Irish Accounting Review (1998), Vol. 5, No.1, 57–82

A COMPARISON OF LAHIRI AND SIEVE SAMPLING WITH TRADITIONAL MUS METHODS

Jane M. Horgan

Dublin City University

ABSTRACT

Lahiri sampling — which has not so far been used in auditing — and sieve sampling — which is relatively new — are compared with unrestricted random, systematic and cell sampling of monetary units, which have been in use for some time. Comparisons are carried out using the Stringer, cell and moment bounds by means of a simulation study based on two actual accounting populations with a realistic range of error rates and amounts. The results show that substantial gains in the precision of the bounds occur with systematic, cell and sieve sampling relative to both unrestricted random and Lahiri sampling in populations with large line items; the resultant decrease in the sample size is illustrated by the design effect. Reliability and tightness of the bounds are unaffected by the method of selection. Relative advantages of the schemes are discussed.

INTRODUCTION

When statistical methods are implemented in substantive testing to determine the accuracy of financial statements, monetary-unit sampling is the most common procedure used to select the sample of line items. A recent survey in England found that, of firms using statistical sampling in substantive testing, the monetary-unit strategy is used predominantly (Abdul-Hamid, 1993). A study carried out by Grimlund and Schroeder (1988) in the United States revealed that monetary-unit sampling is used to some degree or another by the twelve national public accounting firms, including the Big Six. Monetary-unit sampling is popular with auditors because it directs audit efforts towards the items which contain the greatest potential for major overstatement errors (Auditing Practices Board, 1993).

Many monetary-unit selection methods exist of which we examine five: three commonly used methods, unrestricted random, systematic and cell sampling; one relatively new method, sieve sampling; and a fifth, Lahiri sampling, which has not been used in auditing to date. Unrestricted random, systematic and cell sampling are *draw sequential* in the sense that they are carried out by a series of randomised experiments, called draws, from the entire population or from a given subset. With these sampling plans it is necessary to calculate the cumulative subtotals of the book values before selecting the sample. This can be time-consuming, especially if the population is large and if the selection is being carried out without the aid of a computer. Also, these methods ignore the line item structure of the population when selecting the monetary units and, since line items and not monetary units are audited, the selected monetary units must be traced back to their associated line items. Wurst, Neter and Godfrey (1989) maintain that this may at times create implementation problems. Sieve and Lahiri sampling, on the other hand, overcome these problems. They are *list sequential* in the sense that they exploit the natural line item structure of the population. They involve little or no lengthy calculations and may be implemented without the aid of a computer.

Horgan (1996) carried out a statistical comparison of unrestricted random, cell and sieve sampling using the moment bound, and found it to be more precise with cell and sieve sampling than with unrestricted random sampling. In the present study that analysis is extended by including two more sampling methods: (i) the Lahiri selection method, which has not yet been used in auditing and (ii) the well-known systematic strategy. The Stringer and cell bounds, as well as the moment bound, are included in the evaluation process. In order to ensure that this paper is self-standing, some of the descriptive material from Horgan (1996) is repeated, notably information on the accounting populations and the methods used to create the study populations for the simulation study.

The purpose of this paper is to compare the five sampling methods from the practical and statistical points of view. The paper is structured as follows. It begins with an outline of the selection procedures and goes on to describe the design of the simulation study. A comparison of the number of distinct line items obtained with each sampling method follows. Then the Stringer, cell and moment bound estimates of the total error amount are described. Next, three measures of assessing the statistical performance of the sampling methods are defined. The sampling methods are then compared in terms of the performance of the bounds with differing sample sizes drawn from populations with differing error rates and error amounts. Finally sample size planning is investigated using the design effect. The paper concludes with a summary and a discussion of the implications for auditors.

THE SELECTION METHODS

Monetary-unit sampling involves the selection of a sample of monetary units (e.g. £1) from an accounting population in order to estimate the total misstatement amount. It selects line items with probability proportional to book value size (PPS); hence large line items, which have a propensity to large overstatement errors, have a greater chance of selection than small line items.

Unrestricted random sampling

Unrestricted random sampling of monetary units, described in detail in Leslie, Teitlebaum and Anderson (1979, p.100), is a widely used audit selection procedure. A sample of n monetary units is obtained by first forming the book value subtotals and selecting a set of n random numbers between one and the total book amount. The random numbers identify the monetary units and the line items containing the selected monetary units are audited.

Unrestricted random sampling is equivalent to single stage PPS cluster sampling with replacement used in survey applications (see Cochran, 1977, pp. 233-247). To maintain PPS for line items, it is necessary to replace the line item after selecting a monetary unit; any line item may be chosen at any selection and therefore the number of distinct line items in the sample may be less than the target sample size.

Systematic sampling

Systematic sampling was first proposed by Madow (1949) in survey applications and was adapted for auditing by Anderson and Teitlebaum (1973). A systematic sample of *n* monetary units is selected by cumulating the book values, then dividing the total book value amount into *n* equal size intervals — the sampling interval — and finally selecting a set of monetary units systematically, after a random start in the first interval. The line items corresponding to the monetary units are chosen for auditing.

Systematic sampling may be viewed as a form of stratification where one unit is selected from each stratum. The sampling procedure ensures that the selected units are distributed evenly across the entire population; they come from distinct line items provided all the line items are less than the sampling interval. A disadvantage of systematic sampling is that it may result in biased selection if there are trends in the error pattern in the population; Jenne (1982) illustrates how, if the errors or groups of errors in the population are distributed as multiples of the sampling interval, a systematic sample may contain an extremely high or extremely low proportion of errors compared to the true proportion.

