8515631

Hitzig, Neal B.

SUBSTANTIVE AUDIT TESTS OF DETAILS AND THE ESTIMATION OF RARE ERRORS BY DOUBLE SAMPLING WITH PROBABILITY PROPORTIONAL TO SIZE

City University of New York PH.D. 1985

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

SUBSTANTIVE AUDIT TESTS OF DETAILS AND THE ESTIMATION OF RARE ERRORS BY DOUBLE SAMPLING WITH PROBABILITY PROPORTIONAL TO SIZE.

by

NEAL B. HITZIG

A dissertation submitted to the Graduate Faculty in Business in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1985

This manuscript has been read and accepted for the Graduate Faculty in Business in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

2/28/85
date

Martin Benis
Chairman of Examining Committee

2/28/85
date

Sidney I. Lirtzman
Executive Officer

Douglas R. Carmichael

Martin R. Frankel

Matthew Goldstein

Steven Lilien

Maurice S. Newman
Supervisory Committee

The City University of New York

To

Herbert Arkin

Professor Emeritus

Baruch College

City University of New York

PREFACE

An auditor performs a substantive test of details of transactions or account balances in order to assess the amount of monetary error in an accounting population. The assessment is often based on an auditor's consideration of the results of a random sample of items from the accounting population under audit. Sample information is often used as the sole or predominant basis for a such an assessment.

Sampling implies sampling risk, which audit samplers need to control. To control risk, auditors need to be able to measure it. Because auditors act on sample information, they also need to know what courses of action are consistent with the risks that they are willing to run. Finally, because risk and sample size are related, auditors need to know how to determine sample sizes that are consistent with the risks they are willing to run.

This study deals with a common audit sampling problem - the detection and estimation of infrequently occurring overstatement in an accounting population - in a way that satisfies the aforementioned needs. In this study I present a methodology which enables the auditor to plan a sample and to act on the sample's results, subject to risks which he can measure and control.

The methodology is designed to satisfy two audit test objectives: to detect errors in an accounting population if the actual error in the population exceeds some specified maximum tolerable amount; and to estimate the extent of detected error with sufficient confidence and precision to provide the auditor with a reliable basis for an adjustment to the affected account. The amount of adjustment is governed by a rule, which applies whenever error is detected and the auditor's precision requirement is satisfied. The purpose of the adjustment rule is to control the risk that post-adjustment total error in the population exceeds the auditor's specified maximum tolerable amount.

In order to ensure that the auditor's precision requirement is achieved, so that the adjustment rule may be applied, the technique of estimation by double sampling is employed.

In this study, the double sampling procedure is applied to samples which are selected with probability proportional to size (pps). The pps sampling method is commonly used in auditing where the auditor is testing for the existence of errors of overstatement of an accounting population.

In practice, overstatement error totaling five to ten percent of a population's recorded value frequently is large enough to be of audit concern. Reliable interval estimates of relatively rare error are not readily obtainable by conventional statistical methods which depend on assumed asymptotic normality of the sampling distribution. Consequently, this study is also concerned with the development of an improved method of calculating confidence limits for pps samples.

The methods developed in this study are tested by computer-assisted analysis. The analytical approach employed is enumeration of the sample space in single and double sampling under a specified error condition. Forty-two error conditions are subjected to analysis under each of two levels of estimation risk (5 percent and 2 percent), varying amount of population error (in three levels, from 50 percent to 150 percent of the specified maximum tolerable error) and richness of error condition (including binomial, trinomial and quadrinomial distributions).

The results of the analysis demonstrate that the proposed methodology comprises an effective audit test procedure which requires no auditor dependence on either internal accounting control or other audit procedures.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

AUDIT SAMPLING

> The overall objective of an auditor in carrying out the attest function is to make reasonably certain that statements examined by him are materially correct. He therefore wishes to find material errors in the financial statements if they exist. To do so he employs numerous procedures, one of the most important being the examination of documentary evidence.[1]

An important problem in the audit of financial accounting data is the development of procedures which are sufficient in themselves for the detection and/or estimation of monetary error in populations of transactions or account balances. Such procedures are, by definition, capable of satisfying the third standard of field work of the American Institute of Certified Public Accountants (AICPA), which states:

> Sufficient competent evidential matter is to be obtained through inspection, observation, inquiries and confirmations to afford a reasonable basis for an opinion regarding the financial statements under examination.[2]

An audit test of details of transactions and balances is one of the principal procedures by which the evidential matter required by the third standard of field work is obtained.[3] Such a test, which is referred to as a "substantive test of details," may be applied statistically, based on the results of a random sample of the transactions or balances. The purpose of this type of test, which is known as an audit sample, is to evaluate some characteristic (usually error) of the population from which the sample was selected.[4,5]

Statistical, or random, sampling techniques have been used by accountants and auditors at least since the 1950s.[6] Formal recognition of the techniques by the accounting profession first occurred in 1962.[7]

Statement on Auditing Standards Number 39, Audit Sampling (SAS 39),[8] sets forth the accounting profession's requirements for the use of sampling techniques in the performance of audit tests of details. There are four principal requirements that are of interest for this study. They are:

1. Sample items should be selected in such a way that the sample can be expected to be representative of the population from which it was selected.[9] Although no definition of the term "representative" is given, SAS 39 does state that random samples satisfy this requirement.

2. Errors that are identified in the sample should be projected to the population from which the sample was selected.[10] That is, a point estimate of error should be calculated.

3. The projected error should be compared with the maximum tolerable error and consideration should be given to sampling risk. Sampling risk is the risk that "the auditor's conclusions may be different from the conclusions he would reach if the test were applied in the same way to all the items in the account balance or class of transactions."[11]

4. When planning a substantive test, the auditor should consider how much monetary error in the population may exist without causing the financial statements to be materially misstated (maximum tolerable error).[12]

Thus, an audit sampling procedure is viewed in the authoritative literature as, first and foremost a decision-making device. Its primary function is to enable the auditor to decide whether or not the amount of error in a population exceeds the maximum amount that the auditor considers to be tolerable in the circumstances. Another objective of an audit sample is to provide a basis for action by the auditor. If the results of a sample lead to a decision that error may exceed the maximum tolerable amount, then further action is appropriate. Further action may involve additional work or adjustment of the book value of the account under examination.[13,14]

Substantive test of details of transactions or balances are tests of financial statement assertions. Assertions are management's representations as to characteristics of financial statement components. The categories into which assertions are classified include existence or occurrence, rights and obligations, valuation or allocation, presentation and disclosure, and completeness.[15] Of these assertions, only the completeness assertion (all transaction and accounts are recorded that should be recorded) is not readily amenable to testing by examination of the details. The reason is that such tests generally involve selection of the details from the records. If the records are not complete, then there will be transactions or accounts that will not be available for testing. The point of this observation is that no single test procedure can be expected to satisfy all audit test objectives. The test procedure that is the subject of this study is designed to test specific aspects of the existence, rights and obligations, and valuation or allocation assertions, examples of which are given in Table 1.1. Each of the error conditions described in Table 1.1 would, if present, result in overstatement of the population under examination. That is, error amounts are strictly positive. Moreover, each error definition describes a condition in which the magnitude of the overstatement cannot, by definition, exceed the associated recorded value of the account or transaction in which the error occurs.

These assertions and error conditions are common to audit of financial statements, and are the ones to which the sampling procedure of this study is directed.

TABLE 1.1

FINANCIAL STATEMENT ASSERTIONS

Assertion	Error Condition
Existence	• Transaction did not occur during audit period. • Transaction is not valid or is otherwise improper. • Transaction occurred, but should have been subsequently reversed (sales returns not credited).
Rights and obligations	• Entity does not have ownership rights or obligations with respect to items comprising an account balance (consignment inventories improperly included in inventory total).
Valuation or allocation	• Transaction or account not realizable (receivables not collectible; inventories not at lower of cost or market). • Transactions improperly included in the account under examination (capitalized items that should be charged to income).

CHAPTER 1 FOOTNOTES

1. Elliot, R.K., and Rogers, J., "Relating Statistical Sampling to Audit Objectives," The Journal of Accountancy, July 1972, 46-55.

2. AICPA, Codification Auditing Standards of Statements on Numbers 1 to 47, New York: AICPA, 1984, AU 150.02.

3. Ibid., AU 320.74.

4. Ibid., AU 320.79.

5. Ibid., AU 350.01,.03,.05.

6. The following are noteworthy examples of early writings on statistical sampling in auditing: Neter, J., and L.L. Vance, Statistical Sampling for Auditors and Accountants. New York: John Wiley & Sons, 1956; Trueblood, R.M., and R.M. Cyert, Sampling Techniques in Accounting. New Jersey: Prentice-Hall, 1957; Vance, L.L. Scientific Method for Auditing. Berkeley: University of California Press, 1950.

7. A brief history is given in Anderson, R., and Teitelbaum, A.D., "Dollar Unit Sampling," CA Magazine (formerly Canadian Chartered Accountant), April 1973, 30-38.

8. SAS 39 is included in the AICPA Codification, op. cit., as AU 350.

9. Ibid., AU 350.24.

10. Ibid., AU 350.26.

11. Ibid., AU 350.10 and .26.

12. Ibid., AU 350.18.

13. Anderson, R.J., The External Audit, Toronto: Pitman, 1977, V.1, 367-369.

14. Elliot and Rogers, op. cit., 53-54.

15. AICPA, op. cit., AU 326.

CHAPTER II

THE AUDIT TEST FRAMEWORK

> A test is nothing but a rule by which we sometimes reject the hypothesis tested and sometimes accept it..., according to whether or not the observations available possess some properties specified by the rule.[1]

Decision rules

Under the approach considered in this study for conducting a substantive audit test of details of transactions or balances, the auditor relies entirely on the information contained in a random sample to make a decision regarding the extent of error in a population. This implies that a procedure should possess two properties. First, it should be capable of detecting error if the error in the population exceeds some auditor-specified maximum tolerable amount. Second, given that the procedure detects error, it should provide sufficient information for the auditor to choose an appropriate course of action. Courses of action might include 1) accepting the population as-is, despite the existence of error, 2) adjusting the population's recorded amount (book value), or 3) gathering more information.

The substantive test of details has been expressed in the form of a classical parametric test of hypothesis by Elliott and Rogers[2] and Roberts.[3,4] The test may be formulated on either a positive or a negative basis, according to terminology employed by Roberts. On a positive basis the null and alternate hypothesis as to the population error are stated as follows:

<u>Null hypothesis</u> - The amount of error is zero.

<u>Alternate hypothesis</u> - The amount of error is not zero.

On a negative basis the hypotheses are stated as follows:

<u>Null hypothesis</u> - The error exceeds the maximum tolerable amount.

<u>Alternate hypothesis</u> - The amount of error is tolerable.

If, after sampling, the the null hypothesis is not rejected, then no further audit work is necessary. If, on the other hand, the auditor rejects the null hypothesis, then the process becomes one of estimating the amount of error in the population. In this case two possible actions would be considered by the auditor. The auditor could determine the amount of an adjusting journal entry that would be appropriate in the circumstances; if necessary selecting additional sample items in order to obtain a more precise estimate of the error.

Statistical (that is, random) audit samples typically involve the auditor's consideration of the confidence interval.[5] Consequently, there is risk that the auditor's conclusion is incorrect because it is based on the results of a sample, rather than a census.

In audit sampling, sampling risk is defined in the following terms:

> The risk of incorrect rejection (α), that is, the risk that the sample erroneously supports the conclusion that error exceeds the maximum tolerable amount.
> The risk of incorrect acceptance (β), that is, the risk that the sample erroneously supports the conclusion that the amount of error is tolerable.[5]

These risks are related to the auditor's specified maximum tolerable error M and to the standard error of the estimate $s_{\hat{E}}$ via the expression,

$$M = s_{\hat{E}}(z_{\alpha/2}+z_{\beta}), \qquad 2.1$$

where $z_{\alpha/2}$ and z_{β} are the standard normal deviates for specified one-sided risks $\alpha/2$ and β, respectively. The risk of incorrect rejection is also the risk that actual population error will not be within the confidence interval of the estimated error $\hat{E}$.

Under the positive decision approach, the following decision rules would apply.

Sample Result	Decision
(LEL.LE.0).AND.(UEL.GE.0)	The population is not misstated
(LEL.GT.0).OR.(UEL.LT.0)	The population is misstated.

The foregoing sample results are given as FORTRAN logical expressions.[7]

Under this approach, there is no explicit reference in the decision rule to the maximum tolerable error M. Instead, M is implicitly considered in the specification of the risk of incorrect acceptance β. That is, the procedure is designed to give a $(1-\beta)$% probability of a decision that misstatement exists, if the actual error exceeds M. This approach is particularly useful when errors of understatement or overstatement are possible (as, for example, in the case of an inventory price test).

Under the negative decision approach, the following decision rules would apply.

Sample Result	Decision
(LEL.GE.-M).AND.(UEL.LE.M)	Misstatement is no greater than M.
(LEL.LT.-M).OR.(UEL.GT.M)	The magnitude of misstatement may exceed M.

An important special case of the negative decision approach involves such tests of the existence, rights and obligations, and valuation assertions for which errors can only be overstatements, as indicated in Table 1.1. In this case the decision rules are as follows.

Sample Result	Decision
UEL.LE.M	Overstatement is no greater than M.
UEL.GT.M	Overstatement may exceed M.

Neither approach necessarily leads to a categorical determination that error exceeds the maximum tolerable amount, a consequence of which is the use of the word "may" in the negative approach's decision rule.[8] For the auditor to make such a determination the following result would have to occur.

Sample Result	Decision
(UEL.LT.-M).OR.(LEL.GT.M)	Magnitude of error is greater than M.

Unless the preceding result is also obtained when error is detected, some ambiguity arises. The decision process is incomplete, because no consideration is given to the consequences of a decision that misstatement exists (positive approach), or that misstatement may exceed the maximum tolerable amount (negative approach).

Detection/Estimation Procedure

The detection/estimation framework and the associated sampling procedures that are developed in this study for overstatement error conditions eliminate ambiguity by eliminating the need for the auditor to formalize his test in the aforementioned manner and by placing emphasis on obtaining sufficiently precise, reliable internal estimates of error. When testing for overstatement error conditions, as defined in Chapter I, the detection of even one error in a sample, however small its magnitude, means that the population is overstated at least to the extent of the detected error. Thus, the detection of at least one error will always lead to the rejection of any hypothesis that the amount of error is zero. On the other hand, the detection of errors in a sample will often result in a confidence interval which also includes the maximum tolerable error within its range. In this case the auditor will not have sufficient information for a categorical statement that misstatement of the population is or is not tolerable. This indefiniteness can be eliminated by restating the audit sampling process as a detection/estimation procedure.

