Stratified Sampling for Sales and Use Tax Highly Skewed Data – Determination of the Certainty Stratum Cut-off Amount

By Eric Falk and Wendy Rotz, Ernst and Young LLP

Introduction

With the increasing number of states facing continued budget cuts and reduction of resources supporting the audit function, the use of statistical sampling is a welcome alternative for businesses conducting sales and use tax studies in an effort to reduce the cost associated with such studies. Statistical sampling also decreases taxpayer burden in providing states the supporting documentation for these audits. Above all, statistical sampling produces reliable results when estimating a company's sales and use tax liability.

In these studies, purchase amounts are randomly sampled and the tax liabilities of sampled purchased amounts are used to estimate the tax liability of the company.

As with most financial data, the distribution of these purchased amounts is highly skewed with long tails (see Figures 1 and 2). Therefore, stratification is used for sales and use tax studies, and, part of the stratified sample design is a certainty stratum.

The statistical reason for selecting a certainty stratum is to capture and isolate the largest dollar purchases so that their extremely large values do not influence sampling variability. This allows more stable estimates and an overall sample size reduction while meeting the desired confidence and precision goals.

However, some state taxing authorities choose significantly more than the extreme outliers for their certainty stratum in order to gain a higher comfort level. The impetus for this paper is to explore theoretical and simulated estimation results to illustrate a more data driven method for selecting the certainty stratum cut-off. The hope is that states may use this research along with their own to assist them in using statistical sampling as a tool to obtain accurate and efficient estimates.

This paper examines a common state approach compared to a data-driven method of determining certainty stratum cut-offs. The precision obtained and total sample sizes with large certainty strata are examined in theory and verified with hypothetical data and simulation. Theory and simulation show large certainty stratum sizes do improve the relative precision of the estimated tax liability, but there is a point where large increases to the certainty stratum only marginally improve the precision, and are therefore a questionable use of resources.

Background

Forty-five states and the District of Columbia impose sales and use taxes, though the application and tax rates vary greatly among different jurisdictions.

Sales tax is generally levied on sales of tangible personal property and certain specified services. Use tax is a complementary tax imposed by a state on the privilege of using, consuming or storing tangible personal or specified services. Although often believed to be imposed on only retail sales, sales tax also falls on many business purchases of inputs.

Generally, most state statutes presume all purchases of tangible personal property are subject to tax unless a sale is specifically excluded or exempted. Examples of exempted purchases include: groceries, prescription drugs, medical supplies, materials and equipment used in manufacturing, and/or sales to exempt organizations. State auditors review these business purchases to evaluate whether the company correctly identified them as exempt or taxable.

Most companies have internal accounting systems that capture sales and use tax that is either paid or accrued. The audit issue is the accuracy of the system making taxable/exempt determinations.

Some companies automate their use tax accrual process with software, such as Vertex or TaxWare, to create a tax decision matrix that is mapped to their general ledger. When this tax decision matrix is properly configured to the general ledger, and the tax determinations are correct, there are few accuracy issues in audit. However, inaccuracies occur when categories within the tax decision matrix are too broad (i.e., include both taxable and non-taxable determinations), the mapping to the general ledger is corrupted, or the matrix is not updated to incorporate changes in tax law.

Other companies have less sophisticated systems and individual tax determinations are made by accounts payable clerks as they process payments. These companies clearly have more opportunities for error.

Large corporations process tens of thousands of invoices annually and thus it would be both time consuming and inefficient for auditors to review each invoice. Some sort of sampling to determine the taxpayer's level of compliance has been long established.

Historically, state auditors used block sampling in sales and use tax studies. For example, they may have judgmentally selected three "typical representative" months in an audit period, examined all purchases in those months, and extrapolated the three months to the entire audit period. Unfortunately, block sampling can be biased due to non-random selection of the months, fluctuations in purchasing volume, and seasonal patterns in purchases. Furthermore, the bias, the accuracy and precision of estimates derived from non-random block sampling cannot be calculated.

