Sampling Size and Auditors' Judgements: A Simulation

MOHAMAD ALI ABDUL-HAMID, SHAMSER MOHAMED and ANNUAR MD-NASSIR

Department of Accounting and Finance
Faculty of Economics and Management
Universiti Pertanian Malaysia
43400 UPM Serdang, Selangor, Malaysia

ABSTRAK

Auditor gunapakai teknik-teknik kos efektif dan efisyen untuk mendapatkan bukti-bukti yang membantu mereka didalam memberi pendapat keatas penyata kewangan. Salah satu teknik audit yang lazim digunakan ialah persampelan dan di UK auditor menggunakan saiz sampel sekecil 25 item. Kajian ini menggunakan teknik simulasi Monte Carlo untuk menentukan samada pendapat auditor berdasarkan pelbagai saiz sampel dan aras ralat adalah dibawah paras ketepatan bolehterima. Keputusan kajian mendapati firma-firma yang menggunakan saiz sampel yang kurang daripada 25 item tidak cukup besar untuk memberi pelan persampelan yang berjaya kecuali pada tahap nilai ralat yang rendah. Untuk memperbaiki pelan persampelan dan kualiti audit, adalah dicadangkan saiz sampel minimum hendaklah melebihi 50 item.

ABSTRACT

Auditors usually seek cost-effective and efficient techniques to accumulate evidence in an effort to express their opinions on financial statements. One such technique is audit sampling, and in the United Kingdom auditors use sample sizes as small as 25 items. This study uses the Monte Carlo simulation technique to determine whether an auditor's opinion using both different sample size and error levels is within an acceptable degree of accuracy. The results suggest that samples of fewer than 50 items are not large enough to provide a successful sampling plan unless the error value is very low. To improve the sampling plan and the quality of the audit, the sample size should, therefore, be increased to more than 50 items.

INTRODUCTION

The high cost of audit sampling in recent years has forced auditors to reduce the size of audit samples. To be cost-effective, audit samples have been reduced significantly, as reported in the literature (Mohamad-Ali 1993), where a sample of 25 items was used to test accounting populations of several thousand items. However, a small audit sample is subject to the possibility of a lack of credibility and accuracy, in terms of giving a true and fair view of the accounts being audited. This study tests whether small samples do provide the auditor with the degree of assurance he needs to state the accounts under audit give a "true and fair" view of the financial condition of the company.

An auditor faces the challenge of two conflicting objectives in gathering evidential matter. First, the collection of excessive evidence at the expense of the client may lead him to seek the services of a more cost-efficient auditor. Second, the auditor is subject to litigation if the client perceives that the auditor had the means, but did not give the most reasonable opinion. Therefore, an auditor needs to maintain a balance between controlling the cost of evidence gathering and the possible consequences of expressing an opinion based on inadequate data.

One way of determining an optimal size of audit sample is to use a well-tested statistical formula. In a recent survey

(Mohamad-Ali 1993) it was found that the use of statistical sampling is on the increase, with 43% using statistical sampling techniques at some stage of their audit procedures and the majority of medium-sized accounting firms stating that they drew a minimum sample size of 25 items from an account under audit. On average, most firms stipulated a sample size of 20-40 units per account audited. Another study (McRae 1982) noted that statistical sample sizes in the UK appear to be significantly smaller than those in North America, with most firms in the UK imposing a minimum sample size of 25 units and a maximum of 100 units.

Although statistical sampling has been in use as an effective audit tool for more than forty years, there is little published evidence on the issue of sample size. The lack of research on this important practical problem is possibly due to the cost of carrying out a proper test on a large population of data. To test the accuracy of the sample on an actual population of accounts is time consuming and costly as every item in the population must be checked for error.

One possible solution to this problem is to develop a computer program which can generate a series of book and audited values (any differences being an error), thus simulating the audit of a real accounting population. This study attempts to determine whether an auditor's opinion on the sampled population is likely to be within an acceptable degree of accuracy when the auditor uses varying audit sample sizes.

