

Division of Research
Graduate School of Business Administration
The University of Michigan

April 1977

A GENERAL STOCHASTIC MODEL
FOR TESTS OF DETAILS

Working Paper No. 146

by

James T. Godfrey and R. W. Andrews

The University of Michigan

FOR DISCUSSION PURPOSES ONLY

None of this material is to be quoted or
reproduced without the express permission
of the Division of Research.

A General Stochastic Model for Tests of Details

(Auditing Area)

ABSTRACT

In this paper we introduce a stochastic model for the relationship between book and audit values. On the basis of this model we develop a method for finding estimators of the total audit value. They are conditional maximum likelihood estimators since (1) they are a function of the maximum likelihood estimators of the parameters of the underlying population model, and (2) they depend on having full information on the population of book values. Two specific examples of the general model are presented and analyzed. Numerical examples are given to illustrate the use of this approach in calculating required sample size.

I. The General Model

Research into the use of statistical sampling in auditing has been increasing as evidenced by the growth of research reports found in the literature. Two recent significant elements of this growth have been the tenth annual conference on research in accounting held at the University of Chicago in 1975, which featured statistical methodology in auditing [8], and the AICPA sponsored study by Neter and Loebbecke, also in 1975 [5]. The Chicago conference focused on different aspects of auditing, e.g., internal control, compliance and substantive testing, and analytical review. Various statistical methodologies that might be used by auditors--such as regression analysis, decision theory, item sampling and dollar unit sampling--were presented. The study by Neter and Loebbecke focused on simulations of the statistical behavior of various estimators that have been used (and proposed for use) in practice.

These recent research efforts have had the purpose of giving greater rigor to auditing procedures. It should be stressed that the use of statistical methodologies by auditors cannot replace the need for their good judgment. Rather, the judgments required become more refined in that it becomes vital to select statistical methodologies that are appropriate to the accomplishment of specific audit objectives. If the proper statistical methodology is selected, it has the potential to help validate an auditor's conclusions at the final stage of an audit.

The auditing area of inquiry in this paper is that part of the

substantive testing referred to as the tests of details.¹ Some authors refer to this area as variables estimation. In any case, we are considering the problem in which an auditor desires to reach conclusions about the acceptability of specific total dollar balances that are presented by a client. Information obtained from sample observations is used by the auditor to aid in his conclusions regarding the dollar balances.

A variety of statistical procedures have been developed for estimating total population values based upon sample observations. In most cases, these statistical procedures have as their source the sample survey statistical literature. Valid use of these procedures depends upon the underlying properties of the populations from which the samples are drawn. In the study by Neter and Loebbecke [5] it was found that many of the most widely used statistical procedures (estimators) did not behave (probabilistically speaking) as they are purported to behave in theory. One partial explanation may be that a specific procedure (estimator) is not appropriate for a specific auditing population. By the word "appropriate" we are referring to both the probabilistic characteristics of the estimator and the special nature of the functional relationship between book and audited values. While we will not be using the Neter and Loebbecke study as a focal point, the issues surrounding the question of "appropriateness" of statistical procedures will be our main point of concern.

¹ W.R. Kinney, Jr., "Decision Theory Aspects of Internal Control System Design/Compliance and Substantive Tests," which is found [8], pp. 14-29. In Fig. 1, p. 16, Kinney diagrams the routes to an auditor's opinion, reflecting SAS No. 1, Section 320. Analytical review and tests of details are included under the heading Substantive Testing.

The primary theme of this paper is that an auditor should be concerned about an underlying population model when selecting a statistical procedure for the ultimate purpose of making judgments about a total dollar balance presented by a client. We propose that each auditing population of interest with its underlying properties can be represented by an analytical model. A basis for our approach is some work by Kaplan-- [3] and [4]--in which he concluded

Two quantities are especially important for inference in auditing: the error rate and, for those items in error, the distribution of errors. The challenge now is to develop statistically valid techniques that can use sample information on these two quantities....²

The two quantities referred to by Kaplan are incorporated into our proposed analytical model. If an auditor can adequately model the population process by which the audited value of an account is either equal to or not equal to the corresponding book value and also model the form and magnitude of the errors, then presumably the quality of the inference procedures can be improved. In the modern computer setting the total population of book values is usually known; therefore, there is no need to assume a model for the book values.