Since the detection of one error is conclusive as to the existence of overstatement in the population under examination, then such an outcome would impose upon the auditor the need to estimate the extent of error. The confidence interval obtained therefrom would then provide the basis for an adjusting journal entry to the account in question. The objective of such an adjustment would be to reduce the auditor's ignorance as to the extent of error remaining in the population to an amount not exceeding the specified maximum tolerable error, subject to the specified estimation risk (that is, the complement of the confidence level that was used to obtain the interval estimate). In this way the auditor would be able to satisfy the third standard of field work with respect to the population under examination based entirely on the information obtained from the sample.

As an alternative to the risk definitions suggested by SAS 39, the following would be employed:

Detection risk (β) - This is the risk that the audit sample would fail to detect any error, given that the population is overstated by more than the maximum tolerable amount.

Estimation risk (α) - This is the risk that the confidence interval will not contain the actual amount of population overstatement within its limits.

A third risk, adjustment risk, is also implied in any estimation procedure which gives rise to an adjustment. However, as will be seen by the application guidelines set forth below, adjustment risk, which is the risk that the error remaining after adjustment exceeds the maximum tolerable error, will not exceed estimation risk.

The detection/estimation procedure for tests of overstatement is an integrated double sampling procedure having one objective - to yield a population whose book value is in error by no more than the maximum tolerable amount. In order for the procedure to function in accordance with the auditor's specified risk levels the following conditions must be satisfied.

1. An initial sample is selected which is large enough so that at least one error will be detected, subject to a risk of failure not

exceeding the specified detection risk, if the amount of population overstatement exceeds the specified maximum tolerable error.

If no error is detected in the initial sample, the auditor will be able to conclude, subject to risk that is no greater than the specified detection risk, that population overstatement does not exceed the maximum tolerable amount.

2. If at least one error is detected in the initial sample, an interval estimate is obtained at $(1-\alpha)\%$ confidence, where $\alpha=2\beta$.

3a. If the range of the confidence interval is not greater than twice the maximum tolerable error, then an adjustment is proposed equal to any amount that is within the specified maximum tolerable amount of <u>both</u> confidence limits. The auditor could then be $(1-\alpha)\%$ confident that the remaining error in the population would not exceed the maximum tolerable amount.

3b. If the range of the confidence interval is greater than twice the maximum tolerable error, then additional sampling is performed to reduce the length of the interval so as to permit application of step 3a.

The Adjustment Rule

Under the detection/estimation procedure, the confidence interval provides the basis for an adjusting journal entry if the sample discloses error. Furthermore, the auditor needs to be able to conclude that the adjusted book value is not misstated by more than the maximum tolerable amount. Application of the following rule reduces the amount of post-adjustment error to the maximum tolerable amount, subject to the estimation risk.

Required Condition:	(UEL-LEL).LE.2M
Smallest Permitted Adjustment:	max (LEL,UEL-M)
Largest Permitted Adjustment:	min (UEL,LEL+M)

Any value within the range from the smallest to the largest permitted adjustment may be considered to be sufficiently correct, at the specified confidence level. An adjustment is not proposed unless the required condition is satisfied. If the condition is not satisfied in the initial sample, the sample is extended and additional items are examined.

CHAPTER II FOOTNOTES

1. Neyman, J., Lectures and Conferences on Mathematical Statistics and Probability. Washington: United States Department of Agriculture, 1952, p.55.

2. Elliott and Rogers, op. cit., 46-53.

3. Roberts, D.M., "A Statistical interpretation of SAP 54," Journal of Accountancy, March 1974, 47-53.

4. ----------, Statistical Auditing, New York: AICPA, 1977, 39-48.

5. Bailey, A., Statistical Auditing, New York: Harcourt Brace Jovanovich, 1981, 57-78.

6. AICPA, op.cit., AU 350.

7. The following notation will be employed in this and in subsequent sections:

   UEL = upper confidence limit of error (upper error limit);

   LEL = lower confidence limit of error (lower eror limit); and

   M = maximum tolerable error.

   AND = conjunction

   OR = disjunction

   LE = "less than or equal to"

   LT = "less than"

   GE = "greater than or equal to"

   GT = "greater than"

   FORTRAN mnemonics and syntax will be used for logical expressions.

8. Anderson, R., op. cit., p.367.

CHAPTER III

INTERVAL ESTIMATES

Auditors obtain an interval estimate of the amount of error in a population by projecting the errors detected in a sample to the population from which the sample was selected. The estimators that are ordinarily used are the difference, ratio, and pps-ratio estimators.[1] It has been observed that errors in accounting populations typically are relatively rare and that accounting populations are highly skewed. Confidence intervals that are based on assumptions of asymptotic normality in the sampling distribution tend to result in understatement of sampling risk.[2,3]

The errors in an accounting population typically are relatively rare. The behavior of the sampling distribution of estimates of relatively rare monetary error resembles a binomial or Poisson sampling distribution of estimates of attribute data in two key respects: the population is highly skewed, and the standard deviation that is estimated from a sample is highly correlated to the number of errors found in the sample.[4] Consequently, large samples need to be selected in order to calculate confidence intervals that are based on the assumption of asymptotic normality of the sampling distribution. For example, if a 5% error rate were to occur in a sample whose results were to be expressed with

95% confidence, assumed normality of the sampling distribution would be appropriate if the sample size were at least 1,400 items.[5]

Instead of selecting large samples in order to induce normality, auditors and statisticians have sought alternative ways to obtain interval estimates that are less sensitive to the non-normality of the sampling distribution. One suggested alternative is to employ confidence coefficients using Student's t, using the number of errors detected in the sample, instead of the number of degrees of freedom, as the basis for determining t.[6] Another approach is to assume that the sampling distribution may be approximated by a gamma distribution, whose standardized deviate is calculated from sample estimates of skewness and kurtosis.[7] In a third method the standard error is calculated for an estimator that is viewed as the product of two random variables: the number of errors found, and the average amount of error in the erroneous items. The confidence coefficient t is based on Student's t with degrees of freedom equal to the number of detected errors minus one.[8] All three approaches have one major shortcoming: all give trivial or near trivial results when the sample discloses no error or when errors are very small amounts.

Dollar Unit Sampling

This study presents a method for calculating confidence limits based on a sampling procedure that is designed to detect and estimate overstatement in an accounting population. The following conditions apply:

(1) The book value b_i of every item i in the population is positive. This is, $b_i > 0$, for all i.

(2) No understatements exist in the population, nor can the amount of overstatement e_i in an item be greater than the book value of the item. That is, $0 \leq e_i \leq b_i$, for all i.

(3) The population error rate and the aggregate error amount E are small relative to the number of items in the population and the book value B of the population. Error rates in the five to ten percent range would be considered typical of the rates of error that could result in material misstatement of the accounting population under examination.[9]

The sampling method, called dollar unit sampling (dus) by Anderson and Teitlebaum,[10] is derived from survey sampling methodologies in which items are selected with probability proportional to size (pps).[11,12] In an auditing application an item's book value is the measure of its size. Dollar unit sampling was developed for audit use by Stringer.[13] Many of the labeling conventions employed herein were established by Anderson and Teitlebaum. These conventions are based on Deming's conception of an elementary population unit as a "dollar of investment."[14]

The principal distinguishing feature of the dus viewpoint lies in the redefinition of the population from its physical sampling unit basis (such as a sales invoice) to a dollar unit basis (such as Deming's "dollar of investment," as recorded in a sales invoice). Under the dus convention, we substitute for each sampling unit a number of individual dollar units whose total number is equal to the book value of the related sampling unit. For example, a $1,000 invoice is viewed as 1,000 identical $1 invoices under the dollar unit convention. Each dollar unit is a miniature replica of its associated sampling unit.

Under the dollar unit viewpoint any overstatement of the recorded amount of the sampling unit is allocated pro rata to each of its dollar units. For example, if a $1,000 invoice is found to be overstated by $400, then each of the 1,000 dollar units is said to be overstated, or tainted, by 40 cents.

By adopting the dollar unit convention, a conceptual simplification is thereby introduced. Under dus, the sampling process can be viewed as a simple random sample of dollar units. When each dollar unit and each combination of dollar units has an equal chance of selection, the sample can be viewed as a simple random sample of dollar units. A dollar unit sample should then be amenable to conventional variables estimation procedures.

The point estimate $\hat{E}$ of total error is the pps-ratio estimate,

$$\hat{E} = (B/n)r \tag{3.1}$$

where B is the total book value of the population,

n is the sample size.

Reordering the sample so as to sum over the m non-zero tainting ratos,

$$r = \sum_{i=1}^{m} r_i, \tag{3.2}$$

where

r_i is the ratio of the overstatement error to the book value of the ith overstated item,

$0 < r_i \leq 1$,

and

m is the number of errors in the sample, $m \leq n$.

If tainting rates are restricted to 0 or 1 (i.e., the dollar units are either correct or fully overstated), then dus evaluation can be performed using conventional attributes estimation methods. Under the Poisson approximation to the binomial this means solving for the largest upper limit factor $r_u(m)$, such that

$$\sum_{k=0}^{m} r_u(m)^k \exp(-r_u(m))/k! = \alpha/2, \qquad 3.3$$

and the smallest lower limit factor $r_l(m)$ such that

$$\sum_{k=m}^{n} r_l(m)^k \exp(-r_l(m))/k! = \alpha/2, \qquad 3.4$$

where n is the sample size, and m is the number of erroneous dollar units in the sample.

Frequently, tainting rates may take any value on the interval from 0 to 1; as, for example, when a $1,000 account balance is overstated by $50 because of a failure to give credit for returned merchandise. If a dollar unit is selected from this account, its tainting rate would be 0.05. The basic procedure for evaluating dollar unit samples, if fractional taintings are detected (that is, $0<r_i<1$) and ranked, was developed (the Stringer bound) from an extended consideration of the Poisson distribution.[15] Taintings, ranked in descending order, are applied to increments of the Poisson confidence limits, which are then summed. The $(1-\alpha/2)\%$ upper error limit in dollars is

$$UEL = (B/n)(r_u(0)+\sum_{k=1}^{m} r_k(r_u(k)-r_u(k-1))), \qquad 3.5$$

where B = the total book value of the sampled population,

n = sample size,

$r_u(k)$ = Poisson upper confidence limit factor for k sample errors,

$r_u(0) = -\ln(\beta)$,

r_k = kth largest tainting rate in sample, and

m = the number of errors in the sample

The $(1-\alpha/2)\%$ lower error limit in dollars is

$$LEL = (B/n)\sum_{k=1}^{m} r_k(r_l(k)-r_l(k-1)) \qquad 3.6$$

where $r_l(k)$ = Poisson lower confidence limit factor for k sample errors. Note that $r_l(0) = 0$.

Although reliable, the Stringer bound is not efficient. Recent work has concentrated on obtaining better precision than the Stringer bound offers. Fienberg, Neter, and Leitch,[16] and Teitlebaum, McCray and Leslie[17] have developed procedures based on the multinomial and multiple Poisson distributions, and have demonstrated that improvements to the Stringer bound exist. The multinomial (and multiple Poisson) approach is formulated as an optimization problem whereby the objective is to solve for the population proportions $\{p_i, i=1,2,\ldots,100\}$ of tainting rates $r_i=.01, r_2=.02,\ldots,r_{100}=1.00$ which maximize

$$\sum_{i=1}^{100} r_i p_i,$$

subject to 3.7

$$\sum_{S} \frac{n!}{m_0! m_1! \ldots m_{100}!} \prod_{i=0}^{100} (p_i)^{m_i} = \alpha/2,$$

$$p_i \geq 0, \quad i=0,1,\ldots,100,$$

$$\sum_{i=0}^{100} p_i = 1,$$

where m_i is the number of occurrences of each detected tainting rate (for convenience, tainting rates are classified into 100 categories at the values of r_i defined above) and S is a suitably chosen set of possible sample outcomes for which the total overstatement error does not exceed the overstatement observed in the sample. The method of Lagrange multipliers is used to solve for the proportions $\{p_i\}$. This method, while yielding the most precise limits to date, is computationally unwieldy. Thus far, its principal practical value is to serve as a benchmark for comparison with other procedures. Although the multinomial method has mathematical support for only a limited set of conditions, the results suggest that a general proof of their reliability by mathematical induction, may exist. Others, including Cox and Snell[18] and McCray[19] employed a Bayesian approach to interval estimation. All of these aforementioned approaches are computationally complex and are ill-suited to practical application.

Of all the improvements to the Stringer method, the most practical is by Leslie, Teitlebaum and Anderson,[20] who developed a procedure based on the random selection of dollar units within cells of fixed length, which computes bounds based on certain assumption as to the distribution of errors within the cells. Another approach to reducing the length of the estimation interval was developed by Gartska and Ohlson,[21] who calculated confidence intervals by substituting the standardized statistics for the binomial for Student's t statistic. However, the method was found by their developers to overstate the confidence level in populations where the distributions of tainting rates are reverse J-shape, containing both large and small tainting rates.

A Parametric Approach to Interval Estimation

As previously stated, the tainting rates assume only integer values (0 or 1) if sample items are testing assertions that result in the items' audited values being either correct or totally incorrect. Then $r = \sum r_i = m$. That is, the sum of the tainting rates is the number of errors. The sampling distribution (employed in this study) is the Poisson approximation to the binominal. The upper and lower $(1-\alpha/2)\%$ confidence limits are

$$UEL = (B/n)r_u(m)$$

and $$\tag{3.8}$$

$$LEL = (B/n)r_l(m),$$

respectively. The upper and lower confidence limit factors, $r_u(m)$ and $r_l(m)$, are chosen so as to satisfy equations 3.3 and 3.4, respectively.

Molina showed that the Poisson distribution function could be evaluated in terms of the incomplete gamma function and vice versa.[22] Thus,

$$\sum_{k=0}^{m} r_u(m)^k \exp(-r_u(m))/k! = \frac{1}{\Gamma(r+1)} \int_{r_u(m)}^{\infty} x^r \exp(-x)dx = \alpha/2,$$

and $$\tag{3.9}$$

$$\sum_{k=m}^{n} r_l(m)^k \exp(-r_l(m))/k! = \frac{1}{\Gamma(r+1)} \int_{0}^{r_l(m)} x^r \exp(-x)dx = \alpha/2.$$

This isomorphic relationship is important, for it also may be applied when the tainting rates r_i are not restricted to the integers 0 and 1. That is if $0 \leq r_i \leq 1$, then it is conjectured that there exists a continuous probability distribution, which is Poisson-like, whose cumulative distribution function is given by the incomplete gamma function. The plausibility of this conjecture rests on the following observations. First, the incomplete gamma function is tractable for real values of r, $0 \leq r \leq n$. Second, it has been shown that any situation involving fractional r_i may also be transformed into a situation involving only integer r_i.[23] The hypothesized distribution appears to be a continuous distribution for all values of its random variable except 0, at which point it has a probability mass, which is $\exp(-r_u(r))$ for the upper limit and $\exp(-r_l(r))$ for the lower limit.

