Understanding the Differences in Regulatory Tax Sampling and Estimation Methodologies by Federal & State Jurisdictions

Zachary Rhyne,[1] PStat®, Roger Pfaffenberger,[1] PhD, ICR[2]
[1]Ryan LLC, Three Galleria Tower, 13155 Noel Road, Suite 100, Dallas, TX 75420
[1]Mathematician Student Consultant, Dallas, TX 75420

Abstract
While many taxing jurisdictions are increasingly exploring and accepting statistical sampling for tax audits, there remains a large discrepancy in accepted sampling and estimation methodologies by state. Some regulatory agencies, such as the Internal Revenue Service (IRS), have robust statistical sampling procedures closely aligned with the scientific literature of mathematical statistics and to the business requirements of taxpayers. This paper explores the current state of regulatory tax sampling particularly within states conducting sales and use tax audits and compares those states' sampling methods to the IRS sampling guidance.

Key Words: statistical sampling, tax, audit, stratification, sales and use tax, federal tax

1. Introduction[3]

The Internal Revenue Service ("IRS") Revenue Procedure 2011-42 provides "taxpayers with guidance regarding the use and evaluation of statistical samples and sampling estimates" in federal tax [29] which bolstered the IRS 2002 *Field Guidance on the Use of Estimates From Probability Samples* [11] to "promote the efficiency and consistency of the probability samples performed and examined by the IRS." The guidance addressed the development of sampling plans and the sample characteristics required for both variable and attribute sampling. The guidance included requirements for a written sampling plan to describe the sampling process and documentation of the execution of the sample. The document also provided the formulas for the stratified mean per unit estimator, the stratified difference estimator, the stratified combined ratio estimator and the stratified combined regression estimator. In current practice, the generally accepted audit sampling estimators are the four listed above provided in the IRS guidance and the separate ratio estimator and the separate regression estimator. However, results for the separate ratio and separate regression are provided for information purposes only and are not considered the best estimator to choose for reporting purposes. The IRS guidance

[1] See acknowledgment section.
[2] The American Statistical Association's (ASA) special interest group on *Statistical Auditing* provides opportunities for student mathematicians to collaborate with industry statisticians.
[3] While the authors hope this research will be informative, we alone are responsible for the opinions expressed herein which do not necessarily represent the views of our organization or any other institution.

determines the best of the four estimators in the guidance by choosing the estimator with the smallest estimated standard error for reporting the results of the audit.

It should be noted that through out this paper the use of the term audit sample (and its derivatives) is used and denotes any tax sample wither conducted by a regulatory agency or initiated and developed by a taxpayer for a variety of tax purposes.

The IRS sampling and estimation procedures follow the theoretical and applied textbooks on sampling including Cochran [6], Lohr [12], and Scheaffer, Mendenhall and Ott [21]. There have been numerous textbooks published on sampling methods for financial audits beginning with Arkin's text first published in 1963 [3], [2]. In addition to the texts by Arkin, Roberts [15] and Guy [9], [10] have published texts on audit sampling for financial audits. The objective of financial sample audits is to determine if there are material differences between the financial reports issued by firms and the results of the sample audit.

By contrast, there have been few publications on audit sampling for sales and use tax audits. Yancey [26] published a text on sales and use tax audit sampling in 2002. The text emphasizes auditing the sample results rather than placing an emphasis on the theoretical and applied methodology such as sample design and statistical estimation. The objective of sales and use tax audit sampling is to estimate the correct amount of tax that should have been collected in sales transactions and paid in purchase transactions by firms. States with sales tax normally audit medium to large firms in three or four year cycles. If the estimated tax paid in the sample audit is less than the actual tax paid, the state issues an assessment for the difference and the interest amount on the taxes due. If the estimated tax paid in the sample is greater than the actual tax paid, the state may issue a refund or require the taxpayer to file refund claim for tax overpayments.

The sampling methods and the estimators vary widely from state to state. Most states conducting sample audits based on stratified random sample designs use either the separate ratio estimator or the mean per unit estimator. And, most states base assessments or refunds on the sample point estimate though increasingly states are using confidence bound estimates for reporting the sample audit results.

This paper reviews the IRS sampling and estimation methodologies, summarizes the sampling and estimation methodologies used by states in sales and use tax audits, and compares and contrasts the procedures. Section 1 provides an introduction. Section 2 of the paper describes the typical stratified random sample design used in sales and use tax audits. Section 3 of the paper presents a comprehensive review and description of the IRS procedures. Section 4 presents a summary of procedures used by states for sales and use tax sample audits. Section 5 contrasts the similarities and differences between the procedures used by the IRS and the procedures used by states for sales and use tax audits. Finally, Section 6 provides concluding remarks and recommendations for states to adopt IRS sampling procedures for their audit sampling applications.

2. Typical Stratified Random Sample Design for Sales and Use Tax Audits

The typical sampling design used for variables sampling in sales and use tax sample audits is a stratified random sample design. Table 1 shows an example of a stratified random sample plan for an audit of a taxpayer's purchases.