Cell sampling

Cell sampling was developed by Leslie et al. (1979, p. 103) to overcome the risk of bias inherent in systematic sampling. It is implemented in the same way except that an independent selection is made in each sampling interval (or cell). Like systematic sampling, the process of selection ensures that the sample is distributed evenly across the entire population. It is a form of stratification where one unit is selected independently from each cell or stratum. However, unlike systematic sampling, the selected monetary units do not necessarily come from distinct line items; repeated selections of monetary units from the same line item or "multiple hits", as they are referred to by Leslie et al. (1979, p. 105), may occur when a line item straddles two cells. Hence the number of line items audited may be less than the target sample size.

Sieve sampling

Sieve sampling, developed by Rietveld (1978, 1979a, 1979b) in the Netherlands, has gained some acceptance in Europe as an alternative to unrestricted random, systematic and cell sampling (Wurst et al., 1989). A sieve sample is selected by first dividing the total book amount by the sample size; this gives the sampling interval. For each line item a random number is selected independently in the sampling interval and the line item is selected if the random number is less than or equal to its book value.

While unrestricted random, systematic and cell sampling ignore the line item structure of the population when selecting the sample, sieve selection takes a different approach and exploits the natural structure of the population thus avoiding the need to form cumulative subtotals and to trace selected monetary units to their associated line items. The selection process is a form of stratified sampling, where strata of equal size are created and each line item is a subset of one stratum. Sieve sampling has the advantage that selected monetary units come from distinct line items but its problem is that the achieved sample size is variable; it may be greater than or less than the target depending on the random numbers chosen and, at the planning stage, it can only be predicted within certain probable limits.

Lahiri sampling

Lahiri sampling was first proposed as an alternative to single stage PPS cluster sampling "because the work involved in forming cumulative totals may be voluminous" (Lahiri, 1951). It may be adapted for auditing as follows. An item is selected by first choosing a random number between one and the number of line items in the population; this identifies a line item for consideration for auditing. Next a random number is chosen between one and the maximum book value; if the maximum is not known exactly when selecting the line items, it can be replaced by an estimate based perhaps on the auditor's previous experience. When the selected random number is less than or equal to the line item under consideration, it is selected for auditing. The process continues until the required sample size is obtained.

Clearly, this method involves no lengthy calculations and is ideal for use with manual records. Also, it does not require that the book values of the individual line items be determined prior to the audit. Even at the time of the audit, the book values are required only for the line items under trial and so the audit can begin before the complete set is available. A disadvantage is that, in order to maintain PPS, line items are selected with replacement and so the selected monetary units may not come from distinct line items; the number of line items audited may be less than the target sample size.

THE EMPIRICAL INVESTIGATION

A comparative investigation of the different sampling methods was carried out by means of a simulation study based on two real accounting populations of debtors summarised in **Table 1**.

Table 1:Characteristics of Two Accounting Populations

	Summary Statistics (£)	
	Population 1	**Population 2**
Total Book Value Amount	2,833,039.0	3,621,349.4
Mean Book Value	763.4	6,179.8
Standard Deviation	1,801.1	8,220.7
Skewness	6.7	1.9
Kurtosis	64.2	2.8
Minimum	2.0	1.0
First Quartile	87.0	552.8
Median	239.0	2,535.0
Second Quartile	640.0	6,727.1
Maximum	28,000.0	36,213.0
Number of Line Items	3711	586

The populations consist of all positive balances; negative or zero book values are eliminated; it is assumed that if these are important, the auditor would wish to audit them separately. Line items whose book values exceed the sampling interval are also excluded on the assumption that these are likely to be audited on a 100% basis. This is consistent with current audit practice where

> high value items are aggregated separately . . . and reported separately. They are not projected from the sample onto the population but are added to the statistical projection after the errors in the lower stratum have been assessed (Canadian Institute of Chartered Accountants, 1990).

Population 1 contains a relatively large number of small accounts and Population 2 contains a relatively small number of large accounts. These were chosen because previous studies have shown that sample selection methods are sensitive to line item size (see for example Plante, Neter and Leitch, 1985).

Creation of the study populations

Investigative audits were carried out on each of the populations to determine the patterns of line item error rates and taints (misstatement amounts relative to the associated book values). Line item error rates of 5.5% and 8.0% were observed in the samples drawn from Populations 1 and 2 respectively, and all the errors were overstatements; 44% of the misstatements in Population 1 and 75% in Population 2 were 100% overstated. With these data thirty populations were created by varying consistently the error rates and taints found in the investigative audits and generating these into the parent populations.

Five line item error rates were used to generate errors into each population; one exactly as found in the investigative audit, two lower and two higher. The lower error rates were one third and two thirds times the observed error rate and the higher ones were two and three times the observed error rate. The error rates ranged from 1.8% to 16.5% in study populations created from Population 1, and from 2.7% to 24.0% in those created from Population 2.

Three mean taint sizes were used: one exactly as found in the investigative audit, one lower and one higher. In Population 1, a regression model was fitted to the non-zero taints which were less than one, using the corresponding book value as the independent variable. A model with a slope of one third (three times) that of the fitted model was used to generate the lower (higher) taints. In Population 2, it was not possible to model the taints from the sample data because the observed number of non-zero taints less than one was too small. A theoretical distribution was used instead. While empirical evidence indicates that no single assumption about the shape of the taint distribution is appropriate in all audit situations, many studies have shown that most error taints follow a reversed J-shaped distribution with a mass at one and, consistent with previous research (for example Wurst et al., 1989; Leitch, Neter, Plante and Sinha, 1982; Peek, Neter and Warren, 1991), the non-zero taints less than one were modelled using an exponential distribution truncated at one with the parameter estimated from the sample data. The lower (higher) taints were generated using a parameter value of one third (three times) that of the sample estimate. In addition a proportion of line items, corresponding to that found in the investigative audit, were allocated taint values of one. The taint per monetary unit averaged over all monetary units ranged from 0.9% to 10.5% in the study populations created from Population 1, and from 2.1% to 12.3% in those created from Population 2. Thus, for each parent population, study populations with error patterns that might occur in real audit situations were created; these are summarised in **Table 2**.