The sum of the taintings may now be given a new interpretation, that of equivalent error. For example, 20 tainting rates of .1 each amount to 2 equivalent errors. The equivalent error in a sample is not restricted to integer values. For any equivalent error value r, the upper and lower $(1-\alpha/2)\%$ confidence limit factors, $r_u(r)$ and $r_l(r)$ lie on continuous curves which are described by the incomplete gamma function. Values are given in Tables 3.1 and 3.2 for $\alpha/2$ = 5%, 2.5% and 1%. Good approximations may be obtained by linear interpolation.

Useful as the extension of Molina's result is, it is nevertheless conservative, since the confidence limits reflect only the equivalent error r in the sample. They do not reflect the variability of the individual r_i's. For example, 10 taintings at 0.1 each would result in the same confidence limits as would 1 tainting at 1.0, hence the same precision, unless relative variability were taken into account.

The gamma-based method can be enriched to give effect to sample variability by recognizing that the incomplete gamma function ignores the scaling factor that is associated with the two-parameter gamma distribution. Thus, an evaluation that is based on the incomplete gamma implicitly assumes that $\sum r_i^2 = \sum r_i$, which is conservative to the extent that $0 < r_i < 1$. The scaling factor q can be estimated from the sample.[24] Thus,

$$q = \sqrt{\sum r_i^2 / \sum r_i}. \qquad 3.10$$

With the scaling factor thus estimated, the upper and lower $(1-\alpha/2)\%$ precision factors, $r_u(r)-r$ and $r-r_l(r)$ can be adjusted for tainting variability. Thus,

$$pf_u = (r_u(r)-r)\sqrt{\sum r_i^2 / \sum r_i}$$

and 3.11

$$pf_l = (r-r_l(r))\sqrt{\sum r_i^2 / \sum r_i}.$$

This result leads to the following observation. Noting that $r_i = r$, then the expressions for the upper and lower precision factors may be rewritten as

$$pf_u = t_u(r)s_r,$$

and 3.12

$$pf_l = t_l(r)s_r,$$

where $t_u(r)$ and t_l (r) are standardized deviates of the conjectured distribution with q=1 whose standard deviation is estimated from a sample. Tables 3.3 and 3.4 give $t_u(r)$ and $t_l(r)$ for calculating the precision factors.

Thus, we have (before projecting by B/n) an estimator r which has an estimated variance

$$s_r^2 = \sum_{i=1}^{m} r_i^2. \quad 3.13$$

Noting that the number of errors m is a random variable and that

$$s_r^2 = m \sum_{i=1}^{m} r_i^2/m, \quad 3.14$$

we have obtained estimators of the parameters of a compound Poisson distribution. This makes sense, since it is reasonable to assume that the number of errors m that occur in a sample of n items is a Poisson process, and that the average value of the m errors is independent of the number of errors. Thus, it appears that the conjectured sampling distribution may be a limiting case of the compound Poisson. This finding gives a straightforward method for obtaining confidence limits.

The upper and lower confidence limits, UEL and LEL, on the estimated error are

$$UEL = (B/n)(r+pf_u),$$

and 3.15

$$LEL = (B/n)(r-pf_l).$$

In practice, a lower bound will be set for the upper precision factor. This lower bound is $r_u(0) = -\ln(\beta)$, which is the implied precision factor when no error is disclosed. Thus, for purposes of computing the upper limit, it is assumed that the precision factor obtained when errors are detected can never be less than $r_u(0)$. This is discussed further in Chapter IV.

In Table 3.5 a comparison is given between the compound Poisson limits and those given by the multinomial and Stringer methods. Note that the compound Poisson limits are more precise than the Stringer limits and less precise for upper limits than the multinomial limits (lower limits for this method have not been published).

TABLE 3.1
UPPER CONFIDENCE LIMIT FACTORS, $r_u(r)$,[26]
COMPOUND POISSON DISTRIBUTION
(Scale Factor = 1)

Equivalent Errors r	$\alpha/2=5\%$	$\alpha/2=2.5\%$	$\alpha/2=1\%$
0.0	3.00	3.69	4.61
0.5	3.91	4.68	5.68
1.0	4.75	5.58	6.64
1.5	5.54	6.42	7.55
2.0	6.30	7.23	8.41
2.5	7.04	8.01	9.24
3.0	7.76	8.77	10.05
3.5	8.46	9.51	10.84
4.0	9.16	10.25	11.61
4.5	9.84	10.96	12.37
5.0	10.52	11.67	13.11
6.0	11.85	13.06	14.58
7.0	13.15	14.43	16.00
8.0	14.44	15.77	17.41
9.0	15.71	17.09	18.79
10.0	16.97	18.40	20.15
15.0	23.10	24.75	25.45
20.0	29.07	30.89	33.11
25.0	34.92	36.91	39.31
30.0	40.70	42.83	45.41
35.0	46.41	48.68	51.41
40.0	52.07	54.47	57.35
45.0	57.70	60.21	63.24
50.0	63.29	65.92	69.07
75.0	90.89	94.02	97.74
100.0	118.07	121.61	125.84

TABLE 3.2
LOWER CONFIDENCE LIMIT FACTORS, $r_1(r)$,[27]
COMPOUND POISSON DISTRIBUTION
(Scale Factor = 1)

Equivalent Errors r	$\alpha/2=5\%$	$\alpha/2=2.5\%$	$\alpha/2=1.0\%$
0.0	0.0	0.0	0.0
0.5	0.00+	0.00+	0.00+
1.0	0.05	0.02	0.01
1.5	0.17	0.11	0.05
2.0	0.35	0.24	0.14
2.5	0.57	0.41	0.27
3.0	0.81	0.61	0.43
3.5	1.08	0.84	0.62
4.0	1.36	1.08	0.82
4.5	1.66	1.35	1.04
5.0	1.97	1.61	1.27
6.0	2.61	2.20	1.78
7.0	3.28	2.81	2.32
8.0	3.98	3.45	2.89
9.0	4.69	4.11	3.49
10.0	5.42	4.79	4.11
15.0	9.24	8.39	7.45
20.0	13.25	12.21	11.06
25.0	17.38	16.17	14.83
30.0	21.59	20.24	18.72
35.0	25.86	24.37	22.70
40.0	30.19	28.57	26.75
45.0	34.56	32.82	30.86
50.0	38.94	37.11	35.01
75.0	61.33	58.99	56.32
100.0	84.13	81.36	78.20

TABLE 3.3
STANDARDIZED DEVIATES, $t_u(r)$,[28]
COMPOUND POISSON DISTRIBUTION
UPPER $\alpha/2$% POINTS

Equivalent Errors r	$\alpha/2$=5%	$\alpha/2$=2.5%	$\alpha/2$=1.0%
0.0	-	-	-
0.5	4.82	5.91	7.33
1.0	3.75	4.68	5.64
1.5	3.30	4.02	4.94
2.0	3.04	3.70	4.53
2.5	2.87	3.48	4.26
3.0	2.75	3.33	4.07
3.5	2.65	3.21	3.92
4.0	2.58	3.13	3.81
4.5	2.52	3.05	3.71
5.0	2.47	2.98	3.63
6.0	2.39	2.88	3.50
7.0	2.32	2.81	3.40
8.0	2.28	2.75	3.33
9.0	2.24	2.70	3.26
10.0	2.20	2.66	3.21
15.0	2.09	2.52	2.70
20.0	2.03	2.44	2.93
25.0	1.98	2.38	2.86
30.0	1.95	2.34	2.81
35.0	1.93	2.31	2.77
40.0	1.91	2.29	2.74
45.0	1.89	2.27	2.72
50.0	1.88	2.25	2.70
75.0	1.83	2.20	2.63
100.0	1.81	2.16	2.58
oo	1.64	1.96	2.32

TABLE 3.4
STANDARDIZED DEVIATES, $t_1(r)$,[29]
COMPOUND POISSON DISTRIBUTION
LOWER $\alpha/2$% POINTS

Equivalent Errors r	$\alpha/2=5\%$	$\alpha/2=2.5\%$	$\alpha/2=1.0\%$
0.0	0.00	0.00	0.00
0.5	0.70	0.71	0.71
1.0	1.05	1.01	1.01
1.5	1.09	1.13	1.18
2.0	1.17	1.24	1.32
2.5	1.22	1.32	1.41
3.0	1.26	1.38	1.48
3.5	1.29	1.42	1.54
4.0	1.32	1.46	1.59
4.5	1.34	1.48	1.63
5.0	1.36	1.52	1.67
6.0	1.38	1.55	1.72
7.0	1.41	1.58	1.77
8.0	1.42	1.61	1.81
9.0	1.44	1.63	1.84
10.0	1.45	1.65	1.86
15.0	1.49	1.71	1.95
20.0	1.51	1.74	2.00
25.0	1.52	1.77	2.03
30.0	1.54	1.78	2.06
35.0	1.54	1.80	2.08
40.0	1.55	1.81	2.10
45.0	1.56	1.82	2.11
50.0	1.56	1.82	2.12
75.0	1.58	1.85	2.16
100.0	1.59	1.86	2.18
∞	1.64	1.96	2.32

TABLE 3.5
MULTINOMIAL, STRINGER, AND COMPOUND POISSON
CONFIDENCE LIMITS
n = 101; $\alpha/2$ = 5%
B = $1,000

Tainting Rates in Sample	Upper Limits			Lower Limits**	
	Multi-nomial	Stringer	Compound Poisson*	Stringer	Compound Poisson
m=1					
.01	29.2	29.9	29.8	0.0	0.0+
.05	29.2	30.6	30.2	0.0	0.0+
.25	29.6	34.0	32.2	0.1	1.3
.50	31.1	38.4	34.7	0.2	1.6
.75	35.4	42.7	38.0	0.4	1.3
.95	43.8	46.2	45.2	0.5	0.7
.99	45.6	46.9	46.7	0.5	0.5
m=2					
.01,.95	43.8	46.3	45.2	0.0+	0.7
.25,.95	35.8	46.5	39.6	0.1	2.5
.30,.40	31.7	41.2	36.6	0.0+	3.0
.50,.50	35.4	46.0	39.6	0.1	3.2
.75,.90	48.2	56.8	53.4	0.1	3.6

*Adjusted for basic precision. See Chapter IV.

**Lower limits for the multinomial method have not been published

CHAPTER III FOOTNOTES

1. Roberts, D.M. Statistical Auditing, New York: AICPA, 1978, 74-76, 84-86, 116-125.

2. Neter, J and J.K. Loebbecke, Behavior of Major Statistical Estimators in Sampling Accounting Populations: An Empirical Study, Research Monograph Number 2, New York: AICPA, 1975.

3. Teitlebaum, A.D. and Robinson, C.F., "The Real Risks in Audit Sampling," Journal of Accounting Research, Vol. 13, Supplement 1975.

4. Roberts, D.M., op.cit., p.74.

5. Cochran, W.G., Sampling Techniques, 3rd Edition New York: John Wiley & Sons, 1977, p.58.

6. Kaplan, R.S., "Statistical Sampling in Auditing with Auxiliary Information Estimators," Journal of Accounting Research, Autumn 1973, 238-58.

7. Arkin, H., Sampling Methods for the Auditor An Advanced Treatment New York: McGraw-Hill, 1981, 62-74.

8. Roberts, D.M., op. cit., p.77.

9. Stephan, F.F., "Some Statistical Problems Involved in Auditing and Inspection, Proceedings of the American Statistical Association, 1963, p.404.

10. Anderson, R and A.D. Teitlebaum, "Dollar Unit Sampling," CA Magazine (formerly Canadian Chartered Accountant, 102, No. 4, 1973, 30-39.

11. Cochran, W.G., op. cit., 250-254.

12. Hansen, M.H., W.N. Hurwitz, and W.G. Madow, Sample Survey Theory and Methods, Volume 1, New York: John Wiley & Sons, 1953, 341-358.

13. Stringer, K., "Practical aspects of statistical sampling in auditing," Proceedings of the American Statistical Association, 1963, 405-411.

14. Deming, W.E., Sample Design in Business Research, New York, John Wiley & Sons, Inc., 1960, p. 89.

15. Meikle, G.R., Statistical Sampling in an Audit Context, Toronto: Canadian Institute of Chartered Accountants, 1972.

16. Fienberg, S.E., J. Neter, and R.A. Leitch, "Estimating the Total Overstatement Error in Accounting Populations," *Journal of the American Statistical Association*, June 1977, 295-302.

17. Teitlebaum, A.D., McCray, J.H., and Leslie, D.A., "Approaches to evaluating dollar-unit samples," paper presented at the 1978 Annual Meeting of the American Accounting Association.

18. Cox, D.R., and E.J. Snell, "On Sampling and the Estimation of Rare Errors." *Biometrika*, 1979, 66, 1.

19. McCray, J.H., "A Quasi-Baysian Audit Risk Model for Dollar Unit Sampling," *Accounting Review*, January 1984, 35-51.

20. Leslie, D.A., A.D. Teitlebaum, and R.J. Anderson, *Dollar-unit Sampling: A Practical Guide for Auditors*, Toronto: Copp, Clark and Pittman, 1979, 278-291.

21. Garstka, S.J., and Ohlson, P.A., "Ratio Estimation in Accounting Populations with Probabilities of Sample Selection Proportional to Size of Book Values," *Journal of Accounting Research* 17, Spring, 1979, 23-59.

22. Molina, E.C., "An Interpolation Formula for Poisson's Exponential Binomial Limit," *American Mathematical Monthly*, 22, 1915, 223-224.

23. Leslie, D.A., et al. op. cit., 120-127.

24. Holt, C.C, F. Modigliani, J.F. Muth and H.A. Simon, *Planning Production Inventories and Work Force*, Englewood Cliffs: Prentice-Hall, 1960, 251-252.

25. Ross, S.M., *Applied Probability Models with Optimization Applications*, San Francisco: Holden-Day, 1970, 23.

26. The upper limit factors for non-integer values of r are adopted from Thompson, C.M., "Tables of Percentage Points of the Chi-square Distribution,: *Biometrika*, v.32, pt. II, 1941, 187. The upper limit factors for the specified values of r correspond to $X^2/2$ with r = (ndf-1)/2, where ndf is the number of degrees of freedom. Remaining values are upper $(1-\alpha/2)\%$ points of the Poisson distribution. The upper limit factors are rounded up in the second decimal place for conservatism.

27. The lower limit factors are obtained as in the preceeding footnote, except that $X^2/2$ corresponds to r = ndf/2, and lower limit factors are rounded down in the second decimal place.