Table 1: Stratified Random Sample Plan for Audit of Purchases

Stratum Label	Lower Bound (value ≥)	Upper Bound (value ≤)	Population Base Dollars	Population Base Items	Sample Base Dollars	Sample Base Items
1A	min	-$0.01	-$359,192.93	75		
1B	$0.00	$0.00	$0.00	4		
1C	$0.01	$149.99	$151,103.60	3,983		
1D	$150.00	$8,299.99	$11,491,239.24	8,759	$268,991.82	200
1E	$8,300.00	$75,000.00	$17,151,499.22	811	$4,299,717.34	200
1F	$75,000.01	max	$44,331,812.68	168	$44,331,810.68	168
Total			**$72,766,461.81**	**13,800**	**$48,900,519.84**	**568**

Most states do not sample negative valued (i.e., credit) transactions. Rather, as many negatively valued transactions are matched with positive valued transactions (i.e., debits) as possible. The remaining negative valued transactions are then reported in the sampling plan. As seen in the sample design plan in Table 1, 75 negative valued transactions could not be matched with corresponding positive valued transactions.

Transactions with $0 amounts are typically isolated for review to determine the nature of transactions that produce a $0 taxable amount in the accounting system.

States may also determine a floor amount beneath which items are neither sampled nor detailed. Stratum 1C contains 3,983 transactions with a taxable amount of $151,103.60. Assuming an 8% tax rate and that tax was not paid on these transactions and should have been, the maximum assessed amount is (0.08) x ($151,103.60) = $12,088.29. With an expected error rate of 10%, the expected assessment amount is $1,208.83 which is considered *de minimis* from an audit sampling perspective.

A ceiling amount is also typical in sales and use tax audits above which all items are detailed; that is, a complete audit census is completed for these items. In the sampling plan in Table 1, the ceiling is set at $75,000, and all 168 transactions with amounts greater than $75,000 are detailed.

Finally, there are two sample strata in this plan, Stratum 1D and Stratum 1E. Estimates of the strata population taxable adjustment amounts (i.e., taxable amounts in error) are made from the random samples drawn from these two strata. The estimated adjustment amounts are added to the adjustment amount determined in Stratum 1F to determine the audit assessment or refund adjustment amount for reporting purposes.

3. IRS Sampling and Estimation Methodologies

Without loss of generality, IRS sampling and estimation methodologies are described for variables sampling in stratified random sample designs.

The IRS applies a 95% confidence bound estimate for reporting the results of a sample audit. For the combined ratio and combined regression estimators, the following requirements must be met “in order to demonstrate that little statistical bias exists, the following applies after excluding all strata tested on a 100%” (i.e., detail) basis [29]:

- The total sample size for all strata must be at least 100 units.
- Each stratum for which a population estimate is made should contain at least 30 sample units.
- The coefficient of variation of the taxable amount variable must be 15% or less.
- The coefficient of variation of the adjustment amount variable must be 15% or less.
- For the combined ratio, the reported values of the units must be of the same sign.

The best estimator among the mean per unit, difference, combined ratio, and combined regression estimators is determined to be the estimator with the smallest estimated sampling error that meets any demonstration of negligible statistical bias requirements.

An estimator should be considered indeterminate if the estimated sampling error is greater than the absolute value of the point estimate. If an estimator is indeterminate, then the proposed adjustment is equal to zero.

The formulas for the confidence bound estimates of the generally accepted audit sampling estimators are can be found (in addition to many other resources) in the IRS Revenue Procedure 2011-42 [29] and Cochran's *Sampling Techniques* [6]. (Initially they were included in the body of this paper but since different nomenclature exist among publications and software, though with the same statistical estimations, the above literature and guidance has been recommended.)

Among these six estimators, the difference estimator almost always performs worse than the other five estimators and generally by a significant margin. The reason is due to the percentage of items in the population requiring adjustment. In most federal tax audits, the adjustment percentage is relatively small ranging from 5% to 25%. The difference estimator will perform well only if the adjustment percentage is high, generally 80% or greater.

Based on general statistical sampling experience with numerous tax sample audits using the generally accepted audit sampling estimators, the combined regression estimator performs best in the majority of audits in terms of minimizing the estimated sampling error. However, excluding the difference estimator, the differences in the estimated sampling error among the other five estimators generally immaterial and negligible.

The IRS guidance involves the computation of the 95% confidence bound estimates in its sample audits for the estimators. If the relative precision is less than 10%, then the point estimate is used to report the audit results. If the relative precision is greater than 15%, then the confidence bound estimate is used. If the relative precision is between 10% and 15%, a linear interpolated value between the point and confidence bound estimates is used.

When the confidence bound estimate is used for an IRS audit, the least favorable bound to the taxpayer, either the lower bound or the upper bound, is used. Consider, for example, a federal excise tax audit where an underpayment of tax for a transaction is recorded as a positive amount and an overpayment of tax for a transaction is recorded as a positive amount. Assume in a sample audit that the relative precision turns out to be 20% requiring the 95% confidence bound estimate to be used. Suppose that the point estimate represents a net tax underpayment of $100,000 with a lower 95% confidence bound estimate of $80,000 and an upper 95% confidence bound estimate of $120,000.

The least beneficial confidence bound estimate is the upper 95% confidence bound estimate of $120,000. Alternatively, suppose that the point estimate represents a net tax overpayment of $100,000 with a lower 95% confidence bound estimate of $80,000 and an upper 95% confidence bound estimate of $120,000. The least beneficial confidence bound estimate is now the lower 95% confidence bound estimate of $80,000.

The formulas for the confidence bound estimates for the six generally accepted audit sampling estimators (inclusive of the four IRS estimators) can be found in the IRS Revenue Procedure 2011-42 [29] and Cochran's *Sampling Techniques* [6]. The confidence bounds are based on the t-distribution with the degrees of freedom estimated using the Welch-Satterthwaite approximation [19], [23].