Table 2: Error Characteristics of the Study Populations

	Total Overstatement Amount (£)		
Error Rate	**Taint 1**	**Taint 2**	**Taint 3**
Population 1			
1.8%	25,307.4	26,230.7	31,262.1
3.6%	54,177.4	56,233.2	68,179.3
5.5%	82,707.4	85,814.6	103,947.0
11.0%	159,588.2	165,496.2	199,156.2
16.5%	235,417.6	244,498.0	296,728.9
Population 2			
2.7%	75,630.0	77,798.6	79,514.6
5.4%	99,560.9	101,917.6	103,987.4
8.0%	139,608.6	147,593.8	156,188.9
16.0%	279,176.3	282,668.3	287,173.3
24.0%	411,963.5	423,523.7	446,652.0

With this data set, it is possible to investigate the behaviour of the sampling methods with different line item error rates and amounts in the same population based on the same relative error pattern. Full details of the investigative audits and the models used to generate errors into the populations are given in Horgan (1996).

Design of the simulation study

Samples of sizes 30, 60 and 100 were drawn from each of Populations 1 and 2 using each sampling method. The sample sizes represent the range currently used in audit practice (Abdul-Hamid, 1993). From each population one thousand replications were carried out for the various sample sizes and selection methods. The performance of the sampling methods was then examined in terms of (i) the number of the distinct line items obtained, (ii) the performance of upper bound estimates of the total misstatement amount and (iii) the design effect and its implication for sample size planning.

DISTINCT LINE ITEMS

It has already been pointed out that monetary-unit sampling selects line items with probability proportional to book value size and is equivalent to PPS selection in survey sampling. Clearly, line items should be selected without replacement and the number audited should exactly hit the target sample size. In survey applications, there are in fact many PPS designs which return a fixed sample size of distinct items; for example, Brewer and Hanif (1983) list some fifty. Many of these, however, can only be used effectively for samples of size two, and all become steadily more complex and difficult to implement as the sample size increases. This, of course, effectively rules them out for auditing. With the exception of systematic sampling, most practical PPS designs select items with replacement. It was to overcome this problem that Poisson sampling (Hájek, 1964), of which sieve sampling is a special case, was proposed but, although it returns a sample of distinct line items, the number in the sample is not constant; depending on the random numbers selected, it may be greater than or less than the target.

We now examine the number of distinct line items obtained with the different sampling methods. Summary statistics of the empirical sampling distribution of the number of distinct line items obtained with target sample sizes of 30, 60 and 100 drawn from Populations 1 and 2, each replicated 1000 times using unrestricted random, cell, sieve and

Lahiri sampling, are given in **Table 3**. Systematic sampling is not included because it selects a set of monetary units chosen from distinct line items.

Table 3: Number of Distinct Line Items Audited

	Population 1				Population 2			
Target Sample Size	**Mean**	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**
n = 30								
Urs	29.25	0.87	26	30	28.06	1.29	24	30
Cell	29.97	0.33	29	30	29.86	0.37	28	30
Sieve	30.07	5.53	15	45	30.09	4.74	16	45
Lahiri	29.23	0.87	26	30	28.00	1.30	23	30
n = 60								
Urs	57.23	1.59	50	60	52.29	2.39	49	59
Cell	59.76	0.49	57	60	58.82	1.07	55	60
Sieve	60.21	7.66	36	85	60.15	6.60	41	85
Lahiri	57.18	1.66	50	60	52.62	2.35	45	59
n = 100								
Urs	92.20	2.69	81	99	80.84	3.45	70	90
Cell	99.03	0.91	96	100	95.73	1.77	89	100
Sieve	99.78	9.07	74	125	99.82	7.44	78	129
Lahiri	92.29	2.63	83	98	80.83	3.47	67	92

From **Table 3** we see that, for any target sample size, unrestricted random and Lahiri sampling are similar with respect to the number of distinct line items obtained; their averages are always less than the target and substantially so when the target sample size is large. This is not surprising since both methods select line items with replacement and the probability of any line item being included more than once increases as the sample size increases.

With sieve sampling, on the other hand, while the average is near the target in all cases, the standard deviation is very large and so the actual number of distinct line items obtained in any sample may be substantially greater than or substantially less than the target.

Cell sampling overcomes both problems; the average is near the target for all sample sizes and the standard deviation is dramatically reduced;

although the number of distinct line items selected may, in some cases, be less than the target, it will not be very much less since the variation is small. For example, when n = 30, the maximum shortfall with cell sampling is only one in Population 1 and two in Population 2 and, although the shortfall increases as the sample size increases, it is never as great as that of the other sampling methods. When the target sample size n = 100, the shortfall is four in Population 1 and eleven in Population 2. This is in sharp contrast to unrestricted random and Lahiri sampling where, when n = 100, the number of line items selected never reaches the target. In Population 2, the maximum number of line items obtained with unrestricted random sampling for a target of 100 is 90 and it can be as low as 70. For Lahiri sampling the maximum is 92 with a minimum of 67.