28. Computed from Table 3.1 as $(r_u(r)-r)/\sqrt{r}$.

29. Computed from Table 3.2 as $(r-r_l(r))/\sqrt{r}$.

CHAPTER IV

BASIC PRECISION

It has been observed that the sample standard deviation vanishes as the number of errors and/or the sum of the sample tainting rates approaches zero. The consequence is a trivial estimation interval, or at best an understatement of the sampling risk. This has been a principal shortcoming of the assumption of asymptotic normality of the sampling distribution when the sample discloses no error or few errors.

A fundamental property of the Stringer and related procedures for calculating confidence limits is their recognition of basic precision. Basic precision is the maximum amount of error in a population that could be undetected in a sample, at the specified detection risk. Basic precision is included in the calculation of upper confidence bounds even when a sample discloses errors. In doing so the auditor is, effectively, attempting to assess the impact on the sampling precision of two types of errors--those that were detected by the sample, and those that were not detected. Since nothing is known about the errors that were not detected, to the extent that such errors exist, their tainting rates are assigned the maximum possible value under the error assumptions, which is 100%. The basic precision thus provides an upper bound on the extent of these hypothetical errors in the population.

For a specified detection risk, the basic precision factor $r_u(0)$ is obtained by solving the Poisson expression

$$\exp(-r_u(0)) = \beta \tag{4.1}$$

for $r_u(0)$. Thus,

$$r_u(0) = -\ln(\beta). \tag{4.2}$$

Expressed in monetary terms for the population the basic precison is $(B/n)r_u(0)$. For example, if no error is detected in a dollar unit sample of size 100 that was selected from a population of $1,000,000, then the auditor can be 95% confident that the population is not overstated by more than $30,000 (rounded to 3 significant figures). Stated another way, if the population overstatement had exceeded $30,000, then the probability of failing to detect at least one overstated dollar unit in the sample would be no more than 5%. In this example, the basic precision is $30,000, which is the sampling precision associated with a point estimate of $0. The basic precision factor, $r_u(0)$, is 3.00 at 95% confidence (one-sided).

Although the Stringer and related bounds implicitly assume that the basic precision factor is constant, it can be shown that the factor is a function of the number and amount of the detected tainting rates in a sample.[1] Nevertheless, the relationship of the basic precision factor to the detected error condition has not been explored. However, procedures exist for calculating $r_u(0)$ directly in a limited number of situations, those in which there is one tainting or two equal taintings.[2] For one detected tainting r_d,

$$r_u(0) = -\ln(\beta)-\ln(1-r_d)-r_d/(1-r_d), \qquad 4.3$$

For two taintings in the sample, each equal to r_d,

$$r_u(0) = -\ln(\beta)-g+\ln(1+g+(g^2/2)), \qquad 4.4$$

where

$$g = (r_d+[r_d(2-r_d)]^{1/2})/(1-r_d). \qquad 4.5$$

Table 4.1 gives the basic precision factors at 95% confidence (one-sided) for one and two detected errors. Note that $r_u(0)$ is largest when the tainting rate is smallest, and the factor appear to decrease as the number of taintings increases. This suggests that the use of an invariant basic precision factor, as defined in equation 4.2, is conservative.

The confidence bound for the upper limit presented in this study also approaches zero as $r = \sum r_i$ approaches zero. This can be shown as follows. If $r_i < 1$ for $i=1,\ldots,m$, then $\sum r_i^2 < \sum r_i$. The upper limit precision factor, $t_u s_r$, is

$$(r_u(r)-r)/\sqrt{r}) \quad r_i^2 = (r_u(r)-r)\sqrt{\sum r_i^2/\sum r_i}\,. \tag{4.6}$$

The limiting values of the two components of this expression are

$$\lim_{r\to 0}(r_u(r)-r) = -\ln(\beta), \tag{4.7}$$

and

$$\lim_{r\to 0} \sum r_i^2/\sum r_i = 0. \tag{4.8}$$

Thus, the precision factor approaches zero, giving a trivial result. To eliminate this effect, we set the minimum value for the upper limit precision factor to be $-\ln(\beta)$, per equations 4.1 and 4.2.

TABLE 4.1
BASIC PRECISION FACTORS, $r_u(0)$
95% CONFIDENCE

SAMPLE TAINTING RATE	ONE TAINTING	TWO EQUAL TAINTINGS
0.0	2.996	2.996
0.1	2.990	2.973
0.2	2.969	2.912
0.3	2.924	2.799
0.4	2.840	2.605
0.5	2.689	2.274
0.6	2.412	1.688
0.7	1.866	0.560
0.7309	1.592	0.001
0.8	0.605	-
0.8259	0.000	-

CHAPTER IV FOOTNOTES

1. See Teitlebaum, A.D., "Dollar Unit Sampling in Auditing." paper presented at the 1973 annual meeting of the American Statistical Association, and Teitlebaum, A.D. et al., op.cit. Each paper included a formal derivation of the basic precision factor for a special case.

2. Ibid, and Teitlebaum et al., op.cit.

CHAPTER V

CALCULATING SAMPLE SIZE

The primary objective in planning an audit sample is to determine a sample size that is large enough to detect errors whenever the population misstatement exceeds the maximum tolerable amount. Approaches to calculating sample size for dollar unit samples are typically based on the attributes sampling techniques to which they are closely related. These vary from procedures for which the principal concern is the control of detection risk β [1] to procedures which also simultaneously attempt to control the risk of incorrect rejection[2] (which, in context, means the risk of concluding that error exceeds the maximum tolerable amount when, in fact, the error is acceptable, even if not zero). In this study I present an approach to sample size calculation that focuses on the control of β, but which also gives consideration to the estimation risk α and the ability to obtain sufficiently precise estimates when error is detected.

A discovery sample size is the smallest sample size for which detection of error is assured with at least $(1-\beta)\%$ probability, if the population error exceeds the maximum tolerable amount.[3]

Discovery sample sizes are popular among auditors because they are easily calculated and are the least expensive sample sizes that are capable of providing the audit assurance contemplated under the third standard of field work.

When planning a dollar unit sample it is convenient and conservative to assume that all errors are associated with 100% tainting rates. It is convenient because the properties of the Poisson distribution can be used in the sample size calculation. The conservatism of the assumption arises from the fact that, for a given amount of total error, 100% taintings are the most concentrated and, consequently, least likely to be detected. This can readily be shown as follows. The total amount of error E can be expressed in terms of the population book value B, the proportion of erroneous dollars p, and the average tainting rate r for the erroneous dollars as

$$E = prB. \qquad 5.1$$

Alternatively,

$$p = E/(rB) \qquad 5.2$$

Thus, we see that the rate of error p is smallest when the tainting rate r is greatest.

If we define $p_m = M/B$, which we interpret as the smallest proportion of dollars in the population that could contain an amount of error M, then the Poisson probability that a sample of size n will fail to detect any error is

$$= \exp(-np_m). \qquad 5.3$$

Solving for sample size, given detection risk and maximum tolerable error rate p_m, we have

$$n_o = -\ln(\beta)B/M, \qquad 5.4$$

where n_o is in this case the discovery sample size.

A well-known property of the Poisson distribution is that precision of the estimate deteriorates as the number of detected errors increases, because the standard error of the estimate, which is estimated from the sample, is the square root of the number of detected errors. Accordingly, the appropriateness of a particular sample size depends on the auditor's ability to allow for the occurrence of errors in the sample so that achieved precison will not exceed M.

A procedure for calculating a dollar unit sample size when errors are anticipated is readily obtained from a consideration of a basic property of the Poisson distribution. The (1-β)% upper precision of the estimate is given by the identity

$$r_u(m)-m = z_u r_u(m)^{1/2}, \qquad 5.5$$

where m is the number of errors detected (the point estimate) and Z_u is the standardized deviate $(r_u(m)-m)/r_u(m)^{1/2}$.

Thus,

$$(B/n)(r_u(m)-m) = (B/n)r_u m^{1/2}. \qquad 5.6$$

By making two substitutions, we can use this expression to calculate a sample size which is large enough to obtain precision no larger than M for any point estimate that is less than or equal to any specified amount E.

First, the left side of the identity is replaced by the desired precision M. Second, the variance of the distribution is replaced by the variance that would exist if E+M were to be the actual amount of error in the population. These substitutions would give

$$(B/n)(r_u(m)-m) = M \qquad 5.7$$

and

$$n(E+M)/B = s^2 = r_u(m). \qquad 5.8$$

Thus,

$$M = (B/n)\ z_u[n(E+M)/B]^{1/2}, \qquad 5.9$$

and

$$n = z_u^2(B/M)[1+(E/M)]. \qquad 5.10$$

Unlike the normal distribution, the value of the standardized deviate z_u of the Poisson distribution is a function of m, as illustrated in Table 5.1. When sample planning is based on the auditor's allowance for the occurrence of errors in the sample, using an upper bound for z_u^2 would give slightly conservative sample sizes. Upper bounds are given in Table 5.2.

Equation 5.10 shows that the sample size required to achieve precision no greater than M is a linear function of the auditor's allowance for the occurrence of error in the sample. If we are willing to forego the slight conservatism suggested in the previous paragraph and in Table 5.2, then we may set $z_u^2 = -\ln(\beta)$ for all levels of detection risk. In this case the sample size for any auditor error expectation becomes a linear function of the discovery sample size n_o.

Thus,

$$n = n_o[1 + (E/M)]. \qquad 5.11$$

We see that a discovery sample size will achieve the desired precision M without extended sampling only if no error is detected. If the sample size is two times the discovery level, then precision equal to M or better will be achieved, provided that the point estimate itself is no greater than M; and so on.

TABLE 5.1

STANDARDIZED DEVIATES z_u
CORRESPONDING TO (1-β)% UPPER POINTS
OF THE POISSON DISTRIBUTION

Number of Errors	95%	97.5%	99%
0	1.73	1.92	2.15
1	1.72	1.94	2.19
2	1.71	1.95	2.21
3	1.71	1.95	2.22
4	1.70	1.95	2.23
5	1.70	1.95	2.24
10	1.69	1.96	2.26
20	1.68	1.96	2.28
30	1.68	1.96	2.29
40	1.67	1.96	2.29
50	1.67	1.96	2.29
100	1.66	1.96	2.30
∞	1.64	1.96	2.32

TABLE 5.2

UPPER BOUNDS FOR z_u2
WHEN ERRORS ARE EXPECTED

RISK $\beta(\alpha)$	PRECISION FACTOR z_u2
5% (10%)	3.00
2.5% (5%)	3.84
1% (2%)	5.38
3.1443%	$-\ln(\beta)$
3.1443%	The standard normal deviate squared for one-tail probability

CHAPTER V FOOTNOTES

1. See Meikle, G.R., op. cit., and Leslie, D.A., A.D. Teitlebaum and R. Anderson, op. cit. for approaches that are used in practice.

2. Kaplan, R.S., "Sample Size Computation for Dollar Unit Sampling", Journal of Accounting Research: Studies on Statistical Methodology in Auditing, 1975, 126-133.

3. The term "discovery sampling" was coined by Arkin. See Arkin, H., Handbook of Sampling for Auditing and Accounting, 2nd Edition, New York: McGraw-Hill, 1974, p.15.

CHAPTER VI

EXTENDING SAMPLES FOR INTERVAL ESTIMATION

It will be recalled that the auditor's objective is to conclude that error in the population does not exceed the specified maximum tolerable amount. Occasionally, however, the precision obtained from the audit sample will not be sufficiently precise to enable the auditor to achieve the test objective at the specified confidence level. This situation occurs when the population standard deviation estimated from the sample exceeds the estimate that was peliminarily applied in the calculation of sample size. Under the error conditions that are the subject of this study, the estimated standard deviation is highly correlated with the number of errors disclosed in the sample. Thus, the situation arises when errors are more prevalent than assumed by the auditor when he planned the audit sampling procedure. In such an event, this suggests that an appropriate course of action would be to continue sampling until sufficient precision is achieved.

The procedure developed in this study enables the auditor to achieve his test objective by obtaining sufficient information to permit adjustment of the affected account by an amount that will reduce remaining misstatement to a magnitude that does not exceed the maximum tolerable error.

Thus, in addition to developing a method for obtaining reliable confidence limits for estimates of relatively infrequently occurring errors, this study is also concerned with the problem of sequential estimation by double sampling. As Wilks states, "(t)he basic idea of sequential estimation is in general to do just enough sampling to be able to obtain an estimate which has a predetermined degree of precision, in some sense which does not depend on the population parameter being estimated."[1] For the audit sampling problem, the predetermined degree of precision is the auditor's specified maximum tolerable error.

In auditing, sequential sampling procedures typically have been applied in accept/reject decision problems with respect to internal control tests of compliance.[2] The principal objective of such a procedure is a decision to accept or reject some proposition as to the effectiveness of the tested controls. If the errors exceed a specified acceptance number in an initial sample but are less than a specified rejection number, additional sample items are selected. Several sets of additional sample items may be selected before a decision is taken. Sequential sampling plans for proportion tend to be highly structured because the computations are complex, involving consideration of conditional probabilities of the binomial or Poisson distribution. An example of such a procedure is given by

Roberts.[3] In this study, however, we are not seeking an accept/reject decision. Instead, the objective is to estimate the amount of error in the population, whenever error is detected, as the basis for proposing a sufficiently precise adjustment to the population's book value.

The sequential procedure employed in this study is a double sampling plan. The basic methodology is due to Stein, who developed a double sampling procedure for obtaining a confidence interval of fixed length when sampling from a normal population where both mean and variance are unknown.[4] Under Stein's procedure we would proceed as follows. If the auditor's desired precision is M and if the results of an initial sample n, which was selected from a population of N items, are such that

$$Nst(n-1,\alpha)/\sqrt{n} \leq M, \qquad 6.1$$

where $t(n-1,\alpha)$ is Student's-t with n-1 degrees of freedom and estimation risk α, then no additional sampling is necessary. Otherwise, an additional sample of n_1-n items is selected, giving a combined sample of size n_1, such that

$$Nst(n-1,\alpha)/\sqrt{n_1} \leq M. \qquad 6.2$$

The confidence interval of the estimated total value of the population, based on the combined results of both samples, will have precison $\pm M$ at $(1-\alpha)\%$ confidence.[5]

Because this study involves the use of sampling primarily as an interval estimation procedure, Stein's method has considerable appeal, which is enhanced by its simplicity. Nevertheless, there is no indication in the literature that Stein's method has been explored for applicability in accounting and auditing.

This procedure is readily reformulated for our audit sampling application. Under the assumption that the sampling distribution is approximately a compound Poisson distribution, it is observed that the confidence limits are not equidistant from the point estimate. For the cases that are of interest herein, the upper precision factor pf_u is larger than the lower precision factor pf_l. The target achieved precision factor is equal to M/B, by definition. Moreover, the estimation risk α is fixed relative to the detection risk β. That is, $\alpha = 2\beta$.