This discussion leads to some reasonable questions in regard to the use of statistical procedures in auditing.

1. What kind of population model is implied by the use of a particular statistical procedure?

2. What kind of a statistical procedure is most appropriate in a particular auditing situation in which the population of interest has been modeled?

²[4], p. 257.

We would suggest that the second question is the more important one to consider. The first question implies the somewhat naïve approach of a solution looking for a problem. That is, employing a specific statistical procedure may be incorrect because the underlying population model implied by the procedure may be inappropriate for the situation under investigation. The second question, which considers what statistical procedure is more appropriate in a particular auditing situation, might be characterized as a head-on approach. In this approach we model a specific auditing situation and then develop an appropriate statistical procedure.

We list below two of the typical statistical estimation procedures that have been employed in auditing. These will be referred to in the next section, and they also provide a vehicle for the notation in the rest of this report.

$$\hat{Y}_T = \frac{\sum_{j=1}^n y_j}{\sum_{j=1}^n x_j} \sum_{i=1}^N X_i \quad (\text{ratio estimator}) \quad (1.1)$$

$$\hat{Y}_T = \frac{N}{n} \sum_{j=1}^n y_j + b \left[\sum_{i=1}^N X_i - \frac{N}{n} \sum_{j=1}^n x_j \right] \quad (\text{regression estimator}) \quad (1.2)$$

where,

\hat{Y}_T = estimate of unknown auditing population sum

N = number of items in the population

n = number of items in the sample

x_j = book value of j^{th} item in sample

y_j = audited value of the j^{th} item in sample

X_i = book value of i^{th} item in population

Y_i = audit value of i^{th} item in population.

In practice the probabilistic behavior of each of these estimators is assumed to be normal. This probabilistic assumption along with an estimate of the sampling error allows auditors to express judgments as to the correctness of account totals presented by clients. Therefore, a statistical procedure is used in a specific audit situation without direct investigation of the characteristics of the populations from which samples are drawn. These two estimators seem implicitly to take into account relationships between book values and audited values. In (1.1) the relationship is multiplicative, and in (1.2) it is both additive and multiplicative. However, the two critical underlying population properties previously described are still suppressed and are not explicit parts of the relationships. That is, the error rate (the frequency with which $Y_i \neq X_i$) is not explicitly modeled but is taken in account in an average way. Also the distribution and the form of the errors are implicitly averaged in each relationship. One might speculate that, if these characteristics could be modeled well, then the derived estimation procedures could result in an improvement. "Improvement" should be judged on the basis of ability to accomplish auditing objectives which, for example, might translate into statistical properties such as unbiasedness and minimum variance.

In the remainder of this paper we propose analytical models of auditing populations and proceed to develop maximum likelihood estimators for the total audited value. A general form for our analytical models of auditing populations can be stated as follows:

$$Y_i = \begin{cases} X_i & \text{with probability } 1-P(X_i) \\ g(X_i; \tilde{T}, \tilde{\phi}) & \text{with probability } P(X_i) \end{cases} \quad i = 1, 2, \dots, N \quad (1.3)$$

where,

\tilde{T} is a vector of random variables,

$\tilde{\phi}$ is a vector of parameters,

and $P(X_i)$ is the probability of an error when the book value is X_i .

We will refer to this model (1.3) as our general model.

II. Two Examples of the General Model

In this section as specific examples of the general model, given in Section I, we select two models which could represent auditing populations. The selection of these models consists of choosing a specific form for the g function and the $P(X_i)$ function of (1.3). In both of the specific models considered in this section we assume that $P(X_i)$ is constant. Since we want to use the procedures of maximum likelihood estimation, we also must incorporate assumptions concerning the distribution of the random variables into the model.

The purpose of using estimation procedures in a test of details is to estimate the total audited value. The total book value is known since all

of the individual book values are known. Our estimation procedure will be to estimate the function that represents the conditional mean of the audit total, given the individual book values. We will find this function using our assumed underlying model. The conditional mean will be a function of the book values and the parameters of our model. By estimating the parameters of our model we will have estimated the total audited value. The procedure that we will use will be the method of estimation known as Maximum Likelihood Estimation.

