1

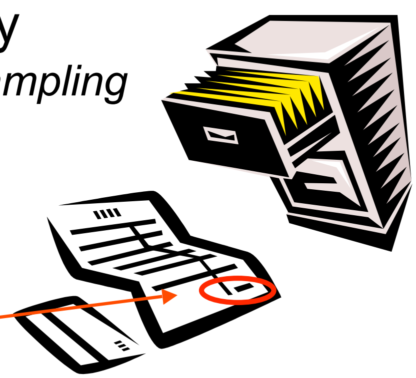
Comments on Stratification in Sales & Use Tax Sampling

Computer Technology Workshop
September 2007

Bob Schauer
Multistate Tax Commission
360-666-3119
rschauer@mtc.gov

2

Terminology
Sales & Use Tax Sampling



- *Invoice Amount*
  - *Recorded Amount*
  - *Examined Amount*
  - *Book Value*
  - known
- *Error Amount*
  - *Difference(s)*
  - *Tax error*
  - *Taxable error*
  - unknown being estimated

Terminology...

3

- *simple random sampling*
  – One random sample drawn from the population
  – May have more two or more strata, but
    - a *probability sample* is drawn from only one
- *stratification*
  = *stratified random sampling* for this discussion
  – Independent random sample taken from two or more strata

Terminology...

4

- Collections of population units:
  – *Strata*
    - My preference
  – *Groups*
  – *Subgroups*
  – *Subpopulations*
  – *Clusters*
    - Not really the same

Terminology…

5

- Strata
  - Plural (more than one)
- Stratum
  - singular

- Stratas
  - Incorrect plural
- Stratums
  - my dictionary says OK, but maybe “strata” is better…

Terminology…

6

- Stratum…
  - basic definition “layer”
  - from Latin, “spread” or “cover”

- Create your own layers
  - use invoice amounts!
  - or, something else that makes sense

Basic Guidelines

7

- Each population unit must belong to only one stratum
- Define strata before sampling
- The number of units in each stratum must be known before sampling
- Once strata established, sample size should be *optimally* distributed to the strata (*Neyman Allocation*)
- Take an independent random sample from each stratum

Neyman Allocation example: Attachment #6

Basic Questions about Stratification

8

- Is it a good idea?
- If it is a good idea, how come this is a good idea (why)?
- How do I stratify?
- How many strata?
- Related questions
  - Overall sample size
  - Distribution of the sample to the strata

Reasons to Stratify

9

- Improve accuracy
  - Usually the reason
  - Does not always result in increased accuracy
    - Criteria used to stratify might have nothing to do with the errors
- Find out something about the individual "layers"
  - Usually not the emphasis
- Legal reasons (& other)
  - State of Minnesota has a court case

Gains in Accuracy (usually) Realized if:

10

- With regard to the criteria used to stratify
  - the elements within each stratum are similar
  - the strata are dissimilar
  - it is correlated with the error amount in the population

Overall *Standard Error*

in a Stratified Population

11

- An estimate the total error amount in a population…
  - is likely precise if the standard error is small
  - is likely not precise if the standard error is large
- Standard errors exist for the total
  - *Invoice Amounts* (known)
  - *Error Amounts* (unknown)
- The standard error of the total *invoice amounts* & the standard error of the total *error amounts* is often…
  - Interrelated

Attachments #1 & 2

Standard Error…

12

- Standard Error influenced by many factors
  - some controllable
  - others not controllable
- Some controllable factors:
  - Overall sample size
  - How overall sample is allocated toward the strata
  - How the population is placed into strata
    - Goes to method of stratifying
  - Number of strata

Standard Error…

13

- Yet another (big) question…
- If the standard error of the *invoice amounts* is (normally) interrelated with the standard error of the *error amounts*, how often will one stratification method have the lowest standard error for both?