Extended Sample Size

If, upon evaluating the initial sample, $UEL-\hat{E}>M$, then additional sample items are selected. The total sample size n_1 is the smallest integer which satisfies

$$n_1 = n[(UEL-\hat{E})/M]^2 = n[Bpf_u/(nM)]^2. \qquad 6.3$$

Point Estimate

When the additional items have been selected and their error amount determined, the total equivalent error of the combined sample is calculated as

$$r_1 = \sum^{m_1} r_i, \qquad 6.4$$

where m_1 is the number of errors detected in both samples. The final point estimate of the total error is

$$\hat{E}_1 = (B/n_1)r_1. \qquad 6.5$$

It was previously stated that the behavior of the sampling distribution of estimates of monetary error resembles the Poisson sampling distribution of estimates of attribute data. When double sampling for attribute estimation, the point estimate is known to be slightly biased, a correction for which was developed by Cox.[6] Cox, working with the estimator of the binomial parameter, showed that uncorrected estimator would be biased above the parameter.[3] This would, of course, raise the lower limit increasing its coverage of the parameter. Using Cox's bias adjustment to his binomial parameter estimator as our basis, the following adjusted estimator $\hat{E}_1'$ was used.

$$\hat{E}_1' = E_1(1 - [\sum^{m} r_i^2/n]) \qquad 6.6$$

Precision

As in the initial sample, precision is calculated separately for the upper and lower confidence limits, based on the intitial upper and lower precision factors pf_u and pf_l, respectively, and the precision reduction factor, $(n/n_1)^{1/2}$.

Thus, the final upper and lower precisons are

$$P_{u_1} = (B/n)pf_u(n/n_1)^{1/2},$$

and 6.7

$$P_{l_1} = (B/n)pf_l(n/n_1)^{1/2},$$

respectively.

Confidence Limits

Finally, the upper and lower confidence limits on the estimated population error are

$$UEL = \hat{E}_1 + P_{u_1}$$

and 6.8

$$LEL = \hat{E}_1 - P_{l_1},$$

respectively.

Teitlebaum noted, in a somewhat different context, that the precision that is ultimately obtained cannot be less than the basic precision without incurring a degradation of the confidence level.[7] The formulation presented herein eliminates this problem by 1) ensuring that the initial sample size is at least as large as $(B/M)(-\ln\beta)$, which is the discovery sample size, and 2) by specifying the target precision for the extended sample, is not less than M.

CHAPTER VI FOOTNOTES

1. Wilks, S.S., Mathematical Statistics, New York: John Wiley & Sons, 1960, 496.

2. Bailey, A.D., Jr., Statistical Auditing, New York: Harcourt Brace Jovanovich, 1981, 235-260.

3. Roberts, D.M. "A Proposed Sequential Sampling Plan for Tests of Compliance," Symposium on Auditing Research, Urbana-Champaign: University of Illinois, 1976, 159-168.

4. Stein, C, "A Two-Sample Test for a Linear Hypothesis Whose Power is Independent of the Variance, Annals of Mathematical Statistics 1945, V.16, 243-258.

5. Wilks, S.S., op.cit., 497-498.

6. Cox, D.R., "Estimation by Double Sampling," Biometrika, 1952, V.39, 217-227.

7. Teitlebaum, A.D. op. cit., Appendix III, 27-31.

CHAPTER VII

TESTING THE SAMPLING PROCEDURE

This chapter describes the conditions and the methodology under which the sampling procedure was tested. The purpose of the test was to measure the achieved risk levels of the procedure under a variety of error conditions. The ability of the procedure to perform with achieved risk levels that are no greater than the specified, or nominal, levels means that the procedure is one that the auditor can rely upon to satisfy the third standard of field work.

Error Characteristics of Accounting Populations

In order to develop appropriate tests of audit sampling procedures, it is desirable to have some knowledge of the error characteristics of the accounting populations to be audited and to which the procedures are to be applied. Yet, little has been published that deals with such error characteristics. Neter and Loebbecke[1] examined four accounting populations. Ramage et al[2] did a broad analysis of accounting errors, but without consideration of their distributional aspects. Johnson et al[3] have performed the most detailed analysis to date, a study of the error characteristics of seventy-seven accounting populations. A significant finding of theirs was that error rate tended to be positively correlated with item value, a condition which favors dollar unit sampling vis-a-vis, say, simple random sampling. Another finding was the tendency of some of the error distributions to resemble reversed J-shaped distributions, containing many small taintings and few large tainting including, possibly, 100% taintings.

Johnson et. al. acknowledged the essentially anecdotal nature of their study and the tenuousness of any generalizations about accounting error conditions. Nevertheless, their identification of the reverse J-shaped distribution is significant, because it is under this error condition that a number of procedures, including especially those based on the assumption of asymptotic normality in the sampling distribution, have been found to fail.[4]

Error Populations Used in This Study

In order to examine the performance of the procedure developed in this study, eighty-four study populations were created. Error conditions in these populations varied as to total amount E and distribution of taintings. Total error amounts were set at three levels relative to a specified maximum tolerable error amount M. These levels were set at 0.5M, 1.0M, and 1.5M in order to observe the detection/estimation properties of the procedure errors of varying seriousness in single and double sampling at $\beta = \alpha/2 = 5\%$, and in single sampling at $\beta = \alpha/2 = 1\%$.

For definiteness, the book value B of each population was set to $1,000,000, and a sample size of 50 dollar units was chosen to be the discovery sample size. Thus, M was set to $(B/n)(-\ln\beta)$; that is, to $59,920 for tests of the procedure at B = 5%, and to $92,120 for tests at β = 1% (the rounded values of $\ln(\beta)$ that were used were -2.996 and -4.606, respectively).

For each total error amount, fourteen error distributions of tainting rates were chosen; including one binomial, four trinomial, and nine quadrinomial distributions. In each case the tainting rates were chosen so as to provide a broad representation of the kinds of error conditions that would exist if there were to be failure of financial statement assertions with respect to existence or valuation. Thus, the presence of 100% taintings reflects failure of the existence assertion. Items that are 100% overstated may be fictitious, fraudulent, wholely worthless, or not owned (such as consigned goods that are included in inventory). At the other extreme are items that are overstated by relatively small amounts, chosen in this study to be 2% taintings. Errors of this type would include overstated freight or storage charges, sales taxes, or unrecorded discounts. Between these extremes are the valuation errors that result from lower of cost or market adjustments, incorrect pricing of sales transactions, adjustments to receivables arising from failure to give credit for returned merchandise, and the like. The tainting rates for these types of errors were varied from 20% to 80%. Table 7.1 gives the error distributions for the single and double sampling test at $\beta = \alpha/2 = 5\%$. Table 7.2 gives the distributions for the single sampling test at $\beta = \alpha/2 = 1\%$.

TABLE 7.1
POPULATION ERROR DISTRIBUTIONS
B = \$1000000; M =\$59920

CASE	TOTAL ERROR	ERROR AMOUNTS BY TAINTING RATE 1.00	.80	.50	.20	.02
1	89880	89880	0	0	0	0
2	89880	88880	0	0	0	1000
3	89880	67500	0	0	21500	880
4	89889	45000	0	0	44000	880
5	89880	22500	0	0	66500	880
6	89880	0	0	0	89000	880
7	89880	67500	0	21500	0	880
8	89880	45000	0	44000	0	880
9	89880	22500	0	66500	0	880
10	89880	0	0	89000	0	880
11	89880	67500	21500	0	0	880
12	89880	45000	44000	0	0	880
13	89880	22500	66500	0	0	880
14	89880	0	89000	0	0	880
15	59921	59921	0	0	0	0
16	59921	59000	0	0	0	921
17	59921	45000	0	0	14000	921
18	59921	30000	0	0	29000	921
19	59921	15000	0	0	44000	921
20	59921	0	0	0	59000	921
21	59921	45000	0	14000	0	921
22	59921	30000	0	29000	0	921
23	59921	15000	0	44000	0	921
24	59921	0	0	59000	0	921
25	59921	45000	14000	0	0	921
26	59921	30000	29000	0	0	921
27	59921	15000	44000	0	0	921
28	59921	0	59000	0	0	921
29	29961	29961	0	0	0	0
30	29961	29000	0	0	0	961
31	29961	22500	0	0	6500	961
32	29961	15000	0	0	14000	961
33	29961	6500	0	0	22500	961
34	29961	0	0	0	29000	961
35	29961	22500	0	6500	0	961
36	29961	15000	0	14000	0	961
37	29961	6500	0	22500	0	961
38	29961	0	0	29000	0	961
39	29961	22500	6500	0	0	961
40	29961	15000	14000	0	0	961
41	29961	6500	22500	0	0	961
42	29961	0	29000	0	0	961

TABLE 7.2
POPULATION ERROR DISTRIBUTIONS
B = $1000000; M =$92120

CASE	TOTAL ERROR	ERROR AMOUNTS BY TAINTING RATE 1.00	.80	.50	.20	.02
43	138180	138180	139180	0	0	0
44	138180	137000	0	0	0	1180
45	138180	102750	0	0	34250	1180
46	138180	68500	0	0	68500	1180
47	138180	34250	0	0	102750	1180
48	138180	0	0	0	137000	1180
49	138180	102750	0	34250	0	1180
50	138180	68500	0	68500	0	1180
51	138180	34250	0	102750	0	1180
52	138180	0	0	137000	0	1180
53	138180	102750	34250	0	0	1180
54	138180	68500	68500	0	0	1180
55	138180	34250	102750	0	0	1180
56	138180	13180	137000	0	0	1180
57	92121	92121	0	0	0	0
58	92121	91000	0	0	0	1121
59	92121	68250	0	0	22750	1121
60	92121	45500	0	0	45500	1121
61	92121	22750	0	0	68250	1121
62	92121	0	0	0	91000	1121
63	92121	68250	0	22750	0	1121
64	92121	45500	0	45500	0	1121
65	92121	22750	0	68250	0	1121
66	92121	0	0	91000	0	1121
67	92121	68250	22750	0	0	1121
68	92121	45500	45500	0	0	1121
69	92121	22750	68250	0	0	1121
70	92121	0	91000	0	0	1121
71	46061	46061	0	0	0	0
72	46061	45000	0	0	0	1061
73	46061	33750	0	0	0	1061
74	46061	22500	0	0	22500	1061
75	46061	1250	0	0	33750	1061
76	46061	0	0	0	45000	1061
77	46061	33750	0	11250	0	1061
78	46061	22500	0	22500	0	1061
79	46061	11250	0	33750	0	1061
80	46061	0	0	45000	0	1061
81	46061	33750	11250	0	0	1061
82	46061	22500	22500	0	0	1061
83	46061	11250	33750	0	0	1061
84	46061	0	45000	0	0	1061

Description of the Testing Approach

The most common approach to testing the reliability of audit sampling procedures has been to employ Monte Carlo simulation of the selection of several hundred samples of specified size from each of a variety of simulated accounting populations.[5] Some researchers have chosen to rely on mathematical analysis.[6] In this study I have chosen enumeration of the sample space as the test procedure. Enumeration entails identifying each sample outcome, calculating the probability of its occurrence, and determining whether the outcome satisfies specified test conditions. This approach eliminates the uncertainty introduced by the sampling error that is associated with simulation.

In the studies published to-date, researchers have examined only the probability that the upper error limit fails to cover the actual population error. No consideration has been given to the performance of the lower error limit. Only one study gave any consideration to the audit consequences of not "accepting" the population book value.[7] Consequently, we have no indication that any of the previously reported methods provides a reliable basis for proposing adjustments, since it is the lower limit that would provide assurance against booking too large an adjustment.

In this study consideration is given to the lower limit and to the consequences of error detection, with respect to the total effort (sample size) required to obtain the sufficient evidential matter required by the third standard of field work. Consideration is also given to the reliability of the procedure as a basis for an adjusting journal entry in double sampling.

Assuming that the non-zero tainting rates $\{r_i\}$ in a population take on a finite number of unique values, each tainting rate occurring with frequency p_i, the joint probability of obtaining exactly m_i taintings of each kind i in a sample of n dollar units is given by the multiple Poisson distribution as

$$\Pr\{m_1, m_2, \ldots\} = \prod (np_i)^{m_i} e^{-n(\sum p_i)} \Big/ \prod m_i! \, , \qquad 7.1$$

where $m_i \leq n$ and $p_i \leq 1$ (typically, $p_i \ll 1$).

If S is a subset of the sample space for which the sample outcomes that are members of S satisfy a specified condition T, then

$$\Pr\{S\} = \sum \Pr\{m_1, m_2, \ldots | T\}. \qquad 7.2$$

For double sampling two conditions determine the subset S_1 of interest in the extended sample space. The condition T is the occurrence of a sample outcome in the initial sample for which UEL - E > M. The condition T is any condition of interest that pertains to the results of the second sample, and is dependent on the occurrence of T. The formation of the outcome probabilities associated with T_1 is based on the multiple Poisson distribution with incremental sample size n_1-n. Denoting m_{1i} as the number of taintings of kind i occurring in the incremental sample, we have

$$\Pr\{S_1\} = \sum [\Pr\{m_1, m_2, \ldots | T\} \sum \Pr\{m_{11}, m_{12}, \ldots | T, T_1\}]. \quad 7.3$$

Outcome probabilities were calculated for each of the test conditions. The test conditions are described in Table 7.3.

TABLE 7.3

TEST CONDITIONS

Single Sampling

	Condition	Description
1.	E.GT.UEL	Upper limit failure.
2.	E.LT.LEL	Lower limit failure.
3.	UEL.LE.M	"Accept" population.
4.	(E.LE.UEL).AND.(E.GE.LEL)	Actual error within interval.
5.	$(UEL-\hat{E})$.LE.M	Additional sampling not needed.

Double Sampling

	Condition	Description
1.	E.GT.UEL_1	Upper limit failure.
2.	E.LT.LEL_1	Lower limit failure.
3.	(E.LE.UEL_1).AND.(E.GE.LEL_1)	Actual error within interval.
4.	$\lvert Max\{UEL_1, LEL_1+M\} - E\rvert$.LE.M .AND. $\lvert Min\{LEL_1, UEL_1-M\}-E\rvert$.LE.M	Post-adjustment error does not exceed maximum tolerable amount.

Note that the double sampling test conditions include the single sampling outcomes where single sampling condition 5 is satisfied. That is, the double sampling results pertain to the entire procedure.

CHAPTER VII FOOTNOTES

1. Neter, J. and J.K. Loebbecke, op. cit.

2. Ramage, J.G., A.M. Krieger, and L.L. Spero, "An Empirical Study of Error Characteristics in Audit Populations," Studies on Auditing, Supplement to Journal of Accounting Research, 1979, 72-102.

3. Johnson, J.R., R.A. Leitch, and J. Neter, "Characteristics of Errors in Accounts Receivable and Inventory Audits," Accounting Review, April 1981, 270-293.

