Regression Sampling in Statistical Auditing: A Practical Survey and Evaluation

J.P.C. Kleijnen
Department of Information Systems and Auditing
Catholic University Brabant
P.O. Box 90153
NL-5000 LE Tilburg, The Netherlands

J. Kriens, H. Timmermans, H. van den Wildenberg
Department of Econometrics
Catholic University Brabant
P.O. Box 90153
NL-5000 LE Tilburg, The Netherlands

Several confidence intervals for the regression estimator are surveyed. A Monte Carlo experiment, based on the NETER and LOEBBECKE (1975) populations, gives estimated coverages and lengths of the different confidence intervals. One interval is exact under the assumption of multivariate normal distributions; it gives longer intervals (hence better coverages) than the interval based on a popular variance estimator. An interval due to ROBERTS (1970) is much too long. Jackknifing gives robust intervals. Rules of thumb for practitioners are given.

Key Words & Phrases: control variates, confidence interval length, coverage, robustness, Neter-Loebbecke populations, Roberts procedure, jackknifing.

1. INTRODUCTION

This research was initiated by some questions arising in the practice of statistical auditing. In particular, we consider the following situation. There is a population of book values y_i, and there is a corresponding population of actual (true) values x_i with $i=1,...,N$. For example, there are N "stock keeping units" (SKU's) which according to the books have values y_i whereas their actual values are x_i. The auditor has to verify that the *total* book value $\Sigma_1^N y_i$ does not deviate "too much" from the total audit value $\Theta=\Sigma_1^N x_i$. To save time, the auditor takes a sample of size n, without replacement, and observes both the true value $\mathbf{x}_j$ and the book value $\mathbf{y}_j$ with $j=1,...,n$ and $n<N$ (random variables are in boldface). This sample yields an estimator $\bar{\mathbf{x}}=\Sigma_1^n \mathbf{x}_j/n$ for the mean audit value $\mu_x=\Sigma_1^N x_i/N=\Theta/N$. The known population mean $\Sigma_1^N y_i/N$ is compared with the estimator $\bar{\mathbf{x}}$ (or, better, with some more efficient estimator, as we shall see in a moment). The total book value $\Sigma_1^N y_i$ is accepted as a

substitute for the total (unknown) audit value $\Sigma_1^N x_i$, if the $1-\alpha$ confidence interval $[L_\alpha, U_\alpha]$ around the estimator $\bar{\mathbf{x}}$ "covers" the mean book value $\Sigma_1^N y_i/N$:

$$1-\alpha = P(L_\alpha \leqslant \bar{\mathbf{x}} \leqslant U_\alpha) = P(nL_\alpha \leqslant n\bar{\mathbf{x}} \leqslant nU_\alpha) \tag{1}$$

and if $L_\alpha < \Sigma_1^N y_i/N < U_\alpha$ then the auditor does not reject the hypothesis

$$H_0: \Theta(= \Sigma_1^N x_i) = \Sigma_1^N y_i. \tag{2}$$

The problem investigated in this paper, is the construction of a confidence interval that has indeed the correct probability $1-\alpha$, even if the statistical assumptions do not hold; and among the intervals with acceptable coverage we look for the shortest interval. If the $\mathbf{x}_j$ are Normally and Independently Distributed with constant mean μ_x and constant variance σ_x^2 - or $\mathbf{x}_j \in NID(\mu_x, \sigma_x^2)$ - then elementary mathematical statistics yields a $1-\alpha$ confidence interval (see eq. 3 later on). If, however, the $\mathbf{x}_j$ are skewly distributed (as they are in auditing; see table 2), then the classical interval has a *coverage probability* smaller than $1-\alpha$; see KLEIJNEN (1987, pp. 18-23). Moreover (whether the $\mathbf{x}_j$ are normal or not) the confidence interval may be *too long* for practical purposes; see §2. Then the $1-\alpha$ confidence interval for the audit values can be made much shorter through the use of information on the book values. We focus on *regression estimators,* including four different confidence intervals for these regression estimators; we also study one classical interval. This paper is organized as follows.