RESEARCH DESIGN

This study utilized the Monte Carlo simulation technique to examine problems with a stochastic or probabilistic basis (Hammersely and Handscomb 1964). Principally, a computer program is used to generate a series of book values and error values. These error values are seeded into the book values to become the accounting population, which is later used to generate a series of matching audited values. The book values and the error values are taken from a series of actual book and error values noted by auditors. The

TABLE 1

Frequency distribution and major characteristics of book values

Class	Book Amount ($)	Number of Accounts
1	$0 < x \leq 90$	1,070
2	$90 < x \leq 230$	715
3	$230 < x \leq 400$	450
4	$400 < x \leq 650$	337
5	$650 < x \leq 1,500$	455
6	$1,500 < x \leq 3,500$	409
7	$3,500 < x \leq 5,000$	149
8	$5,000 < x \leq 10,000$	238
9	$10,000 < x \leq 25,000$	210
Total		4,033

Source: Neter and Loebbecke (1975) p. 26

distribution of the generated book and error values are shown in Table 1 and 2 respectively.

POPULATIONS USED IN THIS STUDY

In order to generate a series of book and audited values several elements in the simulated accounting population need to be specified and explained.

First, to generate the distribution pattern of values (the skewness) in this study, the actual elements found in audited accounting populations were sampled (taken from Population 4 of Neter and Loebbecke's (1975) study[1] of accounting population parameters). Population 4 consists of 4033 trade debtors' accounts and contains only one-sided errors owed to a large US manufacturer. Table 1 illustrates a frequency distribution of these book values representing the trade debtors' accounts. It shows that the distribution is skewed to the right, implying smaller number of items of high value in the population suggests larger number of errors are expected in small value items.

---

[1] Neter and Loebbecke's (1975) study consists of Populations 1, 2, 3 and 4. The Neter and Loebbecke populations are well known in the audit sampling literature and have been widely used by other researchers for comparing the performance of alternative sampling techniques (for example, see Frost and Tamura (1982)).

TABLE 2

Tainting percentages: a classification by relative size of the item in error

| Tainting | Audited Items | | |
	Exceeding $10,000	$2,000-$10,00	Less than $2,000
0 - 1%	35%	19.0%	3%
> 1 - 10%	33%	25.0%	17%
> 10 - 20%	5%	12.0%	19%
> 20 - 99%	17%	19.0%	21%
100%	10%	23.5%	37%
> 100%	0%	1.5%	3%
	100%	100.0%	100%

Source: McRae (1982)

Second, there are errors of principle and operational errors (Taylor 1974). Operational errors can be classified further, into procedural errors and errors of value. This study is concerned with measuring accidental errors of value, which are also referred to as substantive errors; most are monetary errors (McRae 1982). We have ignored deliberate or fraudulent errors in our simulation because the pattern and incidence of such errors are likely to be very different from those of accidental errors and therefore require a separate research study.

The error rate is defined as the proportion of errors in a population. Thus an error rate of 20% means that out of a total population of 100 items, 20 items are in error. The error rate in most accounting populations is very low; however, the acceptable level of error varies from sample to sample. For example, Jones (1947) suggests that error rates below 0.3% are "acceptable" and below 0.9% are considered to be "fair" in clerical work. Vance (1950) used 0.5% as an acceptable rate and 3% as an unacceptable error rate in clerical work. The National Audit Office in the UK applies an unacceptable upper error rate of 2.5% to their audit work on government accounts.

In this study we use three error rates, 1, 2.5 and 5% and define these errors as low, medium and high, and seed them into the population via our simulation program.[2]

Third, the value of the errors and the pattern of distribution of the errors are summarized in Table 2. The term "tainting" used in audit sampling describes the ratio between the value of an error and the value of the item in error. For example, an item of $60 containing a $15 error is said to be 25% "tainted". In actual practice the probability of finding a given tainted percentage appears to be influenced by the relative size of the items in error (McRae 1982). This study classifies the tainting percentage into three groups following McRae's study, that is, audited items exceeding $10,000, those less than $2,000 and those between $2,000 and $10,000.

THE SIMULATION

The simulation program consists of two interrelated BASIC programs. The first program generates 4033 random numbers and stores them on data files. The numbers between 0 and n are generated by using the formula INT[THETA*LOG(RND)], where INT and RND are BASIC functions standing for integer and random number respectively. The second program uses the data inserted into the data file by the first program. Table 3 describes the simulation in detail.