Statistical sampling is a marked improvement over historical state approaches; however, there is still the legacy of block sampling. Because the precision cannot be determined in block sampling, it is still not enough of a focus in statistically sampled state audits. Currently 36 states allow sampling for the purpose of estimating sales and use tax compliance, but only about half of these clearly indicate the precision is a consideration either in the design or estimation. [3 pp 26-62]

About Sales and Use Tax Data

Before examining the certainty stratum, consider the typical sales and use tax data influencing the sampling situation.

The sample design goal is to estimate the tax liability and percent of purchase dollars with tax liability. The sample designs are typically based on the related variable, purchase amount from the invoice. The distribution of purchase amounts is highly skewed to the right (see Figure 1).

In addition, the percent of invoices with tax liability is small, generally ranging between five and twenty percent. Some states assume only 1% of the records will have tax liability. Therefore, the audit situation is sampling for a rare or infrequent event in highly skewed data.

The estimation variable, the taxable amount, is either zero or the entire purchase amount. Therefore, the majority of the estimation variable's values are zero.

Throughout the remainder of this paper we will use the hypothetical setting of a large manufacturing firm with 89,242 purchases totaling $214,712,300 during a particular year.

Figure 1 shows the distribution of the purchase amounts on an equal scale. Figure 2 shows the same distribution of data, however, the scale is unequal to better examine the tail.

Figure 1. Distribution of Purchase Amounts (Equal Scale)

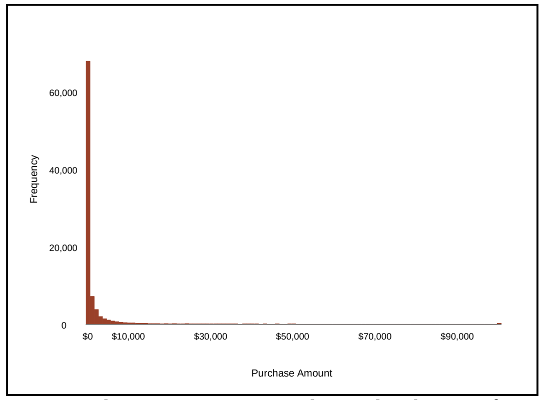


Figure 1 shows a common population distribution of purchase amounts for sales and use tax studies. There are typically thousands of small purchases under $1,000, a substantial number of purchases up to about $10,000 dollars, a long tail, and extreme outliers.

Figure 2. Distribution of Purchase Amounts (Unequal Scale)

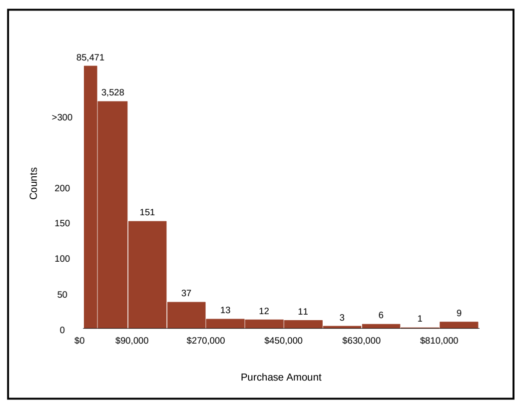


Figure 2, shows the same purchase data as Figure 1, but on a base 10 scale to better examine the long tail. Note: In this example, there are actually several thousand records over $10,000 and even several hundred over $100,000.

State Taxing Authorities and Sampling

In December 2002, the Federation of Tax Administrators (FTA) released a document on its website that was intended to align the business community and tax administrators when sampling is used in sales and use tax studies. [3] The document indicates that of the 45 states (plus the District of Columbia) with a sales tax, 36 allow statistical sampling[1], seven (Arkansas, Hawaii, Maine, Nevada, New Jersey, Oklahoma, and Rhode Island) do not allow statistical sampling, and three (Arizona, New Mexico and the District of Columbia) have not decided whether to allow statistical sampling (see Figure 3).

In addition, the FTA document indicates that 20 states allow taxpayers to initiate a vendor tax overpayment study using statistical sampling. Twenty-four states allow taxpayers to initiate a use tax overpayment study using statistical sampling[2].

The FTA document is specifically written for sampling in sales and use tax audits. Statistical sampling is also being used by state governments and taxpayers with unclaimed funds, property tax, and many other areas where the records to review are far too voluminous to cost effectively review every single transaction.