(i) We study the classical or crude procedure based on $\mathbf{x}_j \in NID(\mu_x, \sigma_x^2)$ in §2. This estimator is efficient if the correlation between $\mathbf{x}$ and $\mathbf{y}$ is "weak".
(ii) KRIENS and PETERSE (1986) use the normal distribution and a popular variance estimator to derive a confidence interval for the regression estimator; see §3.
(iii) ROBERTS (1978) proposes a special confidence interval procedure for the regression estimator in auditing, discussed in §4.
(iv) LAVENBERG et al. (1982) apply the Student distribution and the Ordinary Least Squares variance estimator, which gives an optimal $1-\alpha$ confidence interval, provided $\mathbf{x}_j$ and $\mathbf{y}_j$ are bivariate-normal; see §5.
(v) In §6 we introduce jackknifing combined with the Student distribution. Jackknifing is a very general "trick" - see MILLER (1974) - which has already been applied to a related estimator (namely the ratio estimator) by FROST and TAMURA (1982).

In practice the statistical assumptions do not hold. NETER and LOEBBECKE (1972) developed a set of realistic auditing populations (see §7), which we use in a Monte Carlo simulation, in order to estimate the actual coverages of the five confidence intervals (with nominal coverages of $1-\alpha$) and the interval lengths; see §8. In §9 we give guidelines for practitioners. A more detailed report - namely KLEIJNEN et al. (1988), - is available from the authors.

2. The classical or crude estimator $\bar{\mathbf{x}}$

It is usually assumed that

$$P(\bar{\mathbf{x}}-t_{n-1}^{(\alpha/2)}(1-n/N)^{\frac{1}{2}}\mathbf{s}_x/\sqrt{n}\leqslant\mu_x\leqslant\bar{\mathbf{x}}+t_{n-1}^{(\alpha/2)}(1-n/N)^{\frac{1}{2}}\mathbf{s}_x/\sqrt{n})\approx 1-\alpha \quad (3)$$

where $t_{n-1}^{(\alpha/2)}$ is the $1-(\alpha/2)$ quantile of Student's t statistic with $n-1$ degrees of freedom, $\bar{\mathbf{x}}$ is the sample average or crude estimator, $\mathbf{s}_x$ is the classical estimator of the standard deviation (with denominator $n-1$), and $(1-n/N)^{\frac{1}{2}}$ is the finite-population correction; see COCHRAN (1963, p. 26). As we mentioned in §1, if the mean *book* value $\Sigma_1^N y_i/N$ lies within the $1-\alpha$ confidence interval (3), then the auditor assumes that the total book value $\Sigma_1^N y_i$ represents the unknown true value $\Theta(=\Sigma_1^N x_i=N\mu_x)$. Such a procedure involves the risk of erroneously accepting $\Sigma_1^N y_i$: *type II error.* This risk increases as the length of the confidence interval increases. Therefore auditors may work with estimators more refined than the crude $\bar{\mathbf{x}}$, as we shall see next.

3. The regression estimator $\bar{\mathbf{x}}(\hat{\gamma})$

Estimators can become more efficient than the crude estimator $\bar{\mathbf{x}}$, if we incorporate information on the corresponding book values in the sample, namely $\mathbf{y}_j$. The difference estimator or control variate estimator corrects the crude estimator $\bar{\mathbf{x}}$ for the deviation between the estimator $\bar{\mathbf{y}}$ and the *known* population value $\mu_y=\Sigma_1^N y_i/N$:

$$\bar{\mathbf{x}}(\gamma) = \bar{\mathbf{x}} + \gamma(\mu_y-\bar{\mathbf{y}}). \quad (4)$$

Evidently, for any given constant γ, the estimator $\bar{\mathbf{x}}(\gamma)$ is an unbiased estimator of μ_x.