4. Audit Sampling Procedures Used by States in Sales and Use Tax Sample Audits

States conduct sales and use tax audits to determine tax compliance for sales and purchases made by firms with operations in the states. There are five states that do not have state sales tax: Alaska, Delaware, Montana, Oregon and New Hampshire. The forty-five states that have state sales tax typically conduct sales and use tax audits for firms on a three or four year cycle. Larger firms have voluminous sales and purchase transactions requiring audits based on samples to determine tax compliance.

When states began to use sampling methods for sales and use tax audits, very few states employed statisticians as members of the departments of taxation or departments of revenue. As a result, states often turned to professors in statistics departments to provide assistance in developing audit sampling procedures. Some states modeled their audit sampling procedures based on procedures used by other states that had experience with audit sampling methods.

Initially, most states began their ventures into audit sampling by using block samples where blocks of time during the audit period were chosen for detailing all transactions within the blocks. For illustration purposes, three months may have been chosen in a thirty-six month audit period and all transactions in selected accounts during those three months were audited. The sample audit results for those three months were "projected" to the population covering thirty-six months by multiplying the sample results by a factor of twelve (36/3 = 12). In the late 1980's and into the early 1990's, some states began to use stratified random samples for sample audits in addition to using block samples. Notably, these states had access to consultation advice from university professors in developing sampling methods based on random sampling. These states included California, Florida, Ohio, Tennessee and Washington.

Since states conducting sales and use tax audits did not have an organization to assist with developing common audit sampling methods used by all states, the states were on their own to develop audit sampling procedures. Consequently, there is a wide range of sampling methods, estimation methods, and audit sampling policies employed by the forty-five states conducting sales and use tax audits.

Most of the states selected either the separate ratio estimator or the mean per unit estimator for audit sampling primarily due to the simplicity of training auditors and explaining the estimation methods to taxpayers under audit. The separate ratio estimator had appeal due to the ease of explanation: Form the ratio of the net dollars in error in the sample to the total dollar amount in the sample and multiply this ratio by the total dollar

amount in the population. The mean per unit estimator is also easy to explain to a taxpayer: Calculate the average sample error dollar amount (total dollar amount in error divided by the sample size) and multiply this average by the number of items in the population. The regression estimators were avoided at least initially due to their comparative complexity and the challenges of explaining the estimators (and the more advanced variance calculations) to taxpayers under audit when not familiar with the science of statistical sampling.

Samples in audit sampling are referred to as statistical samples or non-statistical samples. Statistical samples use a measure of sampling risk (e.g., confidence bound, confidence interval, relative precision) whereas non-statistical samples do not provide a measure of sampling risk. The distinction between statistical and non-statistical samples appears to be first due to Arkin [3]. Texas is the largest state not using statistical sampling. In Texas sample audits based on stratified random samples, the final audit report is based on the separate ratio point estimate with no measure of sampling risk provided.

The state of Washington has the most statistically developed audit sampling policies and procedures among states. The four primary IRS estimators are used: mean per unit, difference, combined ratio, and combined regression. In addition, Washington bases the audit results on an 80% confidence bound estimate and not on the point estimate.

Table 2 provides a comparison of audit sampling policies and procedures for the forty-five states with state sales tax. The blank cells in the table are due to insufficient information provided by the states or information that is currently dated. Additional research is required to complete the information in the table. (Table 2 represents information collected year end of 2017 and is generally up to date.)

Based on reviewing Table 2, it is immediately apparent that audit sampling policies and procedures employed by states for sales and use tax vary considerably. Most, but not all states, have gravitated to using stratified random sample designs when audit populations are situationally large (e.g., 10,000 transactions or more). However, most of these states continue to use block sampling methods in some of their audits. The typical block sampling plan based on time periods uses a census of three to six months selected from thirty-six to forty-eight months in the audit period. Block samples to not provide measures of sampling risk but can be convenient, though non-scientific, from an auditing viewpoint if transaction records are in print form and are stored in chronological order.

Also, it is apparent from Table 2 that most states use the mean per unit and separate ratio estimators for sample audits. Both estimators are easy to use and to understand which is particularly important in terms of the ability of state auditors to explain the methods to taxpayers. The mean per unit point estimate is determined by calculating the average sample audited adjustment amount and multiplying this average by the total number of transactions in the population. The separate ratio point estimate is determined by finding the ratio of the audited adjustment amount in the sample to the total transaction amount in the sample, then multiplying this ratio by the total transaction amount in the population. Most state auditors and taxpayers can easily understand these two estimates but that is not the case for the comparatively increased complexity of the regression estimators and the combined ratio estimator.

The number of states that continue not using the science and mathematical solutions inherent in statistical audit sampling methods is both surprising and troubling. In sales

and use tax audit sampling, “statistical sampling” applies to sampling methods that produce a measure of sampling risk (e.g., confidence interval estimate, confidence bound estimate, relative precision estimate) while “non-statistical” sampling methods provide no measure of sampling risk. Seventeen states do not use statistical sampling methods, most notably Texas given the size of the state and the number of sales and use tax audits conducted annually by the Texas Comptroller of Public Accounts. A non-statistical sales and use tax audit sample may produce, for example, an estimated tax assessment of $300,000 with no indication as to the precision of the estimate. If a 90% confidence interval ranged between $280,000 and $320,000, a taxpayer most likely would feel comfortable in accepting the point estimate. If, on the other hand, the 90% confidence interval ranged from -$100,000 to $700,000, the taxpayer understandably should have grave doubts about accepting the point estimate adjustment as the result of the audit. In this sense, it is surprising that so many states continue to use non-statistical sampling methods.