4. Leslie, D.A. and S.J. Aldersley, "Discussants' Response to The Behavior of Selected Upper Bounds of Monetary Error using PPS Sampling," Symposium on Auditing Research IV, Urbana - Champaign: University of Illinois, 1982, 387-400.

5. Neter, J. and J.K. Loebbecke, op.cit.; Reneau, J.H., "CAV Bounds in Dollar Unit Sampling: Some Simulation Results," Accounting Review, July 1978, 669-680; Garstka, S.J. and P.A. Ohlson, op. cit.; MacGuidwin, M.J., D.M. Roberts and M. Shedd, "The Behavior of Selected Upper Bounds of Monetary Error Using PPS Sampling," Symposium on Auditing Research IV, Urbana - Champaign: University of Illinois, 1982, 387-400; and Jenne, S.E., "The Effects of Three Methods of Sample Selection on Dollar Unit Sampling Bounds," Symposium on Auditing Research V, Urbana-Champaign: University of Illinois, 1983, 163-200.

6. Neter, J., R.A. Leitch, and S.E. Fienberg, "Dollar Unit Sampling: Multinomial Bounds for Total Overstatement and Understatement Errors," Accounting Review, January 1978, 77-93; Leslie, D.A., A.D. Teitlebaum and R.J. Anderson, op.cit., p.254; McCray, J.H., op. cit., and Dworin, L. and R.A. Grimlund, "Dollar Unit Sampling for Accounts Receivable and Inventory," Accounting Review, April 1984, 218-241.

7. Garstka, S.J. and P.A. Ohlson, op.cit.

8. Feller, W., An Introduction to Probability Theory and Its Applications, Vol. 1 3rd Ed., New York: John Wiley & Sons, 1968, p.172.

CHAPTER VIII

TEST RESULTS

Performance of the Confidence Limits - Single Sampling

The performance of the compound Poisson assumption as a basis for calculating confidence limits was examined on both a one-sided and two-sided basis. On a one-sided basis, the upper and lower confidence limits were considered separately. Both limits were combined into an interval estimate for consideration of the procedure's performance on a two-sided basis.

For the upper limit UEL, risk is defined as the proportion of sample outcomes for which the actual population error amount E was greater than the upper error limit UEL; that is, $\Pr\{E\} > \mathrm{UEL}$. For the lower limit LEL, the risk was similarly defined as $\Pr\{E\} < \mathrm{LEL}$. Finally, the achieved two-sided confidence level is defined as the proportion of sample outcomes for which the actual population error was between the calculated limits; that is, $\Pr\{(E \geq \mathrm{LEL}).\mathrm{AND}.(E \leq \mathrm{UEL})\}$.

It is well known that, even in a conventional binomial situation where exact procedures are available for construction of confidence limits, the achieved risk levels equal nominal risk levels only in special cases. Generally, the achieved risk levels are always less than the nominal levels.[1] This phenomenon was observed also in this study when all taintings were 100%.

The upper confidence limit was robust over all the error conditions and sample sizes to which it was applied. In all cases the measured risk associated with the upper limit was no greater than the specified one-sided risk.

In the tests of the lower limit the actual risk exceeded the nominal risk under four of the error conditions that were the basis for the 5% risk level tests, and in six cases for the 1% risk level tests. These cases (numbers 6, 20, 24, and 30 for the 5% tests, and numbers 48, 52, 62, 66, 76, and 80 for the 1% tests) involved similar error distributions. None of the affected cases that were tested at the 5% level had tainting rates that were greater than 20%; and none of the cases that were tested at the 1% level had tainting rates that were greater than 50%. Thus, the affected cases were those that had the smallest standard deviations within their respective error amount categories. In all cases the performance of the lower limit was poorest under the discovery sample size of 50 dollar units; when the sample size was double the discovery level (n=100), performance of the lower limit improved. Under discovery sampling the actual risk level for these cases varied as follows: from

6.5% to 7.0%, when the nominal risk was 5%; and from 1.2% to 3.7%, when nominal risk was 1%. When double the discovery sample size was selected the actual risk levels varied as follows: from 5.4% to 6.4%, with 5% nominal; and from 1.1% to 2.2% with 1% nominal.

Despite the aforementioned slight deterioration of the lower limits, the achieved two-sided confidence level was lower than the specified nominal level only in the 98% test (for cases 48, 62, and 76), as shown in Table 8.1. Thus, the compound Poisson assumption appears to provide a suitable, if not perfectly robust, basis for estimating the amount of error in single sampling.

TABLE 8.1
CONFIDENCE INTERVAL
SINGLE SAMPLING
$\alpha/2 = 5\%$; $1-\alpha=90\%$

CASE	n = 100: E > UEL	n = 100: E < LEL	n = 100: E ≥ LEL AND E ≤ UEL	n = 50: E > UEL	n = 50: E < LEL	n = 50: E ≥ LEL AND E ≤ UEL
1	2.1%	4.1%	93.8%	1.1%	4.0%	94.9%
2	2.3	3.8	93.9	1.2	3.8	95.1
3	3.8	3.7	92.5	2.8	3.2	93.9
4	4.4	3.6	91.9	1.5	3.2	95.2
5	3.4	3.6	92.9	0.2	3.4	96.4
6	0.9	5.5	93.6	0.0+	6.5	93.5
7	3.6	3.8	92.7	2.2	3.5	94.3
8	3.3	3.9	92.8	2.0	3.6	94.4
9	3.3	4.1	92.6	1.3	3.9	94.9
10	2.9	4.1	93.0	0.7	3.9	95.5
11	3.5	3.6	92.9	2.1	3.6	94.3
12	3.3	3.9	93.0	2.5	3.4	94.1
13	2.9	4.0	93.1	2.6	4.1	93.3
14	3.5	3.5	93.0	2.5	2.9	94.6
15	1.7	4.2	94.0	5.0	3.3	91.7
16	1.9	3.9	94.2	0.5	3.1	96.4
17	4.7	3.5	91.8	0.0+	3.1	96.8
18	3.3	3.3	93.4	0.0+	3.1	96.8
19	1.1	3.5	95.4	0.0+	3.6	96.4
20	0.1	5.7	94.1	0.0	7.0	93.0
21	3.4	3.7	92.9	0.3	3.4	96.4
22	3.4	3.7	92.8	0.1	3.4	96.5
23	3.1	4.0	92.9	0.1	3.4	96.0
24	2.3	5.4	92.3	0.0+	3.9	96.1
25	3.2	3.6	93.2	0.4	3.8	95.8
26	3.3	3.7	93.0	0.4	3.5	96.2
27	2.9	3.7	93.5	0.3	2.9	96.8
28	2.2	3.8	93.9	0.3	3.5	96.3
29	5.0	3.3	91.7	0.0	1.8	98.2
30	0.0+	2.9	97.1	0.0	1.6	98.4
31	0.0+	3.1	96.9	0.0	3.0	97.0
32	0.0+	3.0	97.0	0.0	3.0	97.0
33	0.0+	3.3	96.7	0.0	3.7	96.3
34	0.0+	6.4	93.6	0.0	7.0	93.0
35	0.0+	3.3	96.7	0.0	2.8	97.2
36	0.0+	3.4	96.6	0.0	3.2	96.8
37	0.0+	3.7	96.3	0.0	3.9	96.1
38	0.0+	3.9	96.1	0.0	4.8	95.2
39	0.0+	3.6	96.4	0.0	2.0	98.0
40	0.0+	3.7	96.3	0.0	2.5	97.5
41	0.0+	3.1	96.9	0.0	3.1	96.9
42	0.0+	3.2	96.9	0.0	3.7	96.3

TABLE 8.2
CONFIDENCE INTERVAL
SINGLE SAMPLING
$\alpha/2 = 1\%$; $1-\alpha = 98\%$

CASE	n = 100			n = 50		
	E > UEL	E < LEL	E > LEL AND E ≤ UEL	E > UEL	E < LEL	E > LEL AND E ≤ UEL
43	0.6%	0.8%	98.6%	0.8%	0.5%	98.7%
44	0.7	0.7	98.6	0.8	0.5	98.7
45	0.9	0.7	98.4	0.7	0.6	98.6
46	0.9	0.7	98.4	0.7	0.6	98.6
47	0.9	0.7	98.4	0.0+	0.6	99.3
48	0.2	1.4	98.2	0.0+	2.5	97.6
49	0.7	0.8	98.6	0.5	0.7	98.8
50	0.7	0.8	98.5	0.6	0.8	98.6
51	0.5	0.9	98.6	0.4	0.9	98.7
52	0.4	1.1	98.5	0.2	1.3	98.4
53	0.6	0.8	98.6	0.6	0.7	98.7
54	0.6	0.8	98.6	0.4	0.7	98.9
55	0.6	0.8	98.6	0.3	0.8	98.9
56	0.5	1.0	98.5	0.2	0.7	94.1
57	0.5	0.7	98.8	1.0	0.9	98.2
58	0.6	0.6	98.8	0.1	0.7	99.2
59	0.9	0.7	98.4	0.0+	0.6	99.4
60	0.1	0.6	98.4	0.0	0.6	99.4
61	0.2	0.7	99.1	0.0	0.7	99.3
62	0.0+	1.7	98.2	0.0	2.5	97.5
63	0.6	0.7	98.6	0.0+	0.7	99.3
64	0.6	0.8	98.6	0.0+	0.8	99.2
65	0.4	0.9	98.7	0.0+	0.9	99.1
66	0.3	1.2	98.6	0.0+	1.2	99.8
67	0.6	0.7	98.7	0.0+	0.7	99.3
68	0.6	0.8	98.6	0.0+	0.7	99.3
69	0.5	0.8	98.7	0.0+	0.9	99.1
70	0.3	0.8	98.9	0.1	0.6	99.2
71	1.0	0.8	98.2	0.0	0.9	99.1
72	0.0+	0.7	99.3	0.0	0.8	99.2
73	0.0+	0.6	99.4	0.0	0.5	99.5
74	0.0	0.6	99.4	0.0	0.5	99.5
75	0.0	0.7	99.2	0.0	0.9	99.1
76	0.0	2.2	97.7	0.0	3.7	96.3
77	0.0+	0.7	99.3	0.0	0.6	99.4
78	0.0+	0.8	99.2	0.0	0.7	99.3
79	0.0+	0.9	99.1	0.0	0.9	99.1
80	0.0+	1.1	98.9	0.0	1.7	98.3
81	0.0+	0.7	99.3	0.0	0.6	99.4
82	0.0+	0.6	99.4	0.0	0.5	99.5
83	0.0+	0.9	99.1	0.0	0.6	99.4
84	0.0+	0.7	99.3	0.0	0.8	99.1

Detection Ability of the Upper Error Limit

There remains considerable interest among auditors in the ability of the calculated upper confidence limit UEL to lead the auditor to correctly conclude as to the existence of error that exceeds the specified maximum tolerable amount M. Consequently, the performance of the upper error limit was also tested in this context, despite its lack of relevance to the overall estimation procedure under study herein. Tables 8.3, and 8.4 give the proportion of "accept" decisions (that is, outcomes for which UEL $\leq$ M) for each case and sample size.

The results are not surprising. The probability of "acceptance" decreases as the amount of error increases. For a given amount of error, the proportion of "acceptance" decreases with the standard error, whether due to population standard deviation or to sample size. In none of the cases where the error amount exceeded M did the proportion of "acceptance" exceed the specified detection risk.

The test results show that the acceptance rule - accept the population as being fairly stated if UEL $\leq$ M - is generally uninformative unless the population is virtually free of error. This is particularly evident where discovery sample sizes were applied to cases in which the error amount was M/2. In only two of those cases (numbers 29 and 71), where all tainting rates were 100%, was there a significant proportion of "acceptances" of the population. Thus, an auditor who uses such a decision rule would generally find that it does not provide a useful basis for discriminating between serious and tolerable error conditions.

TABLE 8.3
"ACCEPTANCE" DECISIONS
$\beta = 5\%$

ERROR AMOUNT	CASE	n = 100 UEL ≤ M	n = 50 UEL ≤ M
1.5M	1	0.1%	1.1%
1.5M	2	0.1	0.1
1.5M	3	0.3	0.0+
1.5M	5	0.0+	0.0
1.5M	6	0.0	0.0
1.5M	7	0.2	0.0+
1.5M	8	0.2	0.0+
1.5M	9	0.1	0.0+
1.5M	10	0.0+	0.0+
1.5M	11	0.3	0.1
1.5M	12	0.2	0.1
1.5M	13	0.2	0.1
1.5M	14	0.1	0.0+
M	15	1.7	5.0
M	16	1.9	0.5
M	17	4.7	0.0+
M	18	3.3	0.0+
M	19	1.1	0.0+
M	20	0.1	0.0
M	21	3.4	0.3
M	22	3.4	0.1
M	23	3.1	0.1
M	24	2.3	0.0+
M	25	3.2	0.4
M	26	3.3	0.4
M	27	2.9	0.3
M	28	2.2	0.3
.5M	29	20.0	22.4
.5M	30	21.5	2.1
.5M	31	34.0	0.6
.5M	32	46.5	0.1
.5M	33	50.0	0.0+
.5M	34	51.5	0.0+
.5M	35	30.8	1.5
.5M	36	36.6	1.1
.5M	37	41.8	0.7
.5M	38	47.8	0.5
.5M	39	29.1	2.0
.5M	40	32.6	1.8
.5M	41	32.3	1.6
.5M	42	29.8	1.5

TABLE 8.4
"ACCEPTANCE" DECISIONS
$\beta = 1\%$

ERROR AMOUNT	CASE	n = 100 UEL ≤ M	n = 50 UEL ≤ M
1.5M	43	0.0+	0.0+
1.5M	44	0.0+	0.0
1.5M	45	0.0+	0.0
1.5M	46	0.0	0.0
1.5M	47	0.0	0.0
1.5M	48	0.0	0.0
1.5M	49	0.0+	0.0+
1.5M	50	0.0+	0.0+
1.5M	51	0.0+	0.0+
1.5M	52	0.0+	0.0
1.5M	53	0.0+	0.0+
1.5M	54	0.0+	0.0+
1.5M	55	0.0+	0.0+
1.5M	56	0.0+	0.0+
M	57	0.5	1.0
M	58	0.6	0.1
M	59	0.9	0.0+
M	60	1.0	0.0
M	61	0.2	0.0
M	62	0.0+	0.0
M	63	0.6	0.0+
M	64	0.6	0.0+
M	65	0.4	0.0+
M	66	0.3	0.0+
M	67	0.6	0.0+
M	68	0.6	0.0+
M	69	0.6	0.0+
M	70	0.3	0.0+
.5M	71	16.2	10.0
.5M	72	17.4	0.7
.5M	73	25.2	0.1
.5M	74	33.6	0.0+
.5M	75	47.1	0.0+
.5M	76	51.3	0.0+
.5M	77	20.7	0.4
.5M	78	23.9	0.2
.5M	79	28.0	0.1
.5M	80	32.4	0.7
.5M	81	17.6	0.6
.5M	82	19.9	0.6
.5M	83	20.6	0.5
.5M	84	18.8	0.4

Double Sampling - Extent of Sampling

The double sampling procedure was proposed as a practical alternative to a rule-based acceptance procedure which gives no consideration to the lower error limit. From the results of the previous section, it is evident that the rule-based approach leaves the auditor in a quandry as to appropriate action to be taken if "acceptance" is not achieved. The approach taken in this study employs interval estimation as the basis for auditor action if the sample discloses error. In one key respect the approach presented herein resembles the rule-based approach: if the sample discloses no error, then no further action is taken. Otherwise, the estimation approach provides the auditor with a course of action that results in a definitive outcome: adjustment of the affected account to reduce remaining error to an amount considered not to exceed M. As stated in Chapter II, in order to achieve this result, the auditor needs to obtain an interval estimate that possesses at least one interior point whose distance from both error limits is no greater than M. It must be expected that the initial sample size may be too small to satisfy this requirement. Hence, the reason for double sampling.