Since the maximum likelihood estimator of a function comprised of a number of parameters is, in turn, a function of the maximum likelihood estimators of the parameters, we will then have found the maximum likelihood estimator of the total audited value. If we did not have a model for the population of interest, we could not use the maximum likelihood estimation procedures. This procedure, which is used extensively in the statistics literature, has not been suggested previously for this type of an auditing application.

Our estimation procedure is given by the following steps:

1. We will find the conditional expected value of $\sum_{i=1}^N Y_i$, given $\tilde{X} = (X_1, \dots, X_N)$, which will be a function of \tilde{X} and the parameters of the model;

2. We will find the maximum likelihood estimates of the parameters of the model; and

3. We will estimate $\sum_{i=1}^N Y_i$ with \hat{Y}_T by substituting the estimators of the parameters into the function which is the conditional mean of $\sum_{i=1}^N Y_i$.

In developing the maximum likelihood estimators we take a bivariate sample of the size $n < N$ represented thus:

$$(x_1, y_1); (x_2, y_2); \dots; (x_n, y_n) \quad (2.1)$$

A similar procedure with more complicated ramifications could be used in the case of stratified book values; however, this report does not address that problem. It should also be mentioned that the auditing unit we are using is a transaction or an accounting record and not a dollar value as in dollar unit sampling.

We now present two specific examples of our general model and derive maximum likelihood estimators for the total audited value.

II.A. Additive Model with Normal Errors

The form of the model is given by the following:

$$Y_i = \begin{cases} X_i & \text{with probability } 1-\gamma \\ X_i + \varepsilon & \text{with probability } \gamma \end{cases} \quad i = 1, 2, \dots, N \quad (2.2)$$

where ε is normally distributed with mean μ and variance σ^2 . The error term is assumed to be independent of the level of the book value. Considering (2.1) as a random sample, the conditional likelihood function for the parameters γ , μ , and σ^2 is:

$$L(\gamma, \mu, \sigma^2 | \tilde{X}) = \binom{n}{k} \gamma^k (1-\gamma)^{n-k} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^k \exp \left\{ -\frac{1}{2} \sum_{j=1}^k \left(\frac{y_j - x_j - \mu}{\sigma} \right)^2 \right\} \quad (2.3)$$

where k = number of observations for which $(x_j \neq y_j)$. Therefore, k is a binominal random variable with parameters n and γ .

The solution to the log of the likelihood function yields the following estimators:

$$\hat{\gamma} = kn^{-1}; \quad \hat{\mu} = k^{-1} \sum_{j=1}^k (y_j - x_j); \quad \hat{\sigma}^2 = k^{-1} \sum_{j=1}^k (y_j - x_j - \hat{\mu})^2, \quad (2.4)$$

In order to use these estimators of the parameters to estimate the total audited value, we must find the conditional expected value of the total audited value as a function of the parameters. Going back to the model in equation (2.2) and taking the expected value of Y_i given the book values (\tilde{X}), we have:

$$E[Y_i | \tilde{X}] = (1-\gamma)X_i + \gamma(X_i + \mu) = X_i + \gamma\mu. \quad (2.5)$$

Therefore, the expected value of the unknown population sum of audit values is

$$E \left[\sum_{i=1}^N Y_i \mid \tilde{X} \right] = \sum_{i=1}^N X_i + N\gamma\mu, \quad (2.6)$$

Substituting the maximum likelihood estimators of γ and μ into equation (2.6) we derive the maximum likelihood estimator (MLE) of the total audited value:

$$\hat{Y}_T = \sum_{i=1}^N X_i + \frac{N}{n} \sum_{j=1}^k (y_j - x_j), \quad (2.7)$$

We recognize equation (2.7) as a specific form of the regression estimator with $b=1$. In this form the regression estimator is usually called the difference estimator. Therefore, we have shown that under the model given by equation (2.2) the maximum likelihood estimator of the total audited value is the difference estimator.