Six states now use a confidence bound instead of the point estimate for reporting the results of a sales and use tax audit. In addition to these six states, Wisconsin has recently chosen to use the confidence bound estimate and another state is now in the process of considering a policy change to permit the use of a confidence bound estimate.

The minimum stratum sample size varies among states, though most states are now settling on a minimum of 200 or 250 transactions per stratum. Some states with minimum stratum sample sizes not reported in Table 1 have interesting approaches to choosing strata sample sizes. Virginia requires a minimum overall sample size of 1,000 transactions. If there are four sample strata, the minimum stratum sample size then becomes 250. If there are five sample strata, the minimum stratum sample size is then 200. Florida requires the stratified random sampling design to have seven sample strata and one detail stratum determined by setting an upper detail stratum dollar amount. These parameter settings do not consider the population size, the population variation or the expected error rate. The minimum stratum sample size was originally set at 30, but now operationally is set at 100.

Some states require a minimum number of errors per stratum to project the sample results in the stratum to the population. Typically, the minimum number of errors is set at 3 per stratum (e.g., California, Washington). If the minimum is not achieved, the auditor has two options: (1) increase the sample size, or (2) do not project the sample results in the stratum to the corresponding stratum population. Texas permits sample expansion in multiples of the original stratum sample size if there is evidence that the results in the sample may not be representative of the corresponding population characteristics.

The variation of state audit sampling policies and procedures presents challenges to firms that may have sample audits from several states occurring at the same time. Harmonizing state audit sampling policies and procedures for sales and use tax audits would simplify the audit process to taxpayers and consequently reduce audit expenses for both states and taxpayers.

There are additional differences in state audit sampling policies that create unnecessary complexity and frustration among taxpayers subjected to state sales and use tax audits. Twenty-two states do not permit the projection of tax overpayments to the audit population in state-initiated audits. Of course, tax underpayments are projected thus leading to tax assessments issued to the taxpayer. In these states, the taxpayer must file

refund claims for the tax overpayments based on a census of transactions identified as containing overpayment tax errors. This policy seems not to be equitable and fair. Rather, it would seem as if the objective of the state sales and use tax sample audit should be to determine the correct amount of tax due such that the sample audit result should be a net tax assessment or a net tax refund.

Table 2: State Audit Sampling Policies and Procedures

State	Type of Samples Used	Estimator Used	Statistical Sampling	Confidence Bound Used	Minimum Stratum Sample Size
Alabama	block; stratified	mean; separate ratio	yes, 90% CI	yes, 90%	
Arizona	block; stratified	separate ratio	no	no	
Arkansas	block; stratified	mean; separate ratio	yes, 80% CI	no	
California	block; stratified	mean	yes	no	300
Colorado	block; stratified	difference, mean, combined ratio, combined regression	no	no	
Connecticut	block; stratified	separate ratio	yes, 80% CI	no	
Florida	block; stratified	mean; separate ratio	yes, 95% CI	no	100
Georgia	block; srs; stratified	mean; separate ratio	no	no	
Hawaii	block	separate ratio	no	no	
Idaho	block; stratified	mean, separate ratio, separate regression	no	no	
Illinois	block; stratified	difference, mean, combined ratio, combined regression	yes, 90% CI	no	200
Indiana	block, stratified	mean, separate ratio, separate regression	yes, 90% CI	no	
Iowa	block; srs; stratified	mean	yes, 80% CI	no	250
Kansas	block; srs; stratified	difference, mean, combined ratio, combined regression	yes	yes; 80%	
Kentucky	block, srs	separate ratio	no	no	250
Louisiana	block; srs; stratified	mean; separate ratio			
Maine	block; systematic	separate ratio	no	no	
Maryland	block; srs; stratified	difference, mean, separate	yes, 90% CI	yes, 90%	
Massachusetts	stratified	separate ratio	yes, 95% CI	no	
Michigan	block; stratified	separate ratio	yes, 90% CI	no	
Minnesota	block; srs; stratified	mean, separate ratio, combined ratio	yes, 95% CI	no	
Mississippi	block; stratified	separate ratio	no	no	
Missouri	block; srs; stratified	difference, mean, combined ratio, combined regression	yes, 95% CI	no	
Nebraska	block; stratified	separate ratio	yes, 95% CI	no	
Nevada	block; stratified	separate ratio	no	no	
New Jersey	block; stratified	separate ratio	no	no	
New Mexico				no	
New York	block, stratified	mean	yes, 90% CI	no	100
North Carolina	block; stratified	separate ratio	yes, 95% CI	no	
North Dakota	block; srs; stratified	mean		no	
Ohio	block; stratified	difference, mean, combined ratio, combined regression	yes, 90% CI	no	
Oklahoma	block; stratified	separate ratio	no	no	
Pennsylvania	block; stratified	separate ratio	yes, 90% CI	no	200
Rhode Island	block; stratified	separate ratio	no	no	
South Carolina	block; stratified	difference, mean, separate ratio	yes, 95% CI	no	
South Dakota	block; srs; stratified	mean; separate ratio	no	no	
Tennessee	block; stratified	mean	yes	yes, 75%	70
Texas	block; stratified	separate ratio	no	no	100
Utah	block; srs; stratified	mean	yes	yes; 80%	
Vermont	block; srs	mean; separate ratio	no	no	
Virginia	block; stratified; systematic	separate ratio	no	no	
Washington	block; stratified	difference, mean, combined ratio, combined regression	yes	yes; 80%	250
West Virginia	block; stratified	separate ratio	yes, 95% CI	no	
Wisconsin	block; stratified	mean	yes, 90% CI	no	
Wyoming	block; srs; stratified	mean; separate ratio	no	no	

Only ten states permit refund claims initiated by taxpayers to be based on sample audits. Texas is the most notable state in this group since Texas has a statute (151.430) specifically permitting refund claims based on samples and has an extensive sampling manual that among other things, delineates in considerable detail the sampling and estimation methods the taxpayer or the taxpayer's consultant must use when basing refund claims on samples. Interestingly, for the states not permitting taxpayer initiated refund claims based on samples, auditors in these states frequently sample the refund claims based on a census of transactions identified by the taxpayer as containing tax overpayment errors.