It was also observed in the previous section that the rule-based approach to audit sampling is uninformative and undiscriminating. However, by adopting the interval estimation approach, the same sample information gives clear indications of the extent of work that would be required to satisfy the third standard of field work. Table 8.5 gives the proportion of samples that would not require extension for each of the cases and sample sizes under study. We see that the discovery sample size typically requires extension in order to improve precision for adjustment purposes. Table 8.6 shows that when the error amount is 1.5M, discovery sample sizes were sufficient for median proportions of 2.7% of sample outcomes at $\alpha/2=5\%$, and 0.3% of sample outcomes at $\alpha/2=1\%$. When the error amount was only 0.5M, the median proportion was 32.4% of outcomes at $\alpha/2=5\%$, and 18.5% of sample outcomes at $\alpha/2=1\%$. However, when the initial sample size was twice the discovery level, the median proportion of outcomes for which single sampling was sufficient was 52.6% for the cases where the error amount was 1.5M. Where the total error was 0.5M the medians were 99.4% at $\alpha/2=5\%$ and 99.1% at $\alpha/2=1\%$.

Table 8.7 gives the average sample sizes required under the double sampling procedure, which are summarized in Table 8.8. Table 8.8 shows that the discovery sample size tends to be more efficient as an initial sample size when the population error is small, but less efficient in the presence of large error amounts. Under discovery sampling, the median required sample size under the double sampling procedure varied from 74 at $\alpha/2 = 5\%$ and 81 at $\alpha/2 = 1\%$, when error was 0.5M, to 125 at $\alpha/2 = 5\%$ and 155 at $\alpha/2$ at 1%, when error was 1.5M. For an initial sample size at twice the discovery level, the median required sample size varied from 100 to 112 at $\alpha/2 = 5\%$ and from 100 to 124 at $\alpha/2 = 1\%$, over the same range of error conditions.

TABLE 8.5
PROPORTION OF SAMPLES
NOT REQUIRING EXTENSION

ERROR		$\alpha/2 = 5\%$			$\alpha/2 = 1\%$	
AMOUNT	CASE	n = 100	n = 50	Case	n = 100	n = 50
1.5M	1	20.8%	1.1%	43	3.5	0.1
1.5M	2	21.7	1.2	44	3.7	0.1
1.5M	3	48.8	3.4	45	19.7	0.6
1.5M	4	83.1	10.5	46	62.1	3.3
1.5M	5	99.2	32.5	47	97.6	18.0
1.5M	6	100.0	100.0	48	99.8	98.2
1.5M	7	39.0	2.9	49	12.2	0.4
1.5M	8	63.3	3.8	50	30.3	0.6
1.5M	9	87.8	3.4	51	63.1	0.4
1.5M	10	100.0-	2.4	52	98.6	0.2
1.5M	11	26.1	0.1	53	5.7	0.1
1.5M	12	34.9	0.7	54	8.4	0.1
1.5M	13	42.6	0.5	55	11.8	0.0+
M	14	56.4	0.4	56	19.3	0.0+
M	15	60.8	5.0	57	30.0	1.0
M	16	62.2	5.2	58	31.2	1.1
M	17	83.1	10.5	59	62.5	3.3
M	18	96.7	22.3	60	90.9	10.3
M	19	99.9	47.2	61	99.8	32.1
M	20	100.0	100.0	62	100.0	100.0
M	21	79.2	10.0	63	54.7	3.0
M	22	91.7	15.1	64	77.9	5.4
M	23	98.9	17.3	65	95.5	6.1
M	24	100.0	16.5	66	100.0	5.2
M	25	76.6	4.4	67	39.4	0.9
M	26	77.2	3.6	68	49.0	0.7
M	27	84.5	3.0	69	59.1	0.6
M	28	92.8	2.5	70	73.3	0.4
0.5M	29	96.7	22.4	71	90.4	10.0
0.5M	30	97.1	23.5	72	91.3	10.5
0.5M	31	99.2	32.5	73	97.8	18.5
0.5M	32	99.9	47.2	74	99.8	32.5
0.5M	33	100.0	86.7	75	100.0	57.0
0.5M	34	100.0	100.0	76	100.0	100.0
0.5M	35	99.1	32.3	77	97.4	18.4
0.5M	36	99.8	44.8	78	99.4	29.9
0.5M	37	100.0-	57.2	79	100.0-	42.7
0.5M	38	100.0	67.5	80	100.0	53.2
0.5M	39	97.9	21.6	81	94.6	9.5
0.5M	40	99.1	19.7	82	97.1	8.6
0.5M	41	99.7	17.9	83	98.8	7.7
0.5M	42	100.0-	16.3	84	99.8	6.9

TABLE 8.6
PROPORTION OF SAMPLES
NOT REQUIRING EXTENSION

ERROR AMOUNT	n = 100			n = 50		
	LOW	MEDIAN	HIGH	LOW	MEDIAN	HIGH
			$\alpha/2 = 5\%$			
1.5M	20.8%	52.6%	100.0%	0.1%	2.7%	100.0%
M	60.8	88.7	100.0	2.5	10.3	100.0
0.5M	96.7	99.4	100.0	16.3	32.4	100.0
			$\alpha/2 = 1\%$			
1.5M	3.5	25.0	99.8	0.0+	0.3	98.2
M	30.0	67.9	100.0	0.4	3.2	100.0
0.5M	90.4	99.1	100.0	0.4	18.5	100.0

TABLE 8.7
AVERAGE SAMPLE SIZE
DOUBLE SAMPLING

ERROR AMOUNT	CASE	$\alpha/2 = 5\%$ n = 100	n = 50	CASE	$\alpha/2 = 1\%$ n = 100	n = 50
1.5M	1	128	158	43	153	189
1.5M	2	127	156	44	152	187
1.5M	3	111	125	45	125	149
1.5M	4	102	95	46	106	112
1.5M	5	100	69	47	100	78
1.5M	6	100	50	48	100	50
1.5M	7	116	137	49	134	163
1.5M	8	107	117	50	118	140
1.5M	9	101	98	51	105	117
1.5M	10	100	79	52	100	94
1.5M	11	122	148	53	145	178
1.5M	12	117	140	54	137	168
1.5M	13	112	133	55	130	159
1.5M	14	108	125	56	123	150
M	15	108	125	57	119	147
M	16	107	123	58	118	145
M	17	102	100	59	106	115
M	18	100	79	60	101	88
M	19	100	61	61	100	65
M	20	100	50	62	100	50
M	21	103	108	63	109	127
M	22	101	93	64	103	109
M	23	100	78	65	100	91
M	24	100	63	66	100	73
M	25	105	117	67	114	138
M	26	103	111	68	110	130
M	27	102	104	69	107	123
M	28	101	99	70	104	117
.5M	29	100	89	71	101	102
.5M	30	100	87	72	101	100
.5M	31	100	75	73	100	82
.5M	32	100	64	74	100	67
.5M	33	100	56	75	100	57
.5M	34	100	50	76	100	50
.5M	35	100	79	77	100	88
.5M	36	100	70	78	100	77
.5M	37	100	61	79	100	66
.5M	38	100	52	80	100	55
.5M	39	100	83	81	100	95
.5M	40	100	79	82	100	90
.5M	41	100	75	83	100	85
.5M	42	100	71	84	100	80

TABLE 8.8
AVERAGE SAMPLE SIZE
DOUBLE SAMPLING

ERROR AMOUNT	n = 100			n = 50		
	MINIMUM	MEDIAN	MAXIMUM	MINIMUM	MEDIAN	MAXIMUM
			$\alpha/2 = 5\%$			
1.5	100	112	128	50	125	158
M	100	103	108	50	100	125
0.5M	100	100	100	50	74	89
			$\alpha/2 = 1\%$			
1.5M	100	124	153	50	155	189
M	100	105	119	50	116	147
0.5M	100	100	101	50	81	102

Performance of the Confidence Limits - Double Sampling

The performance of the confidence limits in double sampling was examined on a one-sided and two-sided basis, following the approach taken for single sampling. The results are given in Table 8.9, by initial sample size, for $\alpha/2 = 5\%$.[2]

The results show that in no case did the upper limit fail to cover the actual error by more than the nominal proportion $\alpha/2$. The lower limit, however, performed better when the initial sample size was twice the discovery sample size. Lower limit coverage of the actual error was less than the nominal 95% proportion in the same four cases and with the same percentages as those reported with respect to single sampling. When a discovery sample was selected as the initial sample size lower limit coverage deteriorated. In 24 cases, the lower limit failed to cover the actual error amount for as much as 7% of the sample outcomes. However, in 18 of these cases the coverage failure rate was between 5% and 6%. These results suggest that the lower limit performs reasonably well in double sampling.

One cause of the degrading of the lower limit was suggested in the analysis of the single sampling results - that t_1 tends to be too small under quadrinomial error conditions for small initial sample sizes.

Another cause is that the point estimate in double sampling is slightly biased. This test was run without bias correction. In order to observe the impact of adjusting the point estimate for bias, by Cox's method,[3] a limited reperformance of the test was undertaken.
The results for n=50, which are given in Table 8.10, show that the bias correction is effective in improving the performance of the lower limit in double sampling without undermining the upper limit. The only cases whose limits exhibited risk levels in excess of the nominal 5% level were three cases whose lower limits also deteriorated in single sampling (numbers 6, 20, and 34) and two cases (numbers 29 and 39) where performance, though improved, was still slightly worse than at the nominal level.

TABLE 8.9
CONFIDENCE INTERVAL
DOUBLE SAMPLING - WITHOUT BIAS ADJUSTMENT
$\alpha/2 = 5\%;\ 1 - \alpha = 90\%$

	n = 100			n = 50		
CASE	E > UEL	E < LEL	E ≥ LEL AND E ≤ UEL	E > UEL	E < LEL	E ≥ LEL AND E ≤ UEL
1	2.1	3.9	94.0	3.0	6.6	90.4
2	2.3	3.7	94.0	3.2	6.0	90.8
3	3.8	3.2	92.9	3.5	5.2	91.3
4	4.4	2.8	92.7	1.5	4.4	94.2
5	3.4	3.6	93.0	0.2	3.5	96.2
6	0.9	5.5	93.6	0.0	6.5	93.5
7	3.6	3.5	92.9	3.2	5.4	91.4
8	3.3	3.3	93.3	2.3	5.2	92.5
9	3.3	3.5	93.1	1.3	5.0	93.6
10	2.9	4.1	93.0	0.7	5.0	94.2
11	3.5	3.7	92.8	3.0	5.8	91.3
12	3.0	3.6	93.3	2.9	5.7	91.4
13	2.9	3.6	93.5	2.8	5.6	91.6
14	3.5	3.2	93.5	2.6	5.5	91.9
15	1.7	3.0	93.3	5.0	5.7	89.3
16	1.9	2.6	95.2	0.5	6.2	93.3
17	4.7	2.6	95.5	0.0+	5.0	95.0
18	3.3	3.0	92.6	0.0+	4.0	96.0
19	1.1	3.5	93.7	0.0+	3.2	96.8
20	0.1	5.7	95.4	0.0	7.0	93.0
21	3.4	2.9	94.1	0.3	5.4	94.3
22	3.4	2.9	93.7	0.1	5.1	94.7
23	3.1	3.9	92.9	0.1	4.9	95.0
24	2.3	5.4	92.3	0.0+	4.8	95.0
25	3.2	3.1	93.7	0.4	5.8	93.7
26	3.3	3.0	93.7	0.4	5.7	93.9
27	2.9	3.0	94.1	0.3	5.5	94.2
28	2.2	3.6	94.1	0.3	4.9	94.9
29	0.0	1.5	98.5	0.0	6.0	94.0
30	0.0	2.9	97.1	0.0	4.9	95.1
31	0.0	3.1	96.9	0.0	4.9	95.1
32	0.0	3.0	97.0	0.0	3.7	96.3
33	0.0	3.3	96.7	0.0	3.2	96.8
34	0.0	6.4	93.6	0.0	7.0	93.0
35	0.0	3.3	96.7	0.0	5.2	94.8
36	0.0	3.4	96.6	0.0	4.5	95.5
37	0.0	3.7	96.3	0.0	3.6	96.4
38	0.0	3.9	96.1	0.0	3.2	96.8
39	0.0	3.6	96.4	0.0	5.8	94.2
40	0.0	3.7	96.3	0.0	5.6	94.4
41	0.0	3.1	96.9	0.0	5.3	94.7
42	0.0	3.2	96.8	0.0	5.3	94.7

Performance of the Adjustment Rule

An adjusting journal entry to the relevant account is the intended outcome of the sampling process if error is detected in the sample. Consequently, it is necessary for the auditor to be able to propose such an adjustment A with reliability at least $(1-\alpha)\%$ in the procedure that gives rise to the adjustment. The adjustment reliability is defined as the proportion of sample outcomes for which $|A-E| \leq M$.

It will be recalled from Chapter II that, in order for the auditor to propose the adjustment with the stated confidence, it was necessary for the length of the estimation interval not to exceed 2M, thus giving at least one interior point that would differ from both limits by no more than M. To achieve this goal, the final sample size in double sampling was calculated to reduce the length of the upper portion of the estimation interval such that $UEL - \hat{E} = M$.

The results of this process, without bias adjustment, are given in Table 8.11. The adjustment reliability was less than the nominal 90% reliability in one case (case 15, n = 50). In this case 89.3% of the adjustments resulted in error not exceeding M.

Table 8.12 gives the results for the cases that were executed with the bias adjustment. In each case the adjustment reliability was at least equal to the reliability level that was achieved for each comparable case without bias adjustment. Furthermore, there was no case in which the adjustment reliability was less than 90%. Thus, even without bias adjustment and with a lower limit that is not as robust as the upper limit, the double sampling procedure is capable of yielding a reliable, actionable result.