By taking the expected value of \hat{Y}_T and showing it is equal to the right-hand side of equation (2.6) we will have demonstrated that \hat{Y}_T is an unbiased estimator for the underlying model in (2.2). This is done as follows:

$$\begin{aligned}
 E[\hat{Y}_T | \tilde{X}] &= \sum_{i=1}^N X_i + \frac{N}{n} E \left\{ E \left[\sum_{j=1}^k (y_j - x_j) \mid \tilde{X}, k \right] \right\} \\
 &= \sum_{i=1}^N X_i + \frac{N}{n} E \{ k\mu \mid \tilde{X} \} \tag{2.8} \\
 &= \sum_{i=1}^N X_i + N\gamma\mu
 \end{aligned}$$

Notice that the first equality sign uses the result that $E\{E[Z_1 | Z_2]\} = E\{Z_1\}$ and the last step uses the fact that k is a binominal random variable with parameters n and γ . The unbiasedness of this estimator should not be taken lightly, because if the true model of the population is (2.2) other alternative estimators may not be unbiased. That is, the expected value of these alternative estimators may not be the right-hand side of equation (2.6).

At this point we are interested in the conditional variance of \hat{Y}_T given the book values. This variance will give us a measure of the accuracy of the estimator and will also enable us to assign the sample size required to achieve a prescribed precision.

The following equations derive³ the conditional variance of \hat{Y}_T :

$$\begin{aligned} \text{Var} [\hat{Y}_T | \tilde{X}] &= \text{Var} \left[\frac{N}{n} \sum_{j=1}^k (y_j - x_j) | \tilde{X} \right] \\ &= \frac{N^2}{n^2} \text{Var} \left\{ E \left[\sum_{j=1}^k (y_j - x_j) | \tilde{X}, k \right] \right\} \\ &\quad + \frac{N^2}{n^2} E \left\{ \text{Var} \left[\sum_{j=1}^k (y_j - x_j) | \tilde{X}, k \right] \right\} \\ &= \frac{N^2}{n^2} \{ \text{Var} [k\mu] + E[k\sigma^2] \} \\ &= \frac{N^2}{n} \{ \mu^2\gamma(1-\gamma) + \sigma^2\gamma \} \end{aligned} \tag{2.9}$$

³ This derivation uses the general result that

$$\text{Var} [Z_1] = \text{Var} \{ E[Z_1 | Z_2] \} + E \{ \text{Var}[Z_1 | Z_2] \} .$$

II.B. Multiplicative Model with Exponential Ratios

The form of this model is given by:

$$Y_i = \begin{cases} X_i & \text{with probability } 1-\gamma \\ \theta X_i & \text{with probability } \gamma \end{cases} \quad i = 1, 2, \dots, N \quad (2.10)$$

where θ has an exponential distribution; i.e., the density of θ is:

$$f(\theta; \lambda) = \begin{cases} \lambda e^{-\lambda\theta} & \theta \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (2.11)$$

We also assume that θ is independent of the level of the book value. This will be recognized as similar to the model that Kaplan [3] used to derive a better estimate of the variance of the ratio and regression estimators.

If we consider (2.1) as a random sample from this model, the conditional likelihood function for the parameters γ and λ is:

$$L(\gamma, \lambda | \tilde{X}) = \binom{n}{k} \gamma^k (1-\gamma)^{n-k} \lambda^k e^{-\lambda \sum_{j=1}^k \frac{y_j}{x_j}} \quad (2.12)$$

where k = number of observations for which $x_j \neq y_j$.

By maximizing the log of the likelihood function we obtain the maximum likelihood estimators for γ and λ to be:

$$\hat{\gamma} = \frac{k}{n} \quad ; \quad \hat{\lambda} = \frac{k}{\sum_{j=1}^k \left(\frac{y_j}{x_j}\right)} \quad (2.13)$$

Since we want to use these maximum likelihood estimators in the conditional expected value of the total audited value, we must find this conditional expected value as a function of the parameters. We proceed under the assumption of the model in (2.10), stating that

$$E [Y_i | \tilde{X}] = (1-\gamma)X_i + \frac{\gamma}{\lambda} X_i \quad , \quad (2.14)$$

since $E[\theta] = \lambda^{-1}$. Therefore,

$$E \left[\sum_{i=1}^N Y_i \mid \tilde{X} \right] = \sum_{i=1}^N X_i [1-\gamma + \gamma\lambda^{-1}] \quad (2.15)$$