The Multistate Tax Commission ("MTC") offers four-day training workshops on audit sampling that cover the IRS sampling and estimation policies and procedures. The MTC has developed a software program that produces point estimates and confidence bound estimates for four estimators that the IRS guidance also addresses. Bob Schauer at the MTC originally developed the software for the State of Washington when he was a member of the audit sampling unit for the Department of Revenue for the state. When he joined the MTC, he continued to develop and modify the software which largely conforms to the IRS audit sampling policies and procedures. Several states have had their audit sampling specialists attend the MTC workshop and are currently considering or using the MTC audit sampling software for their sales and use tax audits. The MTC also promotes supporting sales and use tax audits for two or more states simultaneously. While combining state audits has obvious benefits from a resource efficiency standpoint, tax practitioners see the difficulty with multiple state audits is the challenge of allocating the audit results to the participating states equitably particularly if audit results are based on confidence bound estimates.

Finally, states use a variety of audit sampling software to conduct sales and use tax sample audits. Increasingly, it appears that states are starting to use the MTC software package in place of previously chosen audit software selections. Currently states are using the following software packages: MTC software package, IDEA, ACL, TSEPWin, SAS, and some type of legacy software programs

IDEA is an audit software package produced and marketed by CaseWare Analytics with Audimation as its US distributor (casewareanalytics.com). The software emphasizes the importation of tax audit populations in various formats including print reports, and the manipulation and analysis of the population data. The software does not produce the six generally accepted audit sampling estimators though a programmer or statistician can script statistical sampling macros to produce these estimates.

ACL is comparative to IDEA and is marketed by ACL Analytics (acl.com). ACL and IDEA are competing products with very similar characteristics. These two software products continue to be used by states for sales and use tax sample audits but lack the prepackaged functionality of deep statistical sampling estimation.

TSEPWin was developed by Dr. Richard Kulp for Tennessee and its sales and use tax sample audits. The software is now available for free from the Multistate Tax Commission (mtc.com). Several states in addition to Tennessee have adopted TSEPWin for sales and use tax sample audits. The software produces point and confidence bound estimates for the six generally accepted audit sampling estimators. In addition, the software offers several advanced estimators including empirical likelihood estimation methods.

SAS is a sophisticated statistical software package that provides extensive statistical methods for analyzing data. Scripting in SAS can be used to produce point and confidence bound estimates for the six generally accepted audit sampling estimators. Due primarily to cost and secondarily due to complexity, no states are currently using the software for sales and use tax sample audits. However, this is the software frequently used by statistical sampling groups focused on federal tax sampling.

There are now only a handful of states (five or fewer) using legacy software developed for sales and use tax audit sampling. Most of the legacy software programs were written in Cobol or Fortran.

There are open source softwares such as Python and R that are advanced and versatile allowing for the development of deep statistical sampling and estimation suites. The use of these softwares are frequently seen in academia consulting but are catching on in many environments because of the associated low cost but can require staff statisticians to maintain the statistical scripts and interrupt outputs.

A challenge for taxpayers with sample audits conducted by multiple states is learning and using a variety of sample audit results produced by these software products. Interpreting the results can be a challenge due to formatting and reporting methods and different styles of statistical notation used by these software packages.

5. Comparison Between IRS and State Audit Sampling Policies and Procedures

There are significant differences between audit sampling policies and procedures used by the IRS and by states for sample audits. These differences are *generally summarized* by sampling plan characteristics, choice of estimators, and choice of audit sampling software in this section of the paper.

- Minimum strata sample sizes in stratified random samples
  a) IRS: 30 (based on Central Limit Theorem)
  b) States: Typically, 100 or more with most states using 200 to 250. The requirement of larger strata sample sizes used by states for sales and use tax sampling audits than used by the IRS is due to typically low error rates among certain classes of transactions in tax sample audits. If the error rate is 1%, then 1% of 30 sampled items (IRS minimum) is only 0.3 item.
  c) States: Several states (e.g., CA, WA, UT, MN) require a minimum of 3 errors in each sample stratum to estimate the error amount in the stratum population. The purpose of this requirement is to ensure sufficient information on the error process to estimate the population error dollar amount with acceptable precision.

- Number of sample strata in stratified random samples
  a) IRS: there is not requirement on the number of strata and is situationally determined. Generally, between five to seven sample strata (but can range substantially) is typically appropriate following the guidance on the maximum number of sample strata described in Cochran [6], pp. 132-134.
  b) States: Most states use between two and seven sample strata, but occasionally use more than seven in their stratified random sample designs. Tennessee is an outlier due to using typically more than seven sample strata and up to twenty sample strata in its designs.

- Choice of point estimators
  a) IRS: The IRS guidance uses four estimators but should compute the estimates for the separate ratio and separate regression estimators for comparative and informational purposes. In typical applications of the four estimators the IRS uses for reporting sample audit results, the combined regression estimator is frequently the best estimator of sample audit results.

b) States: Most states use separate ratio or mean per unit estimators. These two estimators are accessible for auditors and taxpayers.