UPPER BOUNDS

We examine the performance of the sampling methods in terms of the behaviour of upper confidence estimates of the total misstatement amount. Since estimators of the total misstatement amount based on large-sample normal distribution theory have been found to have a coverage much less than the stated confidence level (see for example Kaplan, 1973; Neter and Loebbecke, 1975; and Beck, 1980), auditors frequently use heuristic non-classical bound estimates. In this study, three such bounds which are widely used in practice are considered: the Stringer, cell and moment.

The Stringer bound

The Stringer bound (detailed in Stringer, 1963) is the best-known procedure (Felix, Leslie and Neter, 1982). It is calculated by obtaining an upper confidence limit for the line item error rate using the Poisson distribution and combining this with the taints observed in the sample to get an upper bound for the total misstatement amount. The Stringer bound is heuristic; while no proof of its validity exists, numerous empirical studies have confirmed that the coverage in repeated sampling is greater than the nominal confidence level. It tends, however, to be conservative in the sense that its value is usually much larger than the actual misstatement amount (see for example Plante et al., 1985; Wurst et al., 1989).

The cell bound

The cell bound, also widely used in practice, is similar to the Stringer bound in that it is calculated by combining the Poisson upper limit for the error rates with the taints. It was developed by Leslie et al. (1979, pp. 135-140) in an attempt to overcome the conservatism of the Stringer bound but studies have shown that, like the Stringer bound, its value is usually far greater than the total misstatement amount. Its exact form is given in Leslie et al. (1979, p. 142).

The moment bound

The moment bound was developed by Dworin and Grimlund (1984, 1986) to overcome the conservatism of the commonly used bounds, and it has been adopted by Arthur Andersen as a replacement for the Stringer bound (Felix, Grimlund, Koster and Roussey, 1990). The moment bound takes a different approach to the Stringer and cell methods and uses the gamma distribution to obtain an upper confidence limit for the average misstatement. It combines this with the sample line item error rate to get an upper estimate of the total misstatement amount. Its mathematical development is given in Dworin and Grimlund (1984).

PERFORMANCE MEASURES

The Stringer, cell and moment bound estimates of the total error amount are calculated at the 95% confidence level. Samples from each study population corresponding to the items in Populations 1 and 2 selected with the five sampling methods are used in the calculations. The selection methods are then compared in terms of the reliability, tightness and variability of the bounds in repeated sampling.

Reliability

Reliability refers to the coverage, the proportion of the 1000 replications returning estimates which are greater than or equal to the total misstatement amount at a specified nominal confidence level. A sampling method is said to be reliable for a particular bound if the coverage reaches the nominal confidence level; otherwise, it is unreliable.

Tightness

A sampling method may be reliable but the bound estimate may be far in excess of the total misstatement amount, thus possibly incurring extra costs by rejecting populations with less than material misstatement. For this reason a tightness measure is used to calculate how close the bound estimate is to the true overstatement amount; it is defined as the mean of the sampling distribution of the bound for the 1000 replications expressed as a percentage of the total misstatement amount. If this is small for a particular sampling method, the average estimate is near the true misstatement amount and the sampling method is said to be tight; if it is large, the sampling method is said to be conservative. Tightness is a relative measure and is used to facilitate comparisons between study populations with different overstatement amounts.

Variability

For any given tightness a bound which is less variable is preferable for distinguishing between acceptable and unacceptable misstatement amounts. In keeping with previous research (for example, Neter and Loebbecke, 1975; Reneau, 1978; Plante et al., 1985; Wurst et al., 1989) the standard deviation of the empirical sampling distribution of the bounds is used as a measure of variability. If this is small, the sampling method is said to be precise.

PERFORMANCE OF THE BOUNDS

The performance of the sampling methods is examined in **Tables 4** through **7**, where average results are given for each of four factors: error rate; taint level; sample size and bound. Each entry in the table represents the estimate of the total misstatement amount for the particular factor averaged over the other three. Tukey's pairwise comparison of means test at a family significance level of 5% is used to compare the sampling methods.

Performance with differing error rates

Table 4: Performance with Differing Error Rates, Population 1 and Population 2

Population 1					
Error Rate					
	1	2	3	4	5
% Reliability					
Urs	100.00	99.72	99.29	97.44	97.13
Sys	100.00	99.91	99.76	98.80	98.60
Cell	100.00	99.70	99.36	97.64	96.70
Sieve	100.00	99.73	99.36	96.77	97.27
Lahiri	100.00	99.72	99.41	97.34	97.15
% Tightness					
Urs	722.29	412.14	319.78	229.73	106.64
Sys	717.18	412.14	317.73	228.61	105.41
Cell	712.71	406.65	316.11	230.62	106.24
Sieve	717.10	310.32	316.57	229.34	105.98
Lahiri	720.95	409.50	317.98	228.56	105.53
Variability in £000					
Urs	56.53	81.95	96.44	121.70	137.08
Sys	53.80	79.34	90.95	107.08	118.98
Cell	54.20	77.20	91.56	118.54	132.56
Sieve	56.73	80.17	96.41	123.91	138.92
Lahiri	57.74	79.42	93.47	123.04	137.64
Population 2					
Error Rate					
	1	2	3	4	5
% Reliability					
Urs	99.80	99.80	98.84	97.21	97.09
Sys	100.00	100.00	100.00	98.47	99.07
Cell	99.89	99.83	98.23	98.57	98.35
Sieve	99.94	99.91	99.44	98.70	97.97
Lahiri	99.83	99.86	98.45	98.96	98.83
% Tightness					
Urs	416.49	352.27	287.84	214.26	94.38
Sys	415.35	353.26	286.31	214.58	105.48
Cell	416.08	353.27	286.57	216.53	106.01
Sieve	418.61	352.30	286.24	214.78	103.83
Lahiri	409.76	346.57	281.87	213.32	104.33
Variability in £000					
Urs	107.44	118.66	137.04	173.21	192.93
Sys	85.91	102.20	120.36	167.99	154.47
Cell	95.99	110.67	128.45	157.58	180.57
Sieve	90.08	104.63	126.59	161.08	187.30
Lahiri	103.96	119.00	139.52	177.06	200.50