Substituting the maximum likelihood estimators for γ and λ into equation (2.15), we derive the MLE of the total audited value to be

$$\hat{Y}_T = \sum_{i=1}^N X_i \left(1 - \frac{k}{n} + \frac{1}{n} \sum_{j=1}^k \frac{y_j}{x_j} \right) \quad (2.16)$$

In this case the estimator is not one of the usual estimators subscribed to in the auditing literature; however, it is in the form of a ratio estimator since it is the product of the total book value with a quantity

depending on the sample. To investigate the accuracy of this estimator we will find its mean and variance.

$$E[\hat{Y}_T | \tilde{X}] = E\{E[\hat{Y}_T | \tilde{X}, k]\} = \sum_{i=1}^N X_i (1-\gamma + \gamma\lambda^{-1}). \quad (2.17)$$

Therefore, \hat{Y}_T is an unbiased estimator of the total audited value.

Now we will find the $\text{Var}[\hat{Y}_T | \tilde{X}]$.

$$\text{Var}[\hat{Y}_T | \tilde{X}] = \text{Var}\{E[\hat{Y}_T | \tilde{X}, k]\} + E\{\text{Var}[\hat{Y}_T | \tilde{X}, k]\} \quad (2.18)$$

$$= \frac{\left(\sum_{i=1}^N X_i\right)^2}{n\lambda^2} [(1-\lambda)^2\gamma(1-\gamma) + \gamma]$$

We will use this variance of the estimator as well as the variance of the estimator given by (2.9) to calculate sample sizes in the next section. It should be noted that for both these specific models different distributions could have been assumed for the error term ϵ and the ratio θ . For example, if an auditing situation presented itself in which the auditor wanted to guard against overstatements only, a possible model would be (2.10) with the assumption that θ has a uniform distribution over the interval $[a, 1)$. Then an estimator could be derived using the maximum likelihood procedure similar to the above analysis.

In many cases analytical work in this area will not be tractable and therefore simulation studies will be required. The two models we selected

were selected not so much for their ability to model audit populations realistically but rather to demonstrate the procedure. Therefore, these models were simple and tractable for deriving the maximum likelihood estimators and the variance of the maximum likelihood estimators. We have investigated many different specific examples of the general model and we usually find that determining the variance of a reasonable estimator is not tractable.

III. Calculation of Sample Size for Derived Estimators

In this section we will use the results of Section II to calculate the required sample sizes for specific numerical examples. These examples will illustrate an important side benefit of using the estimation procedure outlined above. In order to calculate the required sample size using our procedure, population parameters contained in our model must be initially assessed.

It is our contention that the assessment of these parameters is easier than the assessment of the parameters that is needed when estimator (1.1) and (1.2) are used. For example, in order to calculate the sample size required when using the regression estimator (1.2), the auditor must assess prior estimates of the population variance of the audit values and the population correlation between the book and audit values. However, for our multiplicative model with exponential ratios we only need to assess the percentage of the population in error and the percentage of errors which are overstated. This procedure will be illustrated in our second numerical example below.

III.A. Additive Model with Normal Errors

Let us consider a numerical example with which we will calculate the required sample size using the estimator (2.7). A small wholesaler of drug sundries has 2000 different inventory items. From a computer listing taken from the perpetual records we have the 2000 book values available. The sum of these 2000 book values is \$700,000. We have decided that a materiality (M) for this inventory should be \$15,000 with a reliability level of 95 percent and a maximum β error of 20 percent. From these requirements we calculate a planned precision, A, of \$10,500.⁴ With the 95 percent reliability and the desired precision of \$10,500, we only need to solve the equation $10500 = 1.96 \times$ (standard deviation of the estimator) for n to find the required sample size. The 1.96 is the normal distribution value for a reliability of 95 percent. This implies that the estimator has a normal distribution, which is a valid statement since we know under very general conditions that maximum likelihood estimators are asymptotically normal [6] and [7].

⁴Following the discussion and notation in Hermanson et al, [2], Chap. 7, we define

- a) Alpha (α) to be the probability that the statistical evidence fails to support fair statement even though the book value is fairly stated.
- b) Beta (β) to be the probability that the statistical evidence supports fair statement when the book value is materially misstated.
- c) Reliability (R) is $1-\alpha$.