- Confidence coefficient for confidence bound estimates
  a) IRS: The IRS uses 95% to construct either the lower or upper confidence bound. As a side note, the t-value for a 90% confidence interval (i.e., two-sided) is the same as a 95% confidence bound (i.e., one-sided) making the upper and lower internal *limits* equivalent to the lower and upper confidence *bounds*.
  b) States: There are six states currently using confidence bounds to report the results of sales and use tax sample audits. The confidence coefficient used to construct the bounds varies by state. Tennessee uses 75%; Washington, Utah and Kansas use 80%; and Alabama and Maryland use 90%. States want the confidence bounds closer to the point estimates (which is the "truest" and statistically most likely and accurate estimate) than a 95% confidence bound would produce. This also accommodates relatively heterogeneous populations regularly found in state sales and use tax sample audits without adding undo design complexities.

- Application of confidence bound estimates
  a) IRS: The IRS chooses either lower bound or upper bound that is least advantageous to the taxpayer.
  b) States: Assume that underpayments are positive and overpayments are negative. If the point estimate is net positive, the lower confidence bound is used (advantageous to taxpayer). If the point estimate is net negative, the upper confidence bound is used (advantageous to state).
  c) States: Tennessee is one of the six states that use the confidence bound for reporting sample audit results for sales and use tax audits. Tennessee has developed a decision rule for determining whether the result of the audit is indeterminate (no tax due and no refund). The rule: If the point estimate and the confidence bound estimate are on opposite sides of the origin, the sample audit is deemed indeterminate. Some states employing statistical sampling but not using confidence bounds will conclude that the sample audit is indeterminate if a confidence interval, usually set with 90% confidence, contains 0 between the lower and upper limits. It is interesting to note that this rule can be viewed as a hypothesis test with a 0.05 significance level, where the null hypothesis that the population adjustment (e.g., error) dollar amount is $0 cannot be rejected.

- Acceptable relative precision level
  a) IRS: The IRS target is a relative precision of 10% or less. If achieved, the point estimate rather than the confidence bound estimate can be used to report the sample audit results. If the relative precision falls between 10% and 15%, the reported estimate is the linear interpolated value between the point and confidence bound estimate. If the relative precision is greater than 15%, the confidence bound estimate is used to report the sample audit results. When the relative precision is relatively high (>25%), the taxpayer will often have to accommodate a large monetary difference between the point estimate and the confidence bound estimate.
  b) States: A target of a maximum relative precision set at 20% or 25% is typical with a confidence coefficient of 80% or 90%. With a relative precision of 25% with an 80% confidence level, the confidence bound estimate will be much closer to the point estimate than it would be by using the IRS 95% confidence level setting. If the relative precision exceeds the state's maximum, the auditor and the

taxpayer have two options: (1) Expand the sample size to achieve better relative precision, or (2) Proceed with the estimate though the relative precision exceeds the state targeted maximum value.

- Heterogenous, homogenous, and population size: As a note of interest, and not meant to distinguish between federal and state audit populations' homogeneity, the following example is given for sales and use tax.
  a) States: The populations derived for sales and use tax sample audits can be very heterogeneous with widely different error rates (e.g., assets, procurement card purchases, general expenses) and some populations are massive by most standards (e.g., imagining all of Walmart sales and purchases in Texas for a three-year period). Due to the heterogeneous nature of the audit populations, considerably more time and effort must be invested in crafting the sample design to insure efficient and effective results. Typically, the derived population for the audit is stratified qualitatively by attributes where most often the attribute used is the type of transaction to be audited. Some states (e.g., Texas) in purchase audits require asset purchases and general expense purchases to be placed in separate populations based on the expectation of significantly different error rates to the two types of purchase transactions. The belief is the asset purchases are subjected to greater review and scrutiny than general expense purchases, and therefore the asset purchases will have a lower error rate than the error rate for the general expense purchases. In purchase audits, several states recommend or require that taxed and non-taxed purchases be placed in separate populations. This makes statistical sense because the taxed purchases when audited will produce mostly overpayments as errors while the non-taxed purchases will produce underpayments as errors exclusively. As a means of further addressing remaining heterogeneity within the stratified populations based on attributes, numerical stratification is employed based on the expectation that lower dollar amounts will likely have larger tax error rates than higher dollar amounts.

- Audit sampling software choices
  a) States: The states use a variety of software packages for sales and use tax sample audits. These audit sampling software products were described and discussed in Section 3 in this paper. As previously noted, the MTC software closely mirrors the IRS sampling and estimation methods. The MTC software adoption and use appear to be on the rise with states due in no small part to the four-day MTC audit sampling workshop at which participants who complete the workshop receive a copy of the software without additional cost. The following four audit sampling software packages[4] are commonly used: IDEA, ACL, TSEPWin, and MTC audit sampling software.

6. Conclusions and Recommendations

The past field directive and the most recent IRS guidance on audit sampling established guidelines for conducting and evaluating IRS sample audits. These guidelines were developed to promote the efficiency and consistency of probability samples performed and examined by the IRS. Since the introduction of the past and the use of the current

[4] IDEA (www.casewareanalytics.com), ACL (www.acl.com), TSEPWin (www.mtc.com), and MTC audit sampling software (www.mtc.com).

guidance, the taxpayer audits have proven to be effective and successful in establishing consistency and efficiency in federal tax samples and estimations.