Figure 3. States and Sampling

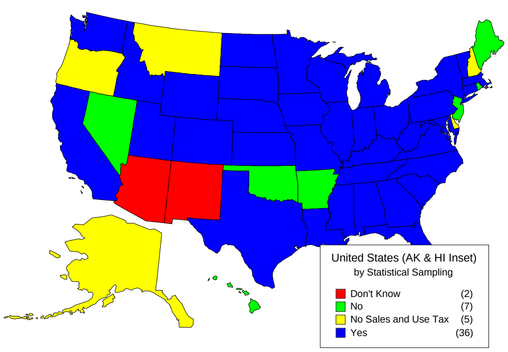


Figure 3 shows that most of the US states allow sampling for sales and use tax. Arkansas, Hawaii, Maine, Nevada, New Jersey, Oklahoma, and Rhode Island do not allow sampling. Arizona, New Mexico and the District of Columbia have not decided. Alaska, Oregon, Montana, Delaware and New Hampshire do not have sales and use tax.

Non-Certainty Strata

Many of the state taxing authorities require large minimum sample sizes within non-certainty strata. These minimum stratum sample size requirements range from 100-300. A statistical reason for this requirement is the low percentage of taxable purchases. The concern is to obtain enough taxable purchases in each stratum to produce stable estimates and use normality assumptions in determining confidence intervals. Another reason for large minimum sample sizes within stratum is that ratio estimators are used in the estimation. Ratio type estimators require larger sample sizes to obtain estimates with negligible bias.

A separate paper will consider whether the minimum stratum sample size requirements adequately meet these statistical concerns and whether they could safely be reduced. The focus of this paper is that on top of the large minimum stratum sample size requirements, some states select a large number of certainty stratum cases to review.

The consequences of both the large certainty and non-certainty stratum sample size requirements are larger samples, higher costs, and longer timelines to make the taxability determinations.

Defining the Certainty Stratum

The statistical goal in defining certainty strata is to identify the extreme values within a population that heavily influence the estimate and its variance. Sampling all of these records in a certainty stratum reduces the sampling error to zero for this stratum and stabilizes the estimates.

The state taxing authorities typically decide their certainty stratum based on judgment. Many of the state taxing authorities decide on the cut-off value based on a comfort level. That is, the choice of the cut-off value may be based on a total percentage of the purchase amount or the highest X purchase amounts. Figure 4 shows a certainty cut-off that many of the states would find comfortable (>= $10,000).

In a data-driven approach, the certainty stratum cut-off value is determined to isolate outliers and is based on statistical relative precision. The frequency distribution of purchase amounts is considered. The certainty stratum cut-off value should be taken at the point where data is sparse and no longer clustered. The remaining clustered data would then be stratified and subject to sampling (see Figure 4 - >= $290,000).

Figure 4 shows a data-driven determination of the certainty stratum cut-off would be much higher. The thick portion below the cut-off would be sampled in the non-certainty strata.

[1] The term statistical sampling versus non-statistical sampling is unclear in several of the states' description of their procedures. Statistical sampling is defined as the type of sampling when each item in a sampling population has a known non-zero probability of selection. In addition, estimates produced from a statistical sample include confidence and precision measures.

[2] These statements may not be reflective of the actual positions currently held by these taxing authorities.

Figure 4. Certainty Stratum Determinations

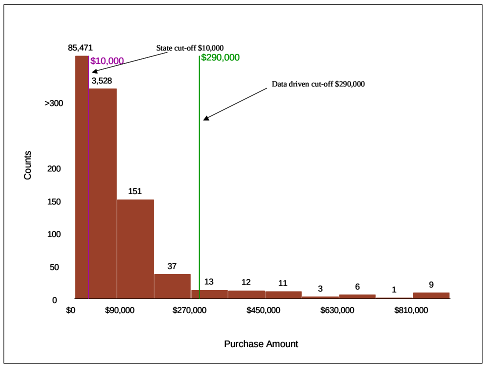


Figure 4 shows the arbitrary certainty stratum cut-off a state taxing authority may use, based on their comfort level. A state sample design may place all purchases over $10,000 in a certainty stratum, even though this may be thousands of invoices. The thickness of the distribution to the right of the cut-off suggests non-certainty strata could be formed in this tail, which would reduce the overall sample size.