It is well-known that the variance of $\bar{\mathbf{x}}(\gamma)$ is minimized by $\gamma_0=\mathrm{cov}(\mathbf{x},\mathbf{y})/\mathrm{var}(\mathbf{y})$. The obvious estimator for γ_0 is

$$\hat{\gamma} = \frac{\widehat{\mathrm{cov}}(\mathbf{x},\mathbf{y})}{\widehat{\mathrm{var}}(\mathbf{y})} = \frac{\hat{\rho}(\mathbf{x},\mathbf{y})\mathbf{s}_x}{\mathbf{s}_y}, \quad (5)$$

which is identical to the Ordinary Least Squares (OLS) estimator $\hat{\gamma}$ in

$$\hat{\mathbf{x}} = \hat{\delta} + \hat{\gamma}\mathbf{y}, \quad (6)$$

so that the estimator $\hat{\gamma}$ can be computed using an algorithm available on any computer. Substitution of $\hat{\gamma}$ into (4) results in a non-linear estimator, called the *regression estimator:*

$$\mathbf{x}(\hat{\gamma}) = \bar{\mathbf{x}} + \hat{\gamma}(\mu_y-\bar{\mathbf{y}}). \quad (7)$$

This estimator is no longer unbiased. Its variance can be estimated in several ways; DENG and WU (1987) prove that many popular variance estimators underestimate the true variance. KRIENS and PETERSE (1986, p. 28) use:

$$s^2\{\bar{x}(\hat{\gamma})\} = \frac{(1-n/N)}{n}\mathbf{s}_x^2[1-(\hat{\rho}(\mathbf{x},\mathbf{y}))^2]. \tag{8}$$

In §5 we shall see that this estimator is biased: it underestimates the variance. If $\mathbf{s}$ denotes $[s^2\{\bar{x}(\gamma)\}]^{\frac{1}{2}}$ and $z^{(\alpha/2)}$ is the $1-(\alpha/2)$ quantile of the standard normal $\mathbf{z}$, then Kriens and Peterse's $1-\alpha$ confidence interval is

$$[\bar{x}(\hat{\gamma})-z^{(\alpha/2)}\mathbf{s}, \bar{x}(\hat{\gamma}) + z^{(\alpha/2)}\mathbf{s}]. \tag{9}$$

They conjecture - also see BAKER and COPELAND (1979) - that this confidence interval gives correct coverage $(1-\alpha)$, if the following conditions hold:

(i) The *sample size n* is at least 200.
(ii) The number of *non-zero* differences $(\mathbf{x}_j \neq \mathbf{y}_j)$, say $\mathbf{m}$, is at least 5% of n, and at least 20.
(iii) There are both *positive* and *negative* differences, to be tested by the sign statistic. (The sign test concerns $H_0 : p = 0.50$ where $p = P(\mathbf{x}-\mathbf{y}>0)$.)

Our study examines these conditions, as we shall see. Condition (ii) is the central issue of the following procedure.

4. THE ROBERTS (1978) PROCEDURE ACCOUNTING FOR m

ROBERTS (1978) proposes a confidence interval for the regression estimator $\bar{x}(\hat{\gamma})$ which accounts for $\mathbf{m}$, the number of non-zero differences. Suppose the population has a low *error percentage,* that is, most book values y_i are identical to the corresponding audited values $x_i (i=1,...,N)$. Then $\mathbf{m}$ is small, with high probability. Roberts's procedure accounts for $\mathbf{m}$. Unfortunately, his procedure is lengthy and difficult to understand intuitively. Because our Monte Carlo experiment shows that his procedure is completely inferior, we do not present his formulas but refer to ROBERTS (1978, pp. 90-91) or KLEIJNEN et al. (1988, pp. 34-37) for details. Further, our initial Monte Carlo study gave extremely long confidence intervals for Roberts's procedure so that we applied an ad hoc tric to the first step of his procedure (his factor $P_U(m)$ is an upper bound for $\mathbf{m}/n$; we replace $P_U(m)$ by m/n; for example, for $\mathbf{m}= 15$ and $n= 180$ Roberts gives $P_U(\mathbf{m})= 0.13$ whereas we use $\mathbf{m}/n= 0.08$ so that in step 5 of the procedure a lower estimated standard deviation follows).