The situation among states conducting sales and use tax sample audits is quite different from the success the IRS has achieved in developing its audit sampling guidance. By contrast, states have developed their own audit sampling policies and procedures often independently and without input from other states. Consequently, there is a wide variety of sampling and estimation methods employed by states for sales and use tax sample audits. Taxpayers involved with sales and use tax sample audits conducted by multiple states, often simultaneously, are confronted with different sampling and estimation methods, and different formats for reports summarizing the sample audit results, all of which can be confusing and frustrating for the taxpayer. In some instances, audit sampling and estimation methods employed by states may be in conflict; that is, one state may treat a transaction with a tax error quite differently than another state with direct implications for estimates of population net adjustment amounts produced by the two states. An obvious example of a significant and consequential difference is when one state uses a point estimate to report sample audit results when another sate uses a confidence bound estimate for reporting purposes

From the taxpayer viewpoint, having common and harmonious audit sampling policies and procedures for conducting sales and use tax sample audits is desirable given the obvious savings in audit costs and sample audit training. The four-day workshop on audit sampling offered by the Multistate Tax Commission has played an important role in harmonizing audit sampling procedures in large measure due to providing workshop participants with the MTC audit sampling software. Increasingly, more states are adopting the use of the MTC audit sampling software with a current count of ten or so states currently using the software. In addition, the MTC has published an extensive and comprehensive sampling manual (*Sampling Policy & Guideline Manual* [19] and a procedures document (*Statistical Sampling Procedures* [23] that represent significant contributions to harmonizing audit sampling procedures for sales and use tax audits.

The Streamlined Sales Tax ("SST") initiative was created in 2000 to harmonize methods used by states for sales and use tax audits (www.streamlinedsalestax.org) with initial attention focused on tax determination issues (e.g., What are the ingredients and percentage of ingredients that define a beverage as fruit drink for tax purposes?). Eventually, the SST may turn its attention to harmonizing audit sampling policies and procedures.

In the interim, recommending to states to accept certain core audit sampling procedures makes sense from a cost-benefit analysis and to leverage the statistical science that is readily accessible to taxpayers and regulatory agencies. Because statistical techniques are mathematically derived and scientifically developed, the general sampling principles can easily be adapted to a multiplicity of circumstances. With that understanding, the following is a list of a reasonable but not exhaustive set of recommendations for states to consider for audit sampling.

The IRS guidance published in 2011 [29] and the MTC sampling manual revised in 2008 [19] provide the basis for harmonizing audit sampling policies and procedures for sales and use tax audits.

The IRS guidance provides extensive guidelines for developing sampling plans, generating samples according to the plans, estimating the population adjustment amount and documenting the sampling procedures and results. However, the IRS guidelines focus on federal tax and that can have characteristics generally different than sales and use tax sample audits. There are two guidelines in the IRS guidance that require adjustment for sales and use tax sample audits.

- The setting of the minimum stratum sample size of 30. For sales and use tax sample audits, a minimum stratum sample size of 200 to 250 items[5] is more reasonable given generally low adjustment error rates in these sample audits.
- The confidence coefficient setting of 95% for the construction of confidence bounds is too high for typical sales and use tax sample audits. States that are using confidence bounds for sales and use tax sample audits are using confidence coefficients that range from 75% to 90% with an 80% confidence coefficient as the most commonly selected confidence coefficient setting.

The MTC audit sampling manual [19] and audit sampling procedures document [23] are directed specifically for application to sales and use tax sample audits. The information in these references are used as the core instruction information and materials for the MTC four-day training program in statistical audit sampling for sales and use tax sample audits. Accordingly, most MTC suggested procedures in the its guidelines make sense in working toward the harmonization of the sampling policies and procedures used by states for sales and use tax sample audits. The following list of guidelines constitute a core set for adoption by states in the quest harmonize sampling policies and procedures for sales and use tax sample audits.

- Derivation of audit sampling population
  a) Match as many negative valued and positive valued transactions as possible. Extract the remaining negative valued transactions, and sample only the remaining positive valued transactions.
  b) Remove items in the initial population that are not in the scope of the audit.
  c) Remove items likely not to be in error.
  d) Set a high dollar detail threshold (“ceiling”) such that transactions with amounts exceeding the detail threshold dollar amount are detailed (i.e., not sampled).
  e) Set a lower dollar exclusion threshold (“floor”) such that transactions with amounts less than the lower dollar exclusion threshold are not ignored and not audited.
- Generally, use three to five strata but no more than seven. Additionally, guidance on the maximum number of sample strata described is seen in Cochran [6], pp. 132-134.
- Determine strata boundaries using the cumulative square root of the frequency method [6] or other generally accepted audit sampling stratification methods [28] such as the equal dollar amount [15] or the equalization of the product of the weight and standard deviation [6].
- Use strata sample sizes in the range from 200 to 400 transactions per stratum. Use a goal of 30% relative precision or better to determine strata sample sizes.

[5] There are statistical formulae that estimate a sample size based on assumed parameters and population characteristics. The minimal 200 – 250 is a general guideline based on a common error rate and the 3 to 5 strata frequently being used in current practices.

- Use the four estimators that the IRS uses (mean per unit, difference, combined ratio, and combined regression).

States should consider using the confidence bound estimate rather than the point estimate to report the sample audit results.

- For confidence bound estimation, use a confidence coefficient of 80% rather than the 95% setting used by the IRS. This accommodates the variation frequently found in audit sampling without adding undo design complexities and provides a confidence bounds closer to the point estimates than a 95% confidence bound would produce.
- Do not adopt the IRS guidance of choosing the confidence bound estimate that is least favorable to the taxpayer. Rather, use the policy already adopted by states that are using confidence bound estimates (e.g., Washington, Utah, Tennessee) which is to use the lower confidence bound for sample audits with an estimated net assessment (favorable to the taxpayer), and use the upper confidence bound for sample audits with an estimated net refund (favorable to the state).