During the course of designing the sample in a data-driven method, the certainty stratum cut-off might be adjusted downward if sample allocation calls for all, or nearly all, records in a non-certainty stratum to be selected. More certainty cases might also be selected to attain a desired precision goal.

There are more sophisticated methods of determining certainty stratum cut-off boundaries, for example, see the work of Hidiroglou. [6] However, the approach of graphing the data and adjusting when necessary is simple and practical.

Case Study - Theoretical Precision

To examine the precision of samples with these large certainty strata, consider sampling this typical sales and use tax population using five equal dollar strata and a sample size of 250 per non-certainty stratum.

For the certainty stratum, consider multiple designs. In the first design consider a rather modest certainty stratum of 50 purchases, and then in subsequent designs, gradually increase the certainty stratum until it encompasses 5,000 records in the last design. As the certainty stratum increases, readjust the five non-certainty strata based on equal dollars, and keep the non-certainty stratum sizes of 250 for every design. Thus the first design has a total sample size of 50+ 5 x 250=1,300 purchases; the last design has 5,000 + 1,250=6,250 purchases.

Theoretically, the precision is expected to continually decrease in each design as this certainty stratum and total sample size is increased (see Figure 5). Indeed, if the entire population were selected in the certainty stratum, the precision would be zero.

However, there is a point of diminishing returns when large increases in the certainty stratum only produce minimal improvements in the precision. Many regulators consider 15% relative precision to be quite good. This example shows that selecting excessive sample sizes in the certainty stratum increases the precision but not significantly.

Figure 5. Theoretical Relative Precision Results

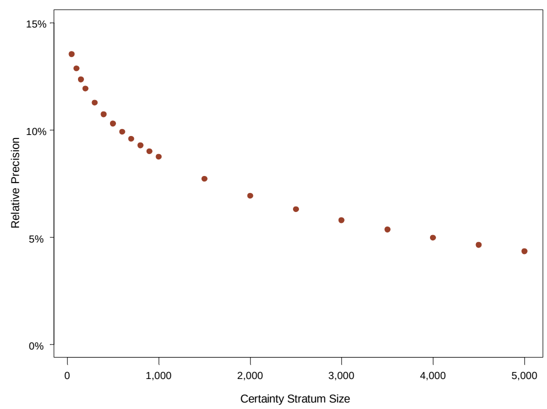


Figure 5 shows the theoretical precision (using 90% confidence) for a typical state sample as the certainty stratum size is increased. At one point, increases to the certainty stratum have only minor increases to precision.

Case Study – Simulated Precision

Using the same typical population data, a simulated population was created by randomly assigning 10% of these invoices to be taxable. For each design, 500 random samples were selected. The taxable amount, and its precision were calculated for each sample.

The simulation confirms the theory. Figure 6 is the average relative precision obtained in 500 random draws for each certainty stratum design considered above. The data are quite consistent with the theory.

Figure 6. Simulation Relative Precision Results

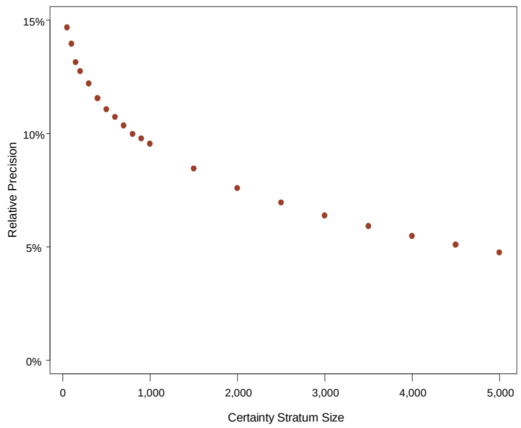
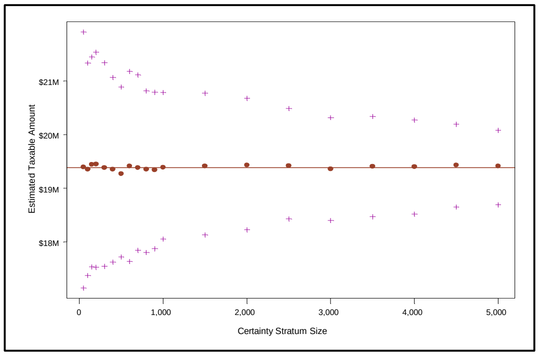


Figure 6. For each design (certainty stratum size), this figure shows the average precision amounts of 500 samples from the simulated population. Note the similarity between Figures 5 and 6. The simulation confirms the theory.