---

[2] According to Neter and Loebbecke's study the number of items in the population is 4033. Thus to make the process simpler, the error items are set to be 50, 100 and 200 errors respectively, that is, for example 2.5% of 4033 is 100 (rounded to ten). This approach creates three populations to be tested.

TABLE 3

The simulation process

STEPS	DESCRIPTION
1. Generate the file holding the population	The intention is to generates 100 files with each file containing 4033 items.
2. Population generation	Within this step we generate a population of 4033 values. The values generated correspond to the book and audited values of each item. The program creates a set of audited values by seeding error values into population of book values.
3. Sampling selection	A sample in now extracted from each population using the monetary unit sampling (MUS) procedure as described below: a) Create a cumulative book value for each population of accounts. b) Randomly select a number = y between 1 and the sampling interval within the cumulated value. We shall call this sampling interval (SI). c) Select the account whose cumulative book values index is just > y + SI = X. d) Repeat C, by $X + SI = X_2$
4. Estimate audit value	Estimate the total audited value of all 4033 accounts based on the samples of 25, 50 and 100 items sampled using the MUS procedure.
5. Decision taken	Decide whether the total audited book value is to be accepted or rejected based on level of tolerable errors.
6. Repeat	Repeat Steps 2, 3, 4 and 5 for 100 runs. This step will measure the probability that the confidence levels claimed by the auditor using this procedure are reasonably accurate.

AUDIT SAMPLING PROCEDURE

Monetary unit sampling (MUS) is a commonly used statistical procedure for expressing an opinion on the validity of the accounts audited from evidence collected from a sample. Mohamad-Ali's (1993) and McRae's (1982) surveys suggest that over 90% of applications of statistical sampling use some form of MUS. The MUS procedure used in this study is a simplified version of the DUS (dollar unit sampling) method described in Leslie *et al.* (1980). This procedure is outlined in Table 4.

This method divides the total population value into equal dollar segments. A dollar unit, sometimes called the "hit" dollar, is then systematically selected from each seg-ment. Thus a sampling interval is calculated as follows: $SI = BV/n$, where BV is the book value and n is the sample size. In our case let us say $BV = \$600,000$ and $n = 88$, then the sampling interval is $6,818 ($600,000/88). The initial step in the sampling selection process is to pick a random number between 1 and 6,818. The auditor then selects the value item that contains every 6,818th dollar thereafter in the population. Assuming a 5,000 random number start, the four sample items selected are as shown in Table 4. It should be noted that though we are sampling individual monetary units in single dollars, the results concern the entire value associated with the "hit" dollar.

TABLE 4

Systematic selection procedure in MUS sampling

Logical unit	Book values	Cumulative values	Numbers selected	Items selected for audit
1	1,200	1,200		
2	6,043	7,243	5,000	6,043
3	2,190	9,433		
4	3,275	12,708	11,818	3,275
5	980	13,688		
6	1,647	15,335		
7	4,260	19,595	18,636	4,260
8	480	20,075		
9	7,150	27,225	25,454	7,150
	600,000	600,000		

EVALUATING THE RESULTS OF THE MUS SAMPLE

The next stage is to evaluate the results of the sampling procedure. Here the auditor considers (1) the projected error value determined by the sample, (2) the degree of error allowed for sampling risk, and (3) the upper error limit determined by the sample. Item (3) is calculated from items (1) and (2). The evaluation process now differs depending on whether any errors are found in the sample.

Sample Selection with No Errors Found

The error results found in the sample are used to estimate the error in the total population. When no errors are discovered in the sample the allowance for sampling risk will equal the upper error limit, which is equal to or less than the level of tolerable error specified in designing the sample. Therefore the auditor can ordinarily conclude, without making additional calculations, that the book value of the population is not overstated by more than the level of tolerable error at the specified risk of incorrect acceptance.

When no errors are found in the sample, the sampling risk factor consists of basic precision (BP). The amount is obtained by multiplying the reliability factor (RF) for zero errors at the specified risk of incorrect acceptance by the sampling interval (SI). In the case under discussion, let us say *that the required level of confidence is 95%, thus RF = 3.0* (derived from the Poisson distribution), then the basic precision is $20,454 (computed as: $BP = RF \times SI = 3.0 \times \$6,818 = \$20,454$). Since the projected error is zero, this amount is also equal to the upper error limit, which is less than the $30,000 tolerable error specified in the sample design. Thus, the auditor may now state that the book value for the population is not overstated by more than $20,454 at the 5% risk of incorrect acceptance.