5. THE LAVENBERG ET AL. (1982) BIVARIATE-NORMAL MODEL

LAVENBERG et al. (1982, pp. 183-185) consider the simulation of a queuing system, each server q with its own known mean service time $\mu_q (q=1,...,Q)$. The simulation program samples actual service times from the Q service time distributions which results in the average service times $\mathbf{c}_q$. So Lavenberg et al. discuss situations with $Q \geqslant 1$ auxiliary or control variables (instead of a single control variate as is the case in our auditing situation). KLEIJNEN et al. (1988, pp. 37-41) gives the details of their procedure and translates their symbols into our symbols. However, we can also apply their idea immediately to the simple regression model (6).

Eqs. (5) and (6) show that we may use the familiar OLS algorithm to compute the estimate $\hat{\gamma}$ for the regression estimator $\bar{x}(\hat{\gamma})$. Now we introduce the *conditional* model

$$E(\mathrm{x}|y) = \delta + \gamma(y-\mu_y) \tag{10}$$

and assume that the distribution of x given y is normal with constant variance σ^2. This model holds, if (but not only if) the pair (x,y) is bivariate normally distributed. Then OLS gives the estimators $\hat{\boldsymbol{\delta}}'$ and $\hat{\boldsymbol{\gamma}}'$, where the dash indicates that the estimators are considered conditionally on the y values; the formulas for the estimates $\hat{\delta}'$ and $\hat{\gamma}'$ are identical to those for $\hat{\delta}$ and $\hat{\gamma}$. Obviously

$$\hat{\boldsymbol{\delta}}' = \bar{\mathbf{x}} + \hat{\boldsymbol{\gamma}}'(\mu_y - \bar{y}) \equiv \bar{x}(\hat{\gamma}|y). \tag{11}$$

Let W denote the $n \times 2$ matrix of explanatory variables corresponding to (10): $w_{i1}=1$ and $w_{i2}=y_i-\mu_y$. Let v_{11} denote the element (1,1) of $(W'W)^{-1}$ and $\hat{\boldsymbol{\sigma}}^2$ the Mean Squared Residuals. Then we obtain for $\delta'=E(\hat{\delta}')=\mu_x$ (where the last equality follows from the conditional normality):

$$P[\hat{\boldsymbol{\delta}}' - t_{n-2}^{(\alpha/2)}(v_{11}\hat{\boldsymbol{\sigma}}^2)^{\frac{1}{2}} \leq \mu_x \leq \hat{\boldsymbol{\delta}}' + t_{n-2}^{(\alpha/2)}(v_{11}\hat{\boldsymbol{\sigma}}^2)^{\frac{1}{2}}] = 1-\alpha. \tag{12}$$

Since the right-hand side does not depend on y, the left-hand side also holds unconditionally, that is, we can make y random. The confidence interval for μ_x follows from (12) if we make y random, so that $\hat{\boldsymbol{\gamma}}'$ and $\hat{\boldsymbol{\delta}}'$ become $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\delta}}$ and v_{11} becomes v_{11}. It is well-known - see RAO (1988, p. 459) - that

$$\mathrm{v}_{11}\hat{\boldsymbol{\sigma}}^2 = \frac{n-1}{n(n-2)}\left[1 + \frac{(\bar{\mathrm{y}}-\mu_y)^2}{\mathrm{s}_y^2}\right]\mathrm{s}_x^2[1-(\hat{\rho}(\mathrm{x},\mathrm{y}))^2], \tag{13}$$

which may be compared with the variance estimator used by KRIENS and PETERSE (1986); see eq. (8). The latter estimate is always smaller than the former. For further comparisons we refer to KLEIJNEN et al. (1988, pp. 10-11).

The bivariate-normal model, which implies constant variances, may be violated in auditing applications; see BECK (1980), FROST and TAMURA (1982) and §7.1.