For a given materiality (M), α , and β , the planned precision, A, can be calculated using the formula on [2], p. 225.

By equating

$$10500 = 1.96 [\text{Var} (\hat{Y}_T | \tilde{X})]^{1/2} \quad (3.1)$$

where $\text{Var}(\hat{Y}_T | \tilde{X})$ is given by (2.9), we have

$$10500 = 1.96 [N^2 n^{-1} (\sigma^2 \gamma + \mu^2 \gamma (1-\gamma))]^{1/2}. \quad (3.2)$$

We know $N = 2000$, and that leaves the problem of assessing estimates for γ , μ , and σ^2 .

Let us say that, on the basis of our evaluation of internal control and other audit procedures short of the test of details, we assess γ to be .15, i.e., we expect 15 percent of the perpetual records to be in error. In assessing μ and σ^2 we must ask the question: If a difference occurs between the book and audit value, on the average how large will that error be and what will be the variation of these errors? We must remember that μ could be positive or negative. Therefore, if we don't have any information as to whether an overstatement or understatement is more likely, proper assessment of μ might be zero. For this audit situation we have reason to believe that there are more overstatements (with larger magnitude) so we know we will assess μ to be negative.⁵

We further reason that of those records which are in error, the average error will be approximately 35 and therefore set $\mu = -35$ in (3.2). By similar reasoning we assess the standard deviation of the errors to be \$35 and set $\sigma = 35$. To achieve this assessment we might have assessed the range of possible errors and divided by four.

⁵In this calculation of sample size, the sign on μ plays no role, since μ enters (3.2) only as μ^2 .

It should be stressed that these assessments of γ , μ , and σ^2 are subjective evaluations based on the limited information we have. They will affect our initial sample size but not the reliability or β of our estimate. The worst sequence of events that could occur is that our sample size turns out to be too small to make a valid statistical conclusion and more sample data would have to be obtained. Sometimes this eventuality is guarded against by increasing the calculated n by an arbitrary percentage, such as 10 percent.

We now have the quantities necessary to solve equation (3.2) for n , yielding $n = 47.4$. Therefore, we would require a sample size of 48.

III.B. Multiplicative Model with Exponential Ratios

In order to demonstrate the use of the ratio type estimator given by (2.16) we will use a different approach. Rather than find only the required sample size for a single specified materiality level, we will develop a table which shows the required sample size corresponding to a varying materiality level. This table will be applicable for given reliability and β error levels and for specified population parameters.

As in the previous example assume we are attempting to find the required sample size for a test of details at a reliability level of 95 percent and a maximum planned β of 20 percent. Therefore, using the results in Chapter 7 of [2], we know that planned precision should be seven-tenths of materiality. We have determined that the population model given in (2.10) is appropriate for this auditing situation. Our next step is to assess λ .

In order to do this we would ask the following question: Of the errors in the population, what proportion are overstatements? If the response to this question is μ ($0 < \mu < 1$), then by solving

$$\mu = \int_0^1 \lambda e^{-\lambda\theta} d\theta \quad (3.3)$$

we get an assessed value of λ . The solution of (3.3) gives $\lambda = -\ln(1-\mu)$. For our specific example let's say the response to the above question was .80, then λ would be assessed to be 1.609. We further assess the proportion of the population in error to be .20 and therefore, set $\gamma = .20$ in our sample size calculation. Since the estimator derived from the model in (2.10) was the maximum likelihood estimator, we are justified in assuming normality and equating

$$A = 1.96 [\text{Var} (\hat{Y}_T | \tilde{X})]^{1/2} \quad (3.4)$$

Let p be the proportion of the total book value that we have determined to be material, then

$$M = p \sum_{i=1}^N X_i \quad (3.5)$$

and therefore,

$$.7p \sum_{i=1}^N X_i = 1.96 [\text{Var}(\hat{Y}_T | \tilde{X})]^{1/2}, \quad (3.6)$$

or

$$.7p \sum_{i=1}^N X_i = 1.96 \lambda^{-1} n^{-1/2} \sum_{i=1}^N X_i [(1-\lambda)^2 \gamma(1-\gamma) + \gamma]^{1/2} \quad (3.7)$$

or

$$.7p = 1.96 \lambda^{-1} n^{-1/2} [(1-\lambda)^2 \gamma(1-\gamma) + \gamma]^{1/2}. \quad (3.8)$$

Upon viewing the last equation we see that for fixed values of reliability,

β , λ and γ , the required sample size depends only on the proportion of $\sum_{i=1}^N X_i$ which is determined to be material. Setting $\gamma = .20$ and $\lambda = 1.609$

we can solve (3.8) to give the results set forth in Table 1.