Sample Selection with Some Errors Found

If some errors are found in the sample, the auditor must calculate both the projected error value in the population and the allowance for sampling risk in order to determine the upper error limit for overstatement errors. The upper error limit is then compared with the tolerable error.

Projected Population Error

A projected error amount for the population is estimated by first calculating the error for each sampled unit containing an error and then adding these errors for the entire population. The projected error is calculated as follows:

TABLE 5

Determination of projected error

Book Value (BV)	Book Value (AV)	Tainting Percentage (TP = (BV-AV)/BV	Projected Error (TP × SI)
950	855	10	682
2,500	0	100	6,818
5,300	5,035	5	341
8,750	5,890		7,841

Tainting percentage = (book value - audit value)/book value

Projected error = tainting percentage × sampling interval

To illustrate, let's assume that the debtors' accounts reveal the following errors as in Table 5. The total error in the sample is $2,860 ($8,750 - $5,890) and the total projected error in the population is $7,841.

Allowance for Sampling Risk

The allowance for sampling risk of samples containing errors has two components: (1) basic precision, and (2) an incremental allowance resulting from the errors. The calculation of basic precision (RF × SI) is the same as explained previously for a sample with no errors. Thus, in the case studied the amount of this component is again $20,454.

The calculation of the incremental allowance involves the following steps:

- Determine the appropriate incremental change in the reliability factor.
- Rank the projected errors from the highest to lowest.

- Multiply the ranked projected errors by the appropriate factor and sum the products.

Table 6 illustrates the first step.

The data in the first two columns are the specified risk of incorrect acceptance (5% in this illustration). Each entry in the third column is the incremental reliability factor. The values in the last column are obtained by subtracting one from each value in the third column. The second and third steps are illustrated in Table 7, which has the projected errors in the first column (taken from Table 5) and incremental reliability factors in the second column (taken from Table 6).

The incremental allowance for sampling risk is the product of columns one and two, and the incremental allowances for the projected errors are then summed to determine the total incremental allowance, which is $5,580 in this example. The total allowance for sampling risk is the sum of basic precision and incremental allowance for projected errors. For example, in the case under study, the total allowance is computed to be $26,034, which is estimated as follows:

TABLE 6

Incremental change in reliability factor minus one 5% risk of incorrect acceptance

Number of Overstatement Error	Reliability Factor (RF)	Incremental Change in RF	Incremental Change in RF Minus One
0	3.00	–	–
1	4.74	1.74	.74
2	6.30	1.56	.56
3	7.75	1.45	.45
4	9.15	1.40	.40

TABLE 7
Incremental allowance for sampling risk

Ranked Projected Errors	Incremental Change in Reliability Factor Minus One	Incremental Allowance for Sampling Risk
$ 6,818	.74	$ 5,045
682	.56	382
341	.45	153
		$ 5,580

Basic precision $20,454
Incremental allowance for
  projected errors 5,580
Total allowance for sampling
  risk $26,034

Upper error limit for overstatement errors. The upper error limit equals the sum of the projected errors plus the allowance for sampling risk, that is, $33,875 ($7841 + $26,034). Thus, the auditor may conclude that there is a 5% risk that the book value is overstated by $33,875 or more.

The figure thus calculated is then compared with the tolerable error for the item under consideration. If the upper error limit is less than the tolerable error the auditor can accept the population. If the opposite is true, the auditor may adjust the upper error limit for any error found (assuming that the client agrees to the adjustment) to determine whether that reduces the upper error limit to below the tolerable error. If the upper error limit remains above the tolerable error the auditor should carry out such procedures as are laid down by the audit firm to deal with such a situation.

Generally, if the upper error limit is less than the tolerable error, the sample results support the conclusion that the population book value is not mis-stated by more than the tolerable error at the specified risk of incorrect acceptance. In the case under review, the upper error limit exceeds the tolerable error of $30,000 specified in designing the sample. Thus, in this case, the population should be rejected.

