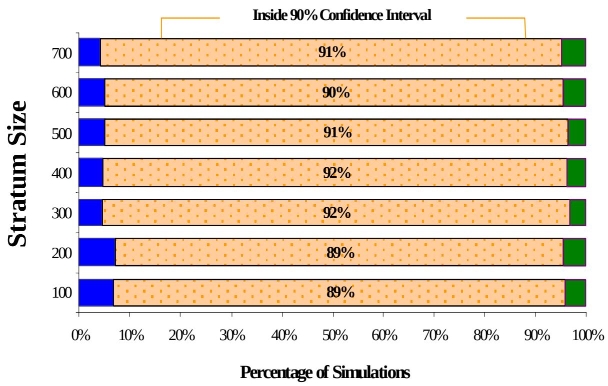
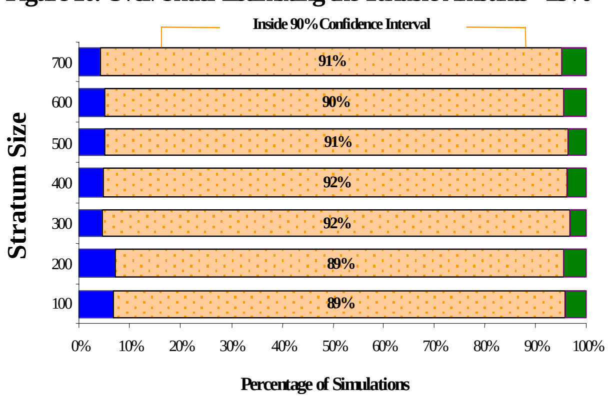


Figure 10. Over/Under Estimating the Taxable Amounts - 25%



Not until sample sizes of 700 is the coverage more symmetrical, and with this low occurrence rate, it would be a questionable use of resources to sample enough for more symmetrical coverage. Regulators and taxpayers may be better off to simply realize their 90% confidence intervals using standard methods relying on normal distribution theory are likely to be somewhat inaccurate.

There are minimal gains in the correct confidence interval coverage from increasing the sample size per stratum, but what about the width of the confidence intervals? This is assessed by exploring relative precision, which is closely related to the confidence interval width.

Figure 11 shows improvement in the average relative precision as the stratum sample sizes increase, given the different taxable distributions. For example, for the 5% taxable distribution, and a sample size of 50 per stratum, the average relative precision over the simulations was approximately 45%. Whereas, the 25% taxable distribution with a stratum sample size of 50 per stratum, has an average relative precision of roughly 20%.

Figure 11. Relative Precision to Sample Size

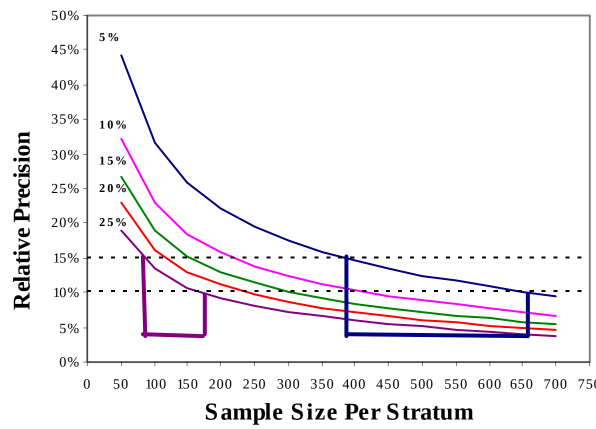


Figure 11 shows the average relative precision per scenario. The y-axis is the average relative precision over the simulations. The x-axis is the minimum sample sizes. The lines represent the five taxable distributions.

In all situations, as expected, the average relative precision improves as the sample size increases, but the significance of the improvement diminishes soon after the stratum sample sizes reaches 100. In the 25% taxable distribution, stratum sample sizes of roughly 75-100 achieve an average relative precision of 15%. For 10% taxable, this is achieved with roughly 175-200 per stratum.

In the 5% taxable distribution scenario, a stratum sample size of 375-400 achieves an average relative precision of 15%. Stratum samples of roughly 675-700 are required for 10% precision. However, is this improvement worth nearly doubling the sample to achieve? How much more comfort does 10% precision give than 15%?

Conclusion

As expected, increasing the minimum sample size per stratum improves the precision but there is a point of diminishing returns. For the 25% taxable distribution, a sample size of 75-200 would achieve reasonable relative precisions. For the 5% taxable distribution, a sample size of 375 or larger would achieve reasonable relative precision.

Minimum stratum sample sizes of 200 to 300 or so purchases may be reasonable for taxable rates as low as 5%. However, the requirements can be relaxed for higher taxable rates.

A one-size fits all sampling approach means that samples must be oversized to accommodate the lower taxable rates. These sizes are unnecessarily large with higher taxable rates.

Using historical information or conducting pilot studies to obtain approximate taxable rates prior to a full study would help the states reduce sample sizes. This would reduce their costs and taxpayer burden.

Sample size requirements should be balanced against the cost of the study, sampling error, and measurement error. The role of the statistical consultant is to determine sample sizes that yield reasonably reliable estimates and to ask whether the marginal statistical benefits of an excessively large sample are worth the increased cost associated with the review of more records. The challenge is to present the cost benefit analysis in terms regulators and taxpayers understand.

Next Steps

Taxpayer statistical consultants and state statisticians could work more closely together to explore sample reduction methods. Minimum sample sizes can be relaxed with the development of sample plans more tailored to each taxpayer's setting.

References

[1] Cochran, W.G. *Sampling Techniques*, Third Edition, John Wiley and Sons, New York, 1977

[2] Falk, Eric and Wendy Rotz, 2003 Joint Statistical Meeting Proceedings.

[3] FTA, *Sampling for Sales and Use Tax Compliance, A Report of the Steering Committee*, Tax Force on EDI Audit and Legal Issues for Tax Administration, Federation of Tax Administrators, Washington D.C., December 2002. Appendix A was updated on January 2004.

[4] Roberts, Donald M. *Statistical Auditing*, American Institute of Certified Public Accountants, New York 1978.