TABLE 8.10
CONFIDENCE INTERVAL
DOUBLE SAMPLING - WITH BIAS ADJUSTMENT
$\alpha/2 = 5\%$; $1 - \alpha = 90\%$
$n = 50$

CASE	$E > UEL_1$	$E < LEL_1$	$E \geq LEL_1$ AND $E \leq UEL_1$
1	3.0%	4.0%	92.9%
2	3.4	3.8	92.9
3	3.6	3.6	92.8
4	1.5	2.9	95.6
5	0.2	2.2	97.5
6	0.0+	6.5	93.5
7	3.2	3.9	92.9
8	2.3	3.6	94.1
9	1.3	3.4	95.2
10	0.7	3.4	95.9
11	3.1	4.1	92.9
12	3.0	4.0	93.0
13	2.9	3.9	93.2
14	2.7	3.8	93.4
15	5.0	4.7	90.3
16	0.5	4.5	94.9
17	0.0+	3.8	96.1
18	0.0+	3.0	97.0
19	0.0+	2.4	97.6
20	0.0	7.0	93.0
21	0.3	4.4	95.3
22	0.1	4.0	95.8
23	0.1	3.5	96.3
24	0.0+	3.4	96.4
25	0.4	4.6	95.0
26	0.4	4.5	95.2
27	0.3	4.2	95.5
28	0.3	4.4	95.4
29	0.0	5.4	94.6
30	0.0	4.5	95.5
31	0.0	4.3	95.7
32	0.0	3.1	96.9
33	0.0	2.8	97.2
34	0.0	7.0	93.0
35	0.0	4.4	95.6
36	0.0	3.7	96.3
37	0.0	3.1	96.9
38	0.0	2.5	97.5
39	0.0	5.2	94.8
40	0.0	4.9	95.1
41	0.0	4.6	95.4
42	0.0	4.2	95.8

TABLE 8.11
APPLICATION OF ADJUSTMENT RULE
DOUBLE SAMPLING - WITHOUT BIAS ADJUSTMENT
$\alpha = 10\%$

CASE	$\lvert E-A \rvert \leq M$ $n = 100$	$\lvert E-A \rvert \leq M$ $n = 50$
1	94.0%	90.4
2	94.0	90.8
3	93.9	91.3
4	95.0	94.2
5	96.2	96.2
6	97.5	93.5
7	92.9	91.4
8	93.3	92.5
9	93.1	93.6
10	93.0	94.2
11	92.8	91.2
12	93.3	91.4
13	93.5	91.6
14	93.3	91.9
15	96.7	89.3
16	97.4	93.3
17	97.3	95.0
18	96.9	96.0
19	96.5	96.8
20	99.5	93.0
21	97.1	94.3
22	97.0	94.7
23	96.1	94.9
24	94.6	95.0
25	96.9	93.8
26	97.0	93.9
27	97.0	94.2
28	96.4	94.9
29	98.5	94.0
30	97.1	95.1
31	96.9	95.1
32	97.0	96.3
33	97.4	96.8
34	100.0-	93.0
35	96.7	94.8
36	96.6	95.5
37	96.3	96.4
38	96.1	96.7
39	96.4	94.2
40	96.3	94.4
41	96.9	94.7
42	96.8	94.7

TABLE 8.12
APPLICATION OF ADJUSTMENT RULE
DOUBLE SAMPLING - WITH BIAS ADJUSTMENT
$\alpha = 10\%$

CASE	$\lvert E-A \rvert \leq M$ $n = 50$
1	92.9%
2	92.9
3	92.8
4	95.6
5	97.4
6	93.5
7	92.9
8	94.1
9	95.2
10	95.9
11	92.9
12	93.0
13	93.2
14	93.4
15	90.3
16	94.9
17	96.1
18	97.0
19	97.6
10	93.0
21	95.3
22	95.8
23	96.3
24	96.4
25	95.0
26	95.2
27	95.5
28	95.4
29	94.6
30	95.5
31	95.7
32	96.9
33	97.2
34	93.0
35	95.6
36	96.3
37	96.9
38	97.5
39	94.8
40	95.1
41	95.4
42	95.8

FOOTNOTES CHAPTER VIII

1. Angus, J.E. and R.E. Schafer, "Improved Confidence Statements for the Binomial Parameter," American Statistician, August 1984, 189-191.

2. Performing the double sampling test for /2 = 1% proved to be impracticable on the computer for which the test software was developed, an IBM XT. The computer could not accommodate the temporary files that were needed to store intermediate results.

3. Cox, D.R., op. cit.

CHAPTER IX

CONCLUSIONS AND RECOMMENDATIONS

When audit sampling is used to perform a substantive test, the procedure should be capable of providing the auditor with evidential matter that is sufficient to form the basis for the expression of an opinion as to the extent of error in the population under examination. If the auditor has chosen not to rely on internal accounting controls, it is important to have available an audit sampling procedure that is sufficient in itself to satisfy this objective.[1] In such situations, adjustment procedures are an integral part of the estimation framework, insofar as it is necessary for the auditor to be able to conclude that errror does not exceed a specified maximum tolerable amount.

This study was concerned with the development of a double sampling procedure that would be applicable to an important class of audit problem, substantive testing of the existence and valuation assertions in which errors are relatively rare and are, by definition, overstatements of accounts or transactions. The procedure also involved sampling with probability proportional to size (dollar unit sampling).

The principal conclusion of this study is that estimation by double sampling is an effective technique which provides the auditor with a clear, unambiguous basis for obtaining evidential matter as required by the third standard of field work. Moreover, because the test results showed that discovery sample sizes could be readily applied, estimation by double sampling is also an efficient approach to audit sampling. The test results also support three subordinate findings, which follow.

First, this study showed that the detection/estimation framework is a viable alternative to the significance testing approach that is suggested in SAS 39, and to related decision approaches. It was also shown that a reliable approach exists for adjusting account balances to give effect to detected errors.

Second, this study demonstrated that, by viewing the sampling distribution as the result of a compound Poisson process, a simple, generally reliable parametric approach to interval estimation could be employed.

Finally, this study showed that the estimation objective could be satisfactorily achieved through double sampling without introducing complexity in the calculations.

The findings also suggest areas in which additional research is appropriate. First, the reliability of the lower confidence limits for the estimated error was below the nominal level in the test cases where all of the tainting rates were relatively small, particularly where a discovery sample size was employed. Whether the cause was due to the use of a standardized deviate that did not properly give effect to data scaling, to small sample considerations, or to other causes is a question of mathematical statistics and is beyond the scope of this study. Nevertheless, it is an issue to be addressed in further research. One alternative to be noted, however, is to define $t_1 = (m - r_1(m))/\sqrt{m}$, that is, to use the standardized deviate of the Poisson distribution as the basis for calculating lower precision.[2]

A second, and related area for further investigation, concerns the concept of basic precision. Since the concept was first introduced by Stringer, basic precision has been a component of all of the reliable procedures for computing upper error limits. As stated in Chapter IV, the basic precision factor is known to vary as a function of the number of taintings and the tainting rates. The function appears to be monotonic and decreasing from its maximum value $-\ln(\beta)$. Beyond these properties, nothing is known about the behavior of basic precision. It may be that the Stringer and related methods implicitly calculate upper error limits based on assumed upper bounds on the standard errors. Roberts[3] has suggested such an approach when samples disclose few errors. Such an approach might be adapted to the compound Poisson situation by defining the upper precision to be $z_u s_u$, where

$$s_u = \sqrt{r_u(0)+(r_u(m)-r_u(0))\sum r^2/m} \tag{9.1}$$

and

$$z_u = (r_u(m)-m)/\sqrt{r_u(m)}. \tag{9.2}$$

One advantage of adopting this approach is that it eliminates the need to perform a separate test to ensure that the precision factor is no less than $-\ln(\beta)$.

Clearly, there is room for further research, particularly with respect to employing parametric methods for obtaining interval estimates.

CHAPTER IX FOOTNOTES

1. AICPA, op. cit., AU 320.53.

2. See Mac Guidwin, M.J., D.M. Roberts, and M. Shedd, op. cit. In this paper, the authors used this definition of t, but only for the upper limit calculation.

3. Roberts, D.M., op.cit., 1978, p.77.

BIBLIOGRAPHY

AICPA, Codification of Statements on Auditing Standards. Numbers 1 to 47, New York: AICPA, 1984.

Anderson, R.J., The External Audit, Toronto: Pitman, 1977.

Anderson, R. J. and Teitlebaum, A.D. "Dollar Unit Sampling", CA Magazine (formerly Canadian Chartered Accountant), April 1973.

Angus, J. E., and R. E. Schafer, "Improved Confidence Statements for the Binomial Parameter", American Statistician, August 1984

Arkin, H., Handbook of Sampling for Auditing and Accounting. 2nd Edition. New York: McGraw-Hill, 1974.

Arkin, H., Sampling Methods for the Auditor: An Advanced Treatment, New York: McGraw-Hill, 1981

Bailey, A.D. Jr., Statistical Auditing, New York: Harcourt Brace Jovanovich, 1981.

Cochran, W. G., Sampling Techniques, 3rd Edition, New York: John Wiley & Sons, 1977.

Cox, D. R., "Estimation by Double Sampling", Biometrika, 1952, 39, 3.

Cox, D. R., and E. J. Snell, "On Sampling and the Estimation of Rare Errors", Biometrika, 1979, 66, 1.

Deming, W. E., Sample Design in Business Research. New York: John Wiley & Sons, Inc., 1960.

Dworin, L., and R. A. Grimlund, "Dollar Unit Sampling for Accounts Receivable and Inventory", Accounting Review, April 1984.

Elliot, R. K., and J. Rogers, "Relating Statistical Sampling to Audit Objectives", The Journal of Accountancy, July 1972.

Feller, W., An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd Ed., New York: John Wiley & Sons, 1968.

Feinberg, S., J. Neter, and R. A. Leitch, "Estimating the Total Overstatement Error in Accounting Populations", Journal of the American Statistical Association, June 1977.

Garstka, S. J., and P. A. Ohlson, "Ratio Estimation in Accounting Populations with Probabilities of Sample Selection Proportional to Size of Book Values", Journal of Accounting Research, Spring 1979, 17.

Hansen, M. H., W. N. Hurwitz, and W. G. Madow, Sample Survey Theory and Methods, Volume 1, New York: John Wiley & Sons, 1953.

Holt, C. C., F. Modigliani, J. F. Muth and H. A. Simon, Planning Production, Inventories and Work Force, Englewood Cliffs: Prentice Hall, 1960.

Jenne, S. E., "The Effects of Three Methods of Sample Selection on Dollar Unit Sampling Bounds", Symposium on Auditing Research V, Urbana-Champaign: University of Illinois, 1983.

Johnson, J. R., R. A. Leitch, and J. Neter, "Characteristics of Errors in Accounts Receivable and Inventory Audits", Accounting Review, April 1981.

Kaplan, R. S., "Statistical Sampling in Auditing with Auxiliary Information Estimators", Journal of Accounting Research, Autumn 1973.

Kaplan, R. S., "Sample Size Computation for Dollar Unit Sampling", Journal of Accounting Research: Studies on Statistical Methodology in Auditing, 1975.

Leslie, D. A., and S. J. Aldersley, "Discussants Response to 'The Behavior of Selected Upper Bounds of Monetary Error using PPS Sampling'", Symposium on Auditing Research IV, Urbana-Champaign: University of Illinois, 1982.

Leslie, D. A., A. D. Teitlebaum, and R. J. Anderson, Dollar-unit Sampling: A Practical Guide for Auditors, Toronto: Copp, Clark and Pitman, 1979.

MacGuidwin, M. J., D. M. Roberts, and M. Shedd, "The Behavior of Selected Upper Bounds of Monetary Error Using PPS Sampling", Symposium on Auditing Research IV, Urbana-Champaign: University of Illinois, 1982.

McCray, J. H., "A Quasi-Baysian Audit Risk Model for Dollar Unit Sampling", Accounting Review, January 1984.

Meikle, G. R., Statistical Sampling in an Audit Context, Toronto: Canadian Institute of Chartered Accountants, 1972.

Molina, E. C., "An Interpolation Formula for Poisson's Exponential Binomial Limit", American Mathematical Monthly, V.22, 1915.

Neter, J., and L. L. Vance, Statistical Sampling for Auditors and Accountants, New York: John Wiley & Sons, 1956.

Neter, J. and J. K. Loebbecke, Behavior of Major Statistical Estimators in Sampling Accounting Populations: An Empirical Study, Research Monograph Number 2, New York: AICPA 1975.

Neter, J., R. A. Leitch, and S. E. Fienberg, "Dollar Unit Sampling: Multinomial Bounds for Total Overstatement and Understatement Errors", Accounting Review, January 1978.

Neyman, J., Lectures and Conferences on Mathematical Statistics and Probability, Washington: United States Department of Agriculture, 1952.

Ramage, J. G., A. M. Krieger, and L. L. Spero, "An Emperical Study of Error Characteristics in Audit Populations, "Studies on Auditing, Supplement to Journal of Accounting Research, 1979.

Roberts, D. M. "A Statistical Interpretation of SAP 54", Journal of Accountancy, March 1974.

Roberts D. M. and M. Shedd, "The Behavior of Selected Upper Bounds of Monetary Error Using PPS Sampling", Symposium on Auditing Research IV, Urbana-Champaign: University of Illinois, 1982.

Roberts, D. M. Statistical Auditing, New York: AICPA, 1978. Roberts, D. M. "A Proposed Sequential Sampling Plan for Tests of Compliance,", Symposium on Auditing Research, Champaign-Urbana: University of Illinois, 1976.

Ross, S. M., Applied Probability Models with Optimization Applications, San Francisco: Holden-Day, 1970.

Stein, C., "A Two-Sample Test for a Linear Hypothesis Whose Power is Independent of the Variance", Annals of Mathematical Statistics 1945, 16.

Stephan, F. F., "Some Statistical Problems Involved in Auditing and Inspection", Proceedings of the American Statistical Association, 1963.

Stringer, K., "Practical Aspects of Statistical Sampling in Auditing", Proceedings of the American Statistical Association, 1963.

Teitlebaum, A. D, and C. F. Robinson, "The Real Risks in Audit Sampling", Journal of Accounting Research, Vol. 13, Supplement 1975.

Teitlebaum, A. D., J. H. McCray, and D. A. Leslie, "Approaches to Evaluating Dollar-unit Samples", paper presented at the 1978 Annual Meeting of the American Statistical Association.

Thompson, C. M., "Tables of Percentage Points of the Chi-square Distribution:, Biometrika, 1941, 32, II.

Trueblood, R.M., and R. M. Cyert, Sampling Techniques in Accounting. New Jersey: Prentice-Hall, 1957. Vance, L. L., Scientific Method for Auditing, Berkeley: University of California Press, 1950.