Table 1

Required Sample Size (n) for Materiality Proportion (p)	
(Reliability = 95%, $\beta \leq 20\%$, $\gamma = .20$, $\lambda = 1.609$)	
<u>p</u>	<u>n</u>
.05	314
.06	219
.07	160
.08	122
.09	97
.10	79
.11	65
.12	54
.13	47
.14	41
.15	35

It should be noted that the required sample size does not depend on the number of total items (N). This is due to the model and the assignment of materiality as a percentage of the total book value. It may be that the auditor's required precision based on a small materiality is so stringent that the required sample size is prohibitive and in fact, it could be calculated to be larger than N. This indicates that this model and the derived estimator, as with most ratio type estimators, would only be suitable for populations with a large number of items. An interesting sidelight of this numerical example is that the required sample size is seemingly a very volatile function of p , the proportion of the total book value that is material.

IV. Concluding Comments and Further Research

We have presented a general model for the population of items in the test of details part of a substantive test. The essence of this model, which is really a generalization of Kaplan's work, [3] and [4], is that it dichotomizes the population into those items in error and those items not in error. It also allows the relationship between the audit and book values to be modeled in a way appropriate for the situation at hand. We have also outlined a general procedure for deriving estimators of the total audited values based on the model we select. The two specific models we analyzed were only for the purpose of demonstrating the procedure and should not be construed to be appropriate for all auditing populations.

We see two immediate areas for further research, which we have classified as the identification stage and the empirical comparison study.

Work needs to be accomplished in which actual audit populations are fitted to the general model. Different functions for $g(X_1; \tilde{T}, \tilde{\phi})$ and $P(X_1)$ need to be investigated in order to see if they model with enough accuracy. This work will have to be an empirical study based on actual audit populations, such as the four actual populations compiled by Neter and Loebbecke [5].

Secondly, a simulation approach would be appropriate in order to study the accuracy of the derived maximum likelihood estimators. Comparisons could also be made with other estimation procedures, such as the usual ratio and regression estimators. This would have to be a simulation study because the determination of the exact variance of most estimators derived through our modeling approach is not tractable. It is anticipated that the maximum likelihood estimators will be better than the competitive estimators when the observations are simulated from the assumed model because of the desirable properties that maximum likelihood estimators usually have.⁶ However, other procedures of finding new estimators, such as Bayes, minimax, and least squares, should not be ignored.

⁶For a survey of the properties of maximum likelihood estimators, see Norden [6] and [7].

References

1. Cochran, W. Sampling Techniques. Second Edition. New York: John Wiley and Sons, Inc., 1963.
2. Hermanson, R.H., S.E. Loeb, J.M. Saada, and R.H. Strawser. Auditing Theory and Practice. Homewood, Ill.: Richard D. Irwin, Inc., 1976.
3. Kaplan, R.S. "A Stochastic Model for Auditing." Journal of Accounting Research 11, Spring 1973.
4. _____, "Statistical Sampling in Auditing with Auxiliary Information Estimators." Journal of Accounting Research 11, Autumn 1973.
5. Neter, J., and J.K. Loebbecke. Behavior of Major Statistical Estimators in Sampling Accounting Populations, An Empirical Study. New York, N.Y.: AICPA, 1975.
6. Norden, R.H. "A Survey of Maximum Likelihood Estimation." International Statistical Review 41, No. 1, 1973.
7. _____, "A Survey of Maximum Likelihood Estimation, Part 2." International Statistical Review 41, No. 1, 1973.
8. "Studies on Statistical Methodology in Auditing." Journal of Accounting Research, Supplement, 1975.