Stratifying on Invoice Amount Methods

14

- Usually by determining *break values* between the strata
  - Example (population with invoices from $0 to $1000):
    - Break values (*stratum boundaries*) of:
      - $0
      - $50
      - $150
      - $350
      - $900
      - $1000
    - Stratum 1: Values from $0 and below $50,
    - Stratum 2: Values from $50 and below $150,
    - Stratum 3: Values from $150 and below $350,
    - Stratum 4: Values from $350 and below $900,
    - Stratum 5: Values from $900 to $1000

Stratifying … Methods… 15

- Determining breaks by "eye" or some rough method
- Basic method
  - described in Roberts book at pages 97-98
  - Equal Invoice Amount in each stratum
  - I call *proportional*
- Cumulative Square Root of the Frequency (*CSRF*) –
  - Cochran pages 127-131 (& a lot of other places)
  - Comes in two varieties:
    - Basic building block – the interval – is always equal (*CSRF-Equal*)
    - Basic building block – the interval – is varied (*CSRF-Unequal*)
- Method Proposed by Will Yancey and a few others – *Geometric Ratio*
- Others

Attachments #3, #4, & #5

Stratifying … Methods… 16

- Whole idea to bring down the standard error
- Different methods, with different break values, will have different standard errors
- In theory, the best is the one with the lowest standard error
- Standard error of the total *Error Amount* is unknown
- However, standard error total *Invoice Amount* is known
- These two standard errors are often interrelated

Stratifying … Methods…

17

- Possible factors in choosing method
  - Lowest standard error in invoice amounts (accuracy)
  - How easy is it to determine break values (implementation)
  - Works across all populations encountered (flexibility)
  - Can it be explained to taxpayers and auditors (explainable)
  - Does the method give a reasonable taxpayer or auditor a concern (acceptance)

Stratifying … Methods…

18

- More on Proportional
  - Considers all elements of the population
  - Very little about the population is assumed
  - Formulas are easy to determine break values
  - Easily programmable
  - Manual calculations take a bit of work
  - Break values are typically odd

Stratifying … Methods…

19

- More on CSRF
  - Considers all elements of the population
  - Very little about the population is assumed
  - Formulas are complex
    - unequal even more complex
  - Easily programmable
  - Manual calculations take a bit of work
    - Unequal even more work
  - Break values are typically uniform (not odd)
  - Sample allocation – or optimizing (Neyman Allocation)
    - Recommended for equal
    - Not required for unequal (already approximately equal if sample size across strata is the same)

Stratifying … Methods…

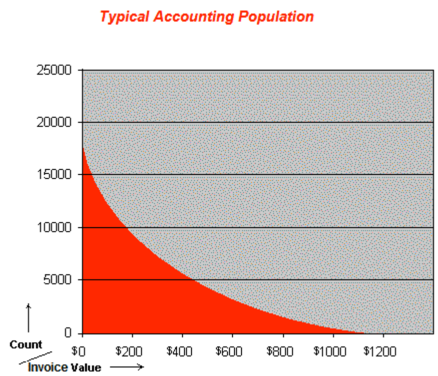
20

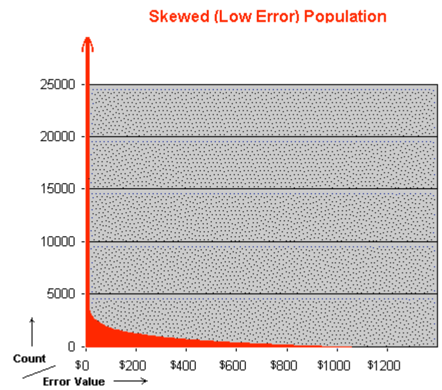
- More on Geometric Ratio
  - Does not consider all elements of the population, only the smallest and largest value
  - A distinct population distribution is assumed
  - Formulas are easy
  - Easily programmable
  - Manual calculations easy
  - Break values are typically very odd
  - Sample allocation – or optimizing – (Neyman Allocation)
    - required

Stratifying … Methods…

21

- More on Geometric Ratio…





Stratifying … Methods…

22

- Comparison of breaks and optimal sample size using one typical example with five strata, overall sample size of 500:

	Proportional		**Geometric Ratio**
1.	$10 to $63.39; sample 99	1.	$10 to $39.35; sample 19
2.	$63.40 to $135.97; sample 72	2.	$39.36 to $158.81; sample 189
3.	$135.98 to $262.50; sample 61	3.	$158.82 to $609.11; sample 201
4.	$262.51 to $565.90; sample 73	4.	$609.12 to $2,392.62; sample 81
5.	$565.91 to 9,429.72; sample 195	5.	$2,392.63 to $9,429.72; sample 10
	CSRF-Equal*		**CSRF-Unequal***
1.	$10 to $59.99; sample 89	1.	$10 to $49.99; sample 100
2.	$60.00 to $139.99; sample 95	2.	$50.00 to $149.99; sample 100
3.	$140.00 to $299.99; sample 90	3.	$150.00 to $349.99; sample 100
4.	$300.00 to $759.99; sample 95	4.	$350.00 to $899.99; sample 100
5.	$760.00 to 9,429.72; sample 131	5.	$900.00 to 9,429.72; sample 100

** "Equal" & "Unequal" refer to the interval used to construct strata (not shown), and not to the samples sizes by stratum – this is confusing as "Unequal" recommends using equal sample size across the strata, and "Equal" will almost always end up with unequal sample sizes!*

Stratifying … Methods…

23

- Comparison of methods

	Proportional	Geometric Ratio	CSRF Equal	CSRF Unequal
Accuracy (lowest standard error)	Don't know until each individual population is analyzed			
Implementation	Easy to program Manually - takes some work Optimizing required	Easy to program Manually - also easy Optimizing required	Easy to program Manually - can be tedious optimizing required	Easy to program Manually - can be very tedious optimizing not required
Flexibility	Don't know	Don't know	Don't know	Don't know
Explainable	Easy to explain & Explaining odd sample sizes not easy	Easy to explain & Explaining odd sample sizes not easy	Difficult to explain & Explaining odd sample sizes also not easy	Very difficult to explain No need to explain sample size calculations
Acceptance	Provides for unusual sample sizes and odd break values	Provides for unusual sample sizes and odd break values	Provides for odd sample sizes but uniform break values	Even sample sizes and uniform break values

Stratification Study

24

- Washington Department of Revenue
  - Actually me…
- Answer some of the questions
- Respond to some of the assertions made in a paper by Will Yancey
- Study available on request (talk to me afterwards)

Stratification Study…

25

- Started with 50 different files of invoice amounts
  - All from actual audits
- Created 8000 different error amount populations connected to these basic 50 populations
  - Random assignment of error

Stratification Study…

26

- Error Mixes:

Injected randomly into 50 populations
Each mix and frame repeated 10 times

Error Mix	Mix Type	FPE %	PPE %	FNE %	Combined %
1	a	2.0	0.0	0.0	2.0
2	a	4.0	0.0	0.0	4.0
3	a	8.0	0.0	0.0	8.0
4	a	20.0	0.0	0.0	20.0
5	b	1.0	1.0	0.0	2.0
6	b	2.0	2.0	0.0	4.0
7	b	4.0	4.0	0.0	8.0
8	b	10.0	10.0	0.0	20.0
9	c	1.5	0.0	0.5	2.0
10	c	3.0	0.0	1.0	4.0
11	c	6.0	0.0	2.0	8.0
12	c	15.0	0.0	5.0	20.0
13	d	1.0	0.5	0.5	2.0
14	d	2.0	1.0	1.0	4.0
15	d	4.0	2.0	2.0	8.0
16	d	10.0	5.0	5.0	20.0

Nonzero Errors are:

Full Positive Error (FPE)
Partial Positive Error (PPE)
Full Negative Error (FNE)

Attachment #7D

Stratification Study…

27

- Flexibility – the downfall of Geometric Ratio
  - Considers only the smallest and largest value
  - and nothing in between!
  - "Failure" means it was just not possible to use the method with the population of recorded amounts at hand
    - "Failure" in Equal and Unequal not really the end of the road, in that all one has to do is change interval sizes – not an option in Geometric Ratio

Strata	Sample Size	Populations	Geometric	Proportional	CSRF-E	CSRF-U
2	100	8,000	2%	0%	0%	0%
3	100	8,000	30%	0%	0%	0%
4	100	8,000	58%	0%	2%	0%
5	100	8,000	68%	0%	4%	0%
2	500	8,000	2%	4%	0%	0%
3	500	8,000	6%	12%	2%	0%
4	500	8,000	28%	14%	6%	4%
5	500	8,000	42%	14%	16%	6%