Figure 7 shows the average estimated value of the 500 samples per design. As expected, these are close to the actual value, which considering the central limit theorem may be just a function of the number of simulations.

The 90th and 10th percentiles of the estimates are also shown in Figure 7. They too demonstrate that, while the spread between the 90th and 10th percentiles does narrow by increasing the certainty stratum sample size, it too has a point of diminishing returns.

Figure 7. Simulation Estimation Results



Conclusion

State taxing authorities may reduce their sample sizes and taxpayer burden by considering a data driven approach of determining certainty strata. Statistical measures of sampling error, such as relative precision; should be the basis of the "comfort zone" to determine whether sample sizes are large enough.

A large certainty stratum containing thousands of records is usually inefficient and wasteful. Increasing the certainty stratum does improve the precision, however, there is a point of diminishing returns when large increases in the certainty stratum have only a minor impact on the precision.

A potential obstacle to using data-driven decisions based on attaining a reasonable degree of precision is the statistical knowledge required to implement this approach. Although selecting an appropriate cut-off using the data driven approach requires very little statistical knowledge, determining the overall sample size necessary to reach desired confidence and precision measures requires more statistical knowledge. Taking a large percentage of the top dollars for the certainty stratum, using equal dollars to define non-certainty strata, then taking very large non-certainty stratum sample sizes requires little statistical training. Tailoring stratum sizes for each company requires more knowledge.

Next Steps

Further simulation of alternative designs to determine certainty boundary cut-offs can be done with the goal of attaining a specified precision level for a lower sample size. Bigger isn't necessarily best.

However, any consideration of reallocating certainty stratum resources to obtain more efficient designs should also consider the large minimum stratum sample sizes required by some states. Testing of the robustness of normality assumptions, coverage of confidence intervals, and performance of the ratio estimator in the setting of rare events and stratified designs could be done to determine whether the states might reasonably relax minimal stratum sizes in conjunction with using a data-driven certainty cut-off.

States could ease into data-driven methods by trying the design approach, evaluating the precision, and taking a larger sample if a better precision level is desired.

References

[1] Arkin, Herbert and Colton, Raymond R. **Statistical Methods,** Harper and Row, New York, 1962

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

[3] 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

[4] Hansen, Morris H., Hurwitz, William N., and Madow, William G. **Sample Survey Methods and Theory, Volume 1,** John Wiley and Sons, New York, 1953

[5] Hansen, Morris H., Hurwitz, William N., and Madow, William G. **Sample Survey Methods and Theory, Volume 2,** John Wiley and Sons, New York, 1953

[6] Hidiroglou, M.A. **The Construction of a Self-Representing Stratum of Large Units in Survey Design**, The American Statistician, Vol. 40, No. 1, February 1986

[7] Kish, Leslie. **Survey Sampling,** John Wiley and Sons, New York, 1965

[8] McCarthy, P.M. and Clickner, R.P. **Optimum Sample Designs for Skewed Populations**, Proceedings of the American Statistical Association Secton on Survey Research Methods, 1985

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

[10] Skinner, C. J., Holt, D., and Smith, T. M. **Analysis of Complex Surveys,** John Wiley and Sons, New York, 1989

[11] Wilburn, Arthur J. **Practical Statistical Sampling for Auditors,** Marcel Dekker, New York, 1984

Acknowledgements: Many thanks to Ru Sun for the graphical analysis, Joomi Kim for SAS programming and analysis assistance, Mary Batcher and David McGrath for their statistical input, and Linda O'Donohue for providing background on sales and use tax in the business setting.