HYPOTHESES TO BE TESTED

In this study, the simulation model used tested the following hypotheses:

H1 *Auditor's conclusion on the population audited: using a 100-sample size*
The hypothesis tested is that this sampling plan, using a sample size of 100 items, accepts the population correctly over 90% of the time at all levels of error rate: 1, 2.5 and 5%.

H2 *Auditor's conclusion on the population audited: using a 50-sample size*

The hypothesis tested is that this sampling plan, using a sample size of 50 items, accepts the population correctly over 90% of the time at all levels of error rate: 1, 2.5 and 5%.

H3 *Auditor's conclusion on the population audited: using a 25-sample size*

The hypothesis tested is that this sampling plan using a sample size of 25 items, accepts the population correctly over 90% of the time at all levels of error rate: 1, 2.5 and 5%.

The various sample sizes used in testing these hypotheses are based on the research conducted in the UK which used sample sizes of 25 and 50 items to test large populations[3] under audit. The error value found in

---

[3] We assume that all accounting populations audited using sampling consist of several hundred and usually several thousand items.

TABLE 8
Auditor's conclusion on the population

Sample Size	Error Rate					
	1%		2.5%		5%	
	Accept	Reject	Accept	Reject	Accept	Reject
100	100	0	96	4	95	5
50	92	8	63	37	64	36
25	55	45	19	81	18	82

accounting populations is reported to be 0.5-5%. An auditor is likely to reject an accounting population thought to contain an error value exceeding 1%. The hypotheses above are intended to ascertain whether populations containing an error value of various magnitudes are likely to be rejected by an auditor using sample sizes of 100, 50 and 25 units.

RESULTS AND DISCUSSION

A simulation test was carried out to ascertain whether the sampling sizes used by auditors are likely to result in correct conclusions being drawn by the auditor on the acceptability of the population under audit, given three error levels and sample sizes.

The end product of the audit is either to reject or to accept the population under audit. If the upper error limit generated by the sample is less than the tolerable error, the sample results support the prior hypothesis that the population book value is not misstated by more than the tolerable error.

The simulation results are then compared with the actual data to ascertain the reliability of the auditor's conclusions. Table 8 shows the auditor's conclusions based on the various sample sizes and the percentage of times the auditor would accept or reject each particular population under the various conditions stated. The auditor's conclusion is that the population book value under audit is, or is not, in error by more than the tolerable error at the specified degree of risk.

If the level of correct decision as to acceptance or rejection generated by our simulation lies below the 90% level (the

auditor makes a wrong decision more than 10% of the time) then the audit procedures used would seem to be inadequate. For example, the audit testing procedure is telling the auditor to reject the population under audit when he should be accepting the population.[4]

The audit sampling plans using a sample size of 100 accepted the audited populations that should have been accepted over 90% of the time at all levels of error rate. The sampling plans using sample sizes of 25 and 50 units provided very different results.

With the error rate at 1% a sampling plan with a sample size of 50 accepts the population correctly more than 90% of the time. However, at an error rate of 2.5 and 5% a sampling plan of 50 provides acceptances far below 90% that is, it only accepts the population (the correct decision) 63% and 64% of the time respectively. The sampling plan based on a sample size of 25 produces an incorrect decision at all levels of error rate, that is, it produces the correct decision less than 90% of the time at all levels.

These findings suggest that firms using samples of fewer than 50 units for auditing accounting populations with low error rates have an unacceptably low probability of arriving at a correct conclusion on the quality of the population under audit and so should increase their minimum sample size per population audited to at least 50 units, and preferably 100 units. The auditor is too

---

[4] If an Auditor rejects a population he should accept this is called Alpha risk. If an Auditor accepts a population he should reject this is called Beta risk.

often rejecting populations he should accept, thus requiring needless extra audit work by both the auditor and the auditee.

However, in practice there are certain other qualitative issues that need to be considered in reaching an overall conclusion on accepting or rejecting an accounting population under audit. These qualitative factors might influence the auditor's conclusions derived from the audit sampling plan. It must also be noted that in this study the simulation was applied only to debtors' account of one particular industry.