Examining the performance of the sampling methods for each error rate, the trend is clear from **Table 4** that, as the error rate increases, the coverage decreases, the tightness improves and the precision decreases for each of the sampling methods. All sampling methods reach the nominal confidence level. No significant differences occur between the sampling methods with respect to tightness; in the low error rate populations the sampling methods are extremely conservative but they become less so as the error rate increases. With respect to the precision, the greatest gains occur in large line item populations (Population 2), where systematic (sys), cell and sieve sampling give consistently more precise estimates than unrestricted random (urs) and Lahiri sampling; in study populations generated from Population 1, systematic and cell sampling are more precise than unrestricted random and Lahiri sampling in all cases. The improvement in precision over unrestricted random and Lahiri sampling is not surprising since, as noted earlier, systematic, cell and sieve sampling are forms of stratification and therefore exclude some extreme combinations.

Performance with differing taints

Table 5: Performance with Differing Taints

	Population 1			**Population 2**		
	Taint Levels			**Taint Levels**		
	1	2	3	1	2	3
	% Reliability			% Reliability		
Urs	98.99	98.91	98.25	98.62	98.51	98.52
Sys	99.63	99.57	99.10	99.66	99.48	99.38
Cell	99.15	99.09	98.40	98.22	98.15	99.15
Sieve	98.94	98.85	98.08	99.22	99.18	99.18
Lahiri	98.95	98.89	98.21	98.34	98.27	98.41
	% Tightness			% Tightness		
Urs	401.73	389.47	337.15	295.41	290.79	286.94
Sys	399.44	387.38	335.83	295.73	290.75	286.53
Cell	397.94	385.72	333.74	198.22	291.66	287.39
Sieve	399.18	387.04	335.38	295.78	290.95	286.73
Lahiri	300.00	389.47	335.71	291.72	286.91	282.89
	Variability in £000			Variability in £000		
Urs	97.13	97.87	101.21	143.33	135.46	147.78
Sys	88.28	89.02	92.79	125.31	125.81	127.45
Cell	93.04	93.88	97.71	133.63	134.46	135.91
Sieve	97.39	98.29	102.62	132.72	133.60	135.47
Lahiri	97.13	98.32	101.00	146.53	147.77	149.71

Turning to the taints, it is clear from **Table 5** that the coverage decreases and the sampling methods become tighter and less precise as the taint level increases. Systematic sampling has the highest coverage for all taint levels but the differences are not of practical importance since the coverage reaches the nominal in all cases. The selection method does not have any effect on the tightness of the estimates for any taint size. Systematic sampling has the highest precision across all taint levels. Gains in precision also occur with cell and sieve sampling over unrestricted random and Lahiri sampling; these are substantial in large line item populations (Population 2).

Performance with differing sample sizes

Table 6: Performance with Differing Sample Sizes

	Population 1			Population 2		
	n = 30	n = 60	n = 100	n = 30	n = 60	n = 100
	% Reliability			% Reliability		
Urs	99.66	98.34	98.15	99.26	98.73	97.65
Sys	99.69	98.27	99.33	99.47	99.32	99.74
Cell	99.60	98.53	98.51	99.44	99.04	99.05
Sieve	99.53	98.28	98.07	99.24	99.03	99.31
Lahiri	99.52	99.18	98.35	99.16	98.45	97.40
	% Tightness			% Tightness		
Urs	530.22	338.22	259.91	378.64	273.63	220.87
Sys	523.93	335.49	263.22	379.50	272.88	220.65
Cell	519.48	337.93	259.99	383.78	269.80	221.50
Sieve	524.61	337.08	259.90	382.45	270.64	220.37
Lahiri	521.22	339.68	261.92	380.00	263.84	217.65
	Variability in £000			Variability in £000		
Urs	135.71	92.60	68.14	195.34	136.56	105.67
Sys	125.74	85.54	61.80	199.89	106.88	71.81
Cell	128.84	90.44	65.37	187.87	127.84	88.35
Sieve	136.48	92.84	68.97	194.98	124.78	82.04
Lahiri	132.85	93.62	68.33	197.70	139.10	107.14

Clearly, the results in **Table 6** illustrate that the sampling methods are reliable for all sample sizes; some significant differences in coverage exist between systematic sampling and the other selection methods but these are not of any practical importance. The sampling method does not have any effect on the tightness of the bound; estimates of the total misstatement amount are extremely conservative with small sample sizes but become less conservative as the sample size increases. With the exception of $n = 30$ in Population 2, systematic sampling is the most precise for all sample sizes and it is significantly more precise than unrestricted random and Lahiri sampling with samples of sizes 60 and 100. Gains in precision over unrestricted and Lahiri sampling also occur with cell and sieve sampling; these gains are significant in Population 2 with n = 100. The precision results are consistent with Wurst et al. (1989), who found that improvements in precision of sieve sampling over unrestricted random sampling are greatest when the sample size is not small.