The purpose of these recommendations is to create an open dialogue encouraging states to accept core audit sampling principles, polices, and procedures for the joint benefit of states and taxpayers when sales and use tax audit samples are employed.

Acknowledgements

A singular recognition goes out to co-author Roger Pfaffenberger who passed away prior to the completion of this work and requested its publication in a JSM proceedings. Our effort on this research began in 2017 and was originally presented on at the 2017 JSM. The initial research we determined to be interesting and exceptionally relevant to further explore what a standardization of sampling methods among the tax regulatory agencies might look like and has formed the foundation of this more extensive publication. Roger and I worked on this paper until July 2019 and minimal revisions have been made to it since then for its current publications in the 2020 JSM proceedings.

References

[01] Audit Manual, Sales and Use Tax Department, California Board of Equalization, 2002. Accessible online at: www.boe.ca.gov. (See Chapter 13, Audit Sampling.)
[02] Arkin, Herbert (1982). Sampling Methods for the Auditor: An Advanced Treatment. New York: McGraw-Hill.
[03] Arkin, Herbert (1984). Handbook of Sampling for Auditing and Accounting, 3rd ed. New York: McGraw-Hill.
[04] Arkin, Herbert and James K. Loebbecke, (1981). Applications of Statistical Sampling to Auditing, Englewood, CA: Prentice Hall, 1981.
[05] CAATS (Computer Assisted Audit Techniques) brochure, Department of Revenue, Massachusetts.
[06] Cochran, W.C. (1977). Sampling Techniques, 3rd ed. New York: John Wiley & Sons.
[07] Computer-Assisted Audits: Guidelines and Procedures for Sales Tax Audits. Publication 132, New York State Department of Taxation & Finance.
[08] Electronic Records and Computer-Assisted Auditing, Publication 107, Illinois Department of Revenue.

[09]Guy, Dan M., Douglas R. Carmichael, and O. Ray Whittington. (1998). Audit Sampling: An Introduction. 4th Edition. New York: John Wiley & Sons.
[10]Guy, Dan M., Douglas R. Carmichael, and O. Ray Whittington. (1977). Practitioner's Guide to Audit Sampling. New York: John Wiley & Sons, 1977.
[11]Field Guidance on the Use of Estimates from Probability Samples. (2002). Internal Revenue Service, Department of the Treasury.
[12]Lohr, Sharon. (1999). Sampling: Design and Analysis. Boson: Duxbury Press
[13]Moors, J.J.A. (1983). "Bayes' Estimation in Sampling for Auditing". Journal of the Royal Statistical Society, Vol. 32, No. 3, pp. 281-288.
[14]Pfaffenberger, Roger and Adina Christian, (2012), "Statistical Sampling – A Step Toward Smarter Government", Journal of State Taxation, March-April 2012, pp. 53-56.
[15]Roberts, Donald A. (1978). Statistical Auditing. New York: American Institute of Certified Public Accountants.
[16]Sahu, S.K., and T.M.F. Smith (2006). "A Bayesian Method of Sample Size Determination with Practical Applications". Journal of the Royal Statistical Society, Vol. 169, pp. 235-253.
[17]Sampling for Sales and Use Tax Compliance, A Report of the Steering Committee, Task Force on EDI Audit and Legal Issues for Tax Administration. Washington, D.C.: Federation of Tax Administrators, 2002, Updated January 2004. (FTA website: www.taxadmin.org. (See Publications Section on the site.)
[18]Sampling Manual, Texas Comptroller's Office, Austin Texas.
[19]Sampling Policy & Guideline Manual, Washington, D.C.: Multistate Tax Commission, July 2008. (See MTC website: www.mtc.gov.)
[20]Satterthwaite, F.E. (1946). "An Approximate Distribution of Estimates of Variance Components," Biometrics Bulletin Vol. 2, pp. 110-114.
[21]Scheaffer, Richard, William Mendenhall, and R. Lyman Ott. (1996). Elementary Survey Sampling.5th Edition. Boston: Duxbury Press.
[22]Statistical Sampling Manual, Department of Revenue, State of Washington.
[23]Statistical Sampling Procedures, Washington, D.C.: Multistate Tax Commission, September 2012. (See MTC website: www.mtc.gov.)`
[24]Task Force on EDI Audit and Legal Issues for Tax Administration (2002). Sampling for Sales and Use Tax Compliance. Washington, D.C.: Federation of Tax Administrators.
[25]Welch, B.L. (1947), "The Generalization of "student's' Problem When Several Different Population Variances are Involved," Biometrika Vol. 34, pp. 28-35.
[26]Yancey, Will (2002). Statistical Sampling in Sales and Use Tax Audits. CCH – Walters Kluwer.
[27]Yancey, Will and Roger Pfaffenberger (1997). Use and Abuse in Sampling and Use Tax Audits. Council on State Taxation State Tax Report, November 1997.
[28]Rhyne, Z. & Pfaffenberger, R. (2019), "Construction of Strata Boundaries in Tax Auditing," 2019 Proceedings of the American Statistical Association.
[29]Rev. Proc. 2011-42, 2011-37 I.R.B. 318 (September 12, 2011).
[30]Falk, E., Petska, R., & Rhyne, Z. (2008), "How to help businesses make regulatory sampling decision," 2008 Proceedings of the American Statistical Association.