However, the type of industry is unlikely to affect the conclusions since the statistical parameters of accounting distributions do not vary much between industries (Neter and Loebbecke 1975). The level of skewness attached to debtor distributions is similar to that attached to most other accounting distributions such as creditors and inventory. The rate of error and the distribution of total error are unlikely to vary in an inventory distribution compared to a debtor's or creditor's distribution. Therefore, we doubt if this parameter variation would have much effect on our conclusions as to the validity of the decisions to be drawn by auditors from small audit samples.

CONCLUSION

The objective of this study was to ascertain whether the different sample sizes drawn by audit firms do provide the auditor with an acceptable level of assurance as to the quality of the population under audit. The auditors must design a cost-effective sampling plan which will minimize both alpha and beta risk, that is an assurance that populations which should be rejected are not accepted, and vice-versa.

The simulation was based on an actual accounting distribution taken from Neter and Loebbecke (1975) Population 4. The sample sizes used were 25, 50 and 100 random items with a required confidence level set at 90 %. The findings are summarized in Table 9.

The results show that within the range of sample sizes normally used by auditors in practice, namely 25-100 units per population audited, the procedures only work consis-

TABLE 9
Summary of results of simulation analysis

Hypothesis	Accept/Reject
Hypothesis 1	*Accepted +*
Hypothesis 2	*Rejected in part* *
Hypothesis 3	*Rejected*

* at error rate of 1%, 2.5% and 5%, it is significantly above 95%

+ at error rate of 1%, the hypothesis is accepted

tently if the sample size is in the region of 100 random items. With samples of 50 random items the results vary somewhat, but for samples of 25 random items, the results are consistently negative. Since many earlier researchers (McRae 1982; Maysmor-Gee *et al.* 1984; Mohamad-Ali 1993) used fewer than 50 sample items per population audited (on average), the findings of this study should alert them in their future audit work. Hopefully, the size of their audit samples in the future would be increased to at least 50 items and preferably 100 items per population audited. This is based on the assumption that the populations under audit consist of several thousand items, though these results might also be true for very small accounting populations consisting of a few hundred items.

IMPLICATIONS FOR THE AUDITORS

Since the study covered only one accounting population, namely debtors, with a relatively low number of simulation runs (100), the conclusions drawn are largely tentative. Nevertheless, the results suggest that an auditor using any form of sampling should be concerned about the validity of the conclusions drawn from the sample when the sample size is below 50 units per population sampled. The findings suggest that audit samples below 50 are not large enough to mitigate alpha and beta risk.

To further validate the findings of this study, it is suggested that a larger number of accounting populations with other error distributions and larger simulation runs are collected and tested. It might also be useful to

run the simulation using other estimators, such as the so-called (MEST) bounds suggested by McCray (1980).

REFERENCES

FROST, P.A. and H. TAMURA. 1982. Jackknifed ratio estimation in statistical auditing. *Journal of Accounting Research* **20**: 103-120.

HAMMERSELY, J.M. and D.C. HANDSCOMB. 1964. *Monte Carlo Method*. London: Methuen.

JONES, H.L. Sampling plans for verifying clerical work. *Industrial Quality Control* **3(4)**: 5-11.

LESLIE, D.A., A.D. TEITLEBAUM and D. ANDERSON. 1980. *Dollar Unit Sampling*. Coop, Clarke & Pitman.

MAYSMOR-GEE, C., J. CRAMER and J. STANTOCKI. 1984. The audit process in the United Kingdom. City of Birmingham Polytechnic.

MCCRAY, J.H. 1980. A quasi-bayesian audit risk model for dollar unit sampling *Accounting Review* **151**: 35-51.

MCRAE, T.W. 1982. A study of the application of statistical sampling to external auditing. ICAEW.

MOHAMAD-ALI, A.H. 1993. A study of audit method: a study of the audit technique used by medium sized accounting firms in England. Ph.D. dissertation, University of Bradford.

NETER, J. and J.K. LOEBBECKE. 1975. Behaviour of major estimators in sampling accounting populations. New York: AICPA.

TAYLOR, R.G. 1974 Error analysis in audit tests. *Journal of Accountancy* **148**: 78-82.

VANCE, L.L. 1950. Auditing use of probabilities in selecting and interpreting test checks. *Journal of Accountancy* **88(3)**: 21-25.

(Received 11 January 1995)