Stratification Study…

28

- Comparison of methods

	Proportional	Geometric Ratio	CRSF Equal	CRSF Unequal
Accuracy (lowest standard error)	Most of the time, this method had the largest standard error	Did well, but not quite is well overall Did as well or better when the error rate was high - better at 4-5 strata	Very close to Unequal	Did the best overall
Implementation	Easy to program Manually - takes some work Optimizing required	Easy to program Manually - also easy Optimizing required	Easy to program Manually - can be tedious optimizing required	Easy to program Manually - can be very tedious optimizing not required
Flexibility	Did Ok	Failed often - least flexible by far sample size of 500 & 5 strata, failed 42% of the time!	Did Ok	Did the best overall
Explainable	Easy to explain & Explaining odd sample sizes not easy	Easy to explain & Explaining odd sample sizes not easy	Difficult to explain & Explaining odd sample sizes also not easy	Very difficult to explain No need to explain sample size calculations
Acceptance	Provides for unusual sample sizes and odd break values	Provides for unusual sample sizes and odd break values	Provides for odd sample sizes but uniform break values	Even sample sizes and uniform break values

Geometric Ratio

- *Would never use as default policy*
  - *one size fits all doesn't work*
- *If taxpayer requested, would be acceptable*
- *Best if you use at least 4-5 strata*
- *Neyman Allocation critical*
- *"Explainable" advantage really not there*

Other thoughts on methods

- *Not a lot of difference in the standard errors across these four methods*
- *Don't knock yourself out!*
- *Equal and Unequal CSRF very very close*
- *Using Equal or Unequal CSRF as default policies appears prudent*
- *By experience, odd sample sizes and break values cause questions…*

Attachment #7E

Stratification Study…

29

- Other thoughts on methods…
  - Geometric Ratio dependent on using Neyman Allocation
  - Best if you **<u>don't</u>** use Neyman Allocation with CSRF-Unequal or Proportional
    - (use same sample size for each stratum)
  - Better if you use Neyman Allocation with CSRF-Equal
    - but not critical
  - Low error rate populations do better with CSRF-Equal and CSRF-Unequal
  - Geometric Ratio does well with high error rate populations
  - < 1% almost no gain in efficiency by stratifying
    - Sometimes (but rarely) there was a decrease in efficiency

Attachment #7C

Standard Error…

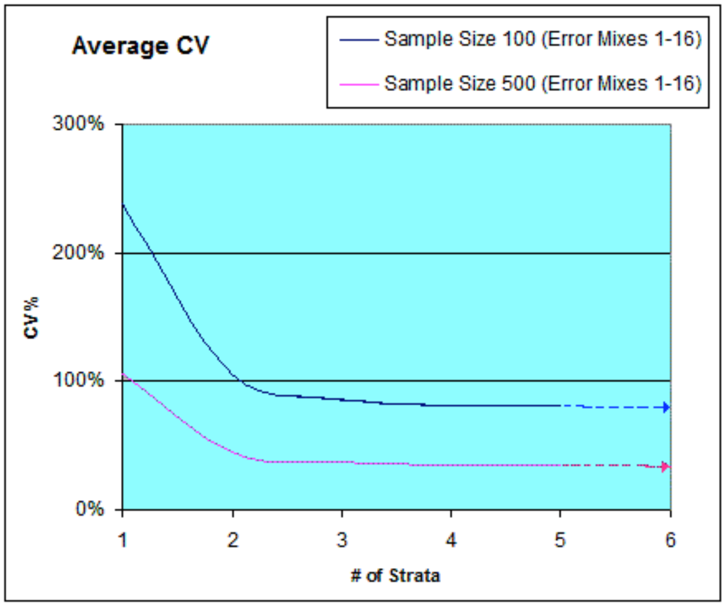
30

- If the standard error of the *invoice amounts* is (usually) interrelated with the standard error of the *error amounts*, how often will one stratification method have the lowest standard error for both?
- Depends on:
  - *the number of strata,*
  - *the combined error rate, and*
  - *the type of error*
- Overall, 40% of the time in the study
  - *If it was 25%, then it would be by chance*
  - *Geometric Ratio paper appears to have presumed 100%*
- *Lesson:*
  - *Use standard error of the invoice amounts as a guide*
  - *Don't presume it will be an exact prediction*

Attachment #7B

Efficiency Gains by Number of Strata

31



Attachment #7A