Performance with different bounds

Table 7: Performance with Different Bounds

	Population 1			Population 2		
	Stringer	**Cell**	**Moment**	**Stringer**	**Cell**	**Moment**
	% Reliability			% Reliability		
Urs	99.22	99.02	97.91	98.87	98.82	97.98
Sys	99.78	99.70	98.81	99.64	99.64	99.64
Cell	99.42	99.18	98.04	99.42	99.40	98.72
Sieve	99.22	98.99	97.67	99.40	99.39	98.78
Lahiri	99.22	99.03	97.76	98.68	98.62	97.72
	% Tightness			% Tightness		
Urs	393.26	392.22	339.86	195.95	195.14	182.04
Sys	393.90	389.85	338.88	295.33	294.61	283.07
Cell	392.78	388.68	335.95	296.83	296.06	282.18
Sieve	395.53	390.13	338.73	296.42	295.67	281.17
Lahiri	391.27	390.31	338.71	292.10	291.34	278.06
	Variability in £000			Variability in £000		
Urs	96.63	97.41	102.17	141.13	141.23	155.20
Sys	88.21	88.90	92.98	122.73	122.69	133.15
Cell	92.96	93.75	97.93	130.26	130.39	143.40
Sieve	97.52	98.25	102.51	129.70	129.81	142.28
Lahiri	96.63	97.01	101.55	143.40	143.53	157.08

With respect to the bounds **Table 7** demonstrates that the comparative performance of the sampling methods is similar for all bounds. There are no substantial differences in the average coverage or tightness between sampling methods for any of the three bounds. Differences in reliability and coverage that exist can be attributed to factors other than the sampling method. Consistent with the findings of Dworin and Grimlund (1984), the moment bound has a lower coverage and it is substantially tighter than the Stringer and cell bounds.

With respect to precision, systematic sampling gives the most precise estimates in all cases. Cell and sieve sampling are substantially more precise than unrestricted random and Lahiri sampling in study populations generated from Population 2. The moment bound is less precise than the Stringer and cell bounds with all sampling methods.

DESIGN EFFECT

The previous section has shown that, although the sampling methods do not appear to have any profound effect on the coverage and tightness of the bounds for any error rate, taint level, sample size or bound, they do affect the precision. The differences in precision of systematic, cell, sieve and Lahiri sampling compared to unrestricted random sampling are now examined in more detail using the design effect suggested by Kish (1965, pp. 257-259); this is defined as the ratio of the variance of the bound estimator with a particular sampling method compared to its variance with unrestricted random sampling. In sample size planning, the design effect may be interpreted as the proportionate increase or decrease in the sample size using the particular sampling method to obtain the same precision as unrestricted random sampling. For example if the estimated design effect of cell sampling is 0.8 (1.2) with a sample of size 60, then a sample of size $n = 60 \times 0.8 = 48$ ($60 \times 1.2 = 72$) will give the same precision as unrestricted random sampling.

Tables 8 and **9** give the design effects of systematic, cell, sieve and Lahiri sampling with target samples of sizes 30, 60 and 100 together with the sample size necessary to obtain the same precision as unrestricted random sampling. The design effects are calculated using the Stringer bound from study populations with taint level 1. The design effects with the other bounds and taints are similar and are not presented here.

Table 8: Design Effects in Population 1

Error Rate	Sampling Method	Design Effect	Sample Size	Design Effect	Sample Size	Design Effect	Sample Size
		n = 30		n = 60		n = 100	
1	Sys	0.97	30	0.92	53	0.92	92
	Cell	0.84	26	1.03	62	0.92	92
	Sieve	1.00	30	1.05	63	0.92	92
	Lahiri	1.01	31	1.12	68	0.99	99
2	Sys	1.02	31	0.95	57	0.82	82
	Cell	0.86	26	0.94	57	0.87	87
	Sieve	0.97	30	1.00	60	0.88	88
	Lahiri	0.92	28	0.86	52	0.99	99
3	Sys	0.98	30	0.87	53	0.82	82
	Cell	0.91	28	0.94	57	0.87	87
	Sieve	0.96	29	1.09	66	0.97	97
	Lahiri	0.88	27	0.93	56	0.97	97
4	Sys	0.66	20	0.87	53	0.69	69
	Cell	0.93	28	0.95	57	0.96	96
	Sieve	1.02	31	0.97	59	1.15	115
	Lahiri	1.00	30	0.99	60	1.06	106
5	Sys	0.75	23	0.63	38	0.90	90
	Cell	0.94	29	0.92	56	0.93	93
	Sieve	1.07	33	0.95	57	1.07	107
	Lahiri	1.00	30	0.99	60	1.05	105

Table 9: Design Effects in Population 2

Error Rate	Sampling Method	Design Effect	Sample Size	Design Effect	Sample Size	Design Effect	Sample Size
		n = 30		n = 60		n = 100	
1	Sys	1.02	31	0.54	33	0.26	26
	Cell	0.89	27	0.83	50	0.63	63
	Sieve	0.87	27	0.70	42	0.43	43
	Lahiri	0.92	28	0.99	60	0.97	97
2	Sys	1.27	39	0.61	37	0.27	27
	Cell	0.98	30	0.87	53	0.68	68
	Sieve	0.96	29	0.77	47	0.51	51
	Lahiri	1.01	31	1.02	62	0.98	98
3	Sys	1.09	33	0.65	39	0.44	44
	Cell	0.96	29	0.88	53	0.74	74
	Sieve	1.00	30	0.85	51	0.61	61
	Lahiri	1.02	31	1.08	65	1.02	102
4	Sys	1.28	39	0.84	51	0.53	53
	Cell	0.86	26	0.88	53	0.71	71
	Sieve	1.02	31	0.85	51	0.63	63
	Lahiri	1.04	32	1.04	63	1.06	106
5	Sys	0.76	23	0.92	56	0.78	78
	Cell	0.96	29	0.90	54	0.62	62
	Sieve	1.09	33	0.91	55	0.74	74
	Lahiri	1.10	33	1.06	64	1.11	111

From **Table 8** we see that, in study populations generated from Population 1, the design effect of systematic sampling is less than one in most cases; with cell, sieve and Lahiri sampling, it is sometimes less than one and sometimes greater than one but the fluctuations show no consistent pattern and could possibly be attributed to random fluctuation.

In study populations generated from Population 2 (**Table 9**), the pattern is clear: the design effects of systematic, cell and sieve sampling decrease as the sample size increases with the greatest gains in the precision occurring with large sample sizes in low error rate populations. With systematic sampling, the design effect is as low as 0.26 indicating that a sample of 26 will give the same precision as an unrestricted sample of size 100. We see from **Table 9** that substantial decreases in sample size also occur with cell and sieve sampling.

Tables 8 and **9** reinforce the results observed previously and provide further evidence of the superiority of systematic, cell and sieve sampling over unrestricted random and Lahiri sampling.

SUMMARY

This study examined the comparative performance of five monetary-unit selection methods, of which three are commonly used in practice, one has gained some acceptance in Europe, and a fifth has not yet been used in auditing. The sampling methods were compared by means of a large-scale simulation study using two real accounting populations with a range of likely line item error rates and amounts.

The results show that the number of distinct line items selected in the sample for a given target sample size varies from sampling method to sampling method. With cell sampling, it may be less than the target but not substantially so for any target sample size. This contrasts with sieve sampling where the number of distinct line items may be very much greater or very much less than the target. It also contrasts with unrestricted random and Lahiri sampling where, although the number of monetary units is constant, nonetheless the sample contains an *a priori* unpredictable number of distinct line items which may be very much less than the target sample size. With systematic sampling, the number of distinct items is constant and always hits the target.

The sampling methods were also compared in terms of the reliability, tightness and variability of the Stringer, cell and moment bound estimates of the total misstatement amount measured at the 95% level of confidence. The results show that, while there were no substantial differences in the sampling methods with respect to coverage and tightness of the bounds, differences in precision did occur. Systematic sampling was found to be the most precise method of selection with significant gains in precision occurring with large sample sizes in large line item populations. While not as precise as systematic sampling, cell and sieve sampling were also found to be more precise than unrestricted random and Lahiri sampling; the greatest gains occurred in large line item populations when the sample size was not small. The precision of Lahiri sampling was similar to unrestricted random sampling in all cases.

IMPLICATIONS FOR AUDITORS

In general, when the reliability and tightness are the same, the auditor chooses the selection method with the greatest precision. Hence, if one were to choose based solely on the statistical performance, systematic sampling would be the preferred option; its coverage and tightness are similar to the other selection methods and its variability is lowest in most cases. It has the added advantage that it returns a sample of target size consisting of distinct line items. However, systematic sampling may result in biased estimates if there are trends in the data. The auditor may be unwilling to take this risk and so may opt for either cell or sieve sampling. Both are more precise than unrestricted random and Lahiri sampling but neither one uniformly dominates the other. Sieve sampling is convenient to implement because it uses the line item structure of the population when selecting the sample and thus avoids the need to obtain cumulative subtotals. It has the added advantage that it returns a sample of distinct line items. The main disadvantage of sieve sampling is that the number of line items audited is variable and it may be less than or greater than the target. When the number of line items is less than the target, the credibility of the audit may be decreased, and when it is greater than the target, the cost is increased. If the auditor finds that this may cause problems when planning the audit, then cell sampling is preferred where, although the number of distinct line items in the sample may be less than the target, the discrepancy is not substantial. Also, cell sampling is not affected to the same extent as systematic sampling by a periodic trend in the error pattern in the population. Leslie et al. (1979, p. 110) contend that it is preferable to systematic sampling because "it is not significantly harder to do and it avoids any doubt as to rigorousness".

Unrestricted random and Lahiri sampling would be discarded if the decision were based purely on statistical grounds; both are consistently less precise than the other three selection methods, especially in large line item populations, and the actual number of line items audited may be substantially less than the target sample size. However, Lahiri sampling is useful when the selection process is being carried out without the aid of a computer; it is ideal for the auditor who, for whatever reason, may need to carry out the sampling process manually. The selection procedure involves little or no lengthy calculations and can be readily invoked when the data consist of manual records. It has the added advantage that the audit may begin before the complete set of book values is available and, even at the time of the audit, the book values are required only for the line items under trial. This will almost certainly avoid delays since, as Leslie et al. (1979, p. 101) point out, "often the total book value is not known accurately during the planning stage, nor is it known for transaction streams prior to the end of the year". Unrestricted random sampling has neither statistical nor practical advantages; it would be used only when nothing else is available.

Finally, it should be noted that the conclusions are based on empirical studies using two actual accounting populations. Although the study populations were created with a wide range of error patterns in an effort to generate populations with characteristics representative of the variety found in real situations, it should not be forgotten that the principal weakness of empirical studies is that the conclusions apply only to the specific populations actually studied. While the results presented here are consistent with other comparative studies of sample selection methods (for example Plante et al., 1985; Wurst et al., 1989), an obvious next step would be to test the sampling methods further on other populations with different error patterns. Also, larger sample sizes (for example 150 and 200) and different confidence levels (for example 85% and 70%) could be used if deemed appropriate.

REFERENCES

Abdul-Hamid, M.A. (1993). 'A Study of the Audit Method: Audit Techniques Employed by Medium-Sized Accounting Firms in England', unpublished Ph.D Thesis, University of Bradford.

Anderson, R. and Teitlebaum, A.D. (1973). 'Dollar-Unit Sampling: A Solution to the Audit Dilemma', *CA Magazine (Canada)*, Vol. 106, pp. 30-38.

Auditing Practices Board (1993). *Draft Guidelines: Auditing Sampling*, London: Institute of Chartered Accountants in England and Wales.

Beck, P.J. (1980). 'A Critical Analysis of the Regression Estimator in Audit Sampling', *Journal of Accounting Research*, Vol. 50, pp. 16-37.

Brewer, K.R.W. and Hanif, M. (1983). *Sampling with Unequal Probabilities*, New York: Springer-Verlag.

Cochran, W.G. (1977). *Sampling Techniques*, New York: John Wiley.

Canadian Institute of Chartered Accountants (1990). *Interactive Data Extraction Analysis*, Version 4.0, Users Manual, Canada.

Dworin, L.D. and Grimlund, R.A. (1984). 'Dollar Unit Sampling for Accounts Receivable and Inventory', *The Accounting Review*, Vol. 59, April, pp. 218-241.

Dworin, L.D. and Grimlund, R.A. (1986). 'Dollar-Unit Sampling: A Comparison of the Quasi-Bayesian and Moment Bounds', *The Accounting Review*, Vol. 61, January, pp. 36-57.

Felix, W.L., Grimlund, R.A., Koster, F.J. and Roussey, R.S. (1990). 'Arthur Andersen's New Monetary Unit Sampling Approach', *Auditing: A Journal of Practice and Theory*, Vol. 9, Fall, pp. 1-16.

Felix, W.L., Leslie, D.A. and Neter, J. (1982). 'University of Georgia, Centre for Audit Research Monetary-Unit Sampling Conference, March 24, 1981', *Auditing: A Journal of Practice and Theory*, Vol. 1, Winter, pp. 92-103.

Grimlund, R.A. and Schroeder, M.S. (1988). 'On the Current Use of the Stringer Method of MUS: Some New Directions', *Auditing: A Journal of Practice and Theory*, Vol. 7, Fall, pp. 53-62.

Hájek, J. (1964). 'Some Contributions to the Theory of Probability Sampling', *Annals of the Institute of Mathematical Statistics*, Vol. 36, pp. 127-133.

Horgan, J.M. (1996). 'The Moment Bound with Unrestricted Random, Cell and Sieve Sampling of Monetary Units', *Accounting and Business Research*, Vol. 26, Summer, pp. 215-223.

Jenne, S.E. (1982). 'A Critical Examination of Dollar-Unit Sampling in the Audit of Accounts Receivable', unpublished Ph.D thesis, University of Illinois at Urbana-Champaign.

Kaplan, R.S. (1973). 'Statistical Sampling in Auditing with Auxiliary Information Estimators', *Journal of Accounting Research*, Vol. 43, Autumn, pp. 238-258.

Kish, L. (1965). *Survey Sampling*, New York: John Wiley.

Lahiri, D.B. (1951). 'A Method of Sample Selection Providing Unbiased Ratio Estimators', *Bulletin of the International Statistics Institute*, Vol. 33, pp. 133-140.

Leitch, R.A., Neter, J., Plante, R. and Sinha, P. (1982). 'Modified Multinominal Bounds for Larger Numbers of Errors in Audits', *Accounting Review*, Vol. 57, pp. 384-401.

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

Madow, W.G. (1949). 'On the Theory of Systematic Sampling', *Annals of Mathematical Statistics*, Vol. 20, pp. 333-354.

Neter, J. and Loebbecke, J.K. (1975). *Behaviour of Major Statistical Estimators in Sampling Accounting Populations — An Empirical Study*, New York: AICPA.

Peek, L.E., Neter, J. and Warren, C. (1991). 'AICPA Nonstatistical Audit Sampling Guidelines: A Simulation', *Auditing: A Journal of Practice and Theory*, Vol. 10, Fall, pp. 33-48.

Plante, R., Neter, J. and Leitch, R.A. (1985). 'Comparative Performance of the Multinominal, Cell and Stringer Bound', *Auditing: A Journal of Practice and Theory*, Vol. 4, Fall, pp. 40-56.

Reneau, J.H. (1978). 'CAV Bounds in Dollar Unit Sampling: Some Simulation Results', *Accounting Review*, Vol. 53, July, pp. 669-680.

Rietveld, C. (1978). 'De Zeefmethode Als Selectiemethode Voor Statistiche Steekproeven in de Controlepraktijk (I)', *Compact: Computer en Accountant*, Amsterdam: Klynveld Kraayenhof & Co., Vol. 15, pp. 2-11.

Rietveld, C. (1979a). 'De Zeefmethode Als Selectiemethode Voor Statištiche Steekproeven in de Controlepraktijk (II) en (III)', *Compact: Computer en Accountant*, Amsterdam: Klynveld Kraayenhof & Co., Vol. 16, pp. 2-13.

Rietveld, C. (1979b). 'De Zeefmethode Als Selectiemethode Voor Statistiche Steekproeven in de Controlepraktijk (IV)', *Compact: Computer en Accountant*, Amsterdam: Klynveld Kraayenhof & Co., Vol. 17, pp. 9-18.

Stringer, K.W. (1963). 'Practical Aspects of Statistical Sampling in Auditing', *Proceedings of the Business and Economics Statistics Section, American Statistical Association*, December, pp. 405-411.

Wurst, J., Neter, J. and Godfrey, J. (1989). 'Comparison of Sieve Sampling with Random and Cell Sampling of Monetary Units', *The Statistician*, Vol. 38, pp. 235-249.