6. JACKKNIFE

We first explain jackknifing in general; next we apply the jackknife to the regression estimator. Suppose we have a sample of size n for the random variable $\mathbf{z}$. This sample yields the estimator (say) $\hat{\Theta}$; for example, $\hat{\Theta}$ denotes the sample average $\bar{\mathbf{z}}$. Next we delete the jth variable and from the $(n-1)$ remaining variables we compute the same estimator $\hat{\Theta}_{-j}$; for example, $\bar{\mathbf{z}}_{-1} = \sum_{j_1'=2}^{n} \mathbf{z}_{j_1'}/(n-1)$. Permutation yields n estimators $\hat{\Theta}_{-1},\ldots,\hat{\Theta}_{-n}$; for example $\bar{\mathbf{z}}_{-1},\ldots,\bar{\mathbf{z}}_{-n}$. Now we define the *n pseudo-values* J_j:

$$\mathbf{J}_j = n\hat{\Theta} - (n-1)\hat{\Theta}_{-j} \qquad (j=1,\ldots,n). \tag{14}$$

It can be proved that $\mathbf{J}_j$ has smaller bias if $\hat{\theta}$ had any bias. Moreover, if the variable **z** is not normally distributed, then the jackknife confidence interval may still hold. To compute this interval we treat the $\mathbf{J}_j$ as if they were NID (also see eq. 3):

$$1-\alpha \approx P\left(\frac{|\bar{\mathbf{J}}-\theta|}{s_J/\sqrt{n}} \leqslant t_{n-1}^{(\alpha/2)}\right) \tag{15}$$

where $\bar{\mathbf{J}}=\Sigma_1^n \mathbf{J}_j/n$ and $s_J^2=\Sigma_1^n(\mathbf{J}_j-\bar{\mathbf{J}})^2/(n-1)$. See MILLER (1974) and also KLEIJNEN (1987, p. 424).

Jackknifing the *regression estimator* is now very simple. We had a sample of n independent pairs $(\mathbf{x}_j,\mathbf{y}_j)$ from which we computed the regression estimator $\bar{x}(\hat{\gamma})$, using (7). Now we delete pair $j(j=1,...,n)$ and from the remaining $(n-1)$ pairs we again compute the regression estimator $\bar{x}(\hat{\gamma}_{-j})$. This yields the pseudo-values; see (14) with $\hat{\theta}=\bar{x}(\hat{\gamma})$ and $\hat{\theta}_{-j}=\bar{x}(\hat{\gamma}_{-j})$. Finally (15) yields the $1-\alpha$ confidence interval for μ_x. LAVENBERG et al. (1982) also examined the jackknifed regression estimator, but not in an auditing situation.

DENG and WU (1987) prove that the jackknifed regression estimator yields an overestimated variance; NELSON (1988) gives a recent survey of regression estimators including jackknifing; these authors do not use an auditing context. FROST and TAMURA (1982) apply jackknifing to a related estimator, namely the ratio estimator; also see KLEIJNEN et al. (1988, pp. 13-15).

7. MONTE CARLO INPUT OR EXPERIMENTAL DESIGN

In §7.1 we describe the populations from which we sample in our Monte Carlo experiment; in §7.2 we specify some other inputs to the Monte Carlo experiment.

7.1 The Neter and Loebbecke (1975) accounting populations

We use a set of realistic data developed by NETER and LOEBBECKE (1975, pp. 11-32), which are also used in several other studies on statistical auditing. The data are nicely characterized by FROST and TAMURA (1982, pp. 109-111) from which table 1 is reproduced. Note that populations 3 and 4 have positive errors only. Only population 3 has approximately constant error variances ($\hat{\gamma}\approx 0$).

BECK (1980) characterizes not the errors $\mathbf{x}-\mathbf{y}$ but the book values y; see table 2. In the Monte Carlo experiment different error rates create different audit values x_i, not different book values $y_i(i=1,...,N)$.

NETER and LOEBBECKE (1975) further modify accounting population 1: book values higher than \$950 are not sampled but are investigated individually; so only the population of accounts with book balances of \$950 or less are studied, on a sampling basis. We also utilize this modified population, denoted by 1M in the tables of the next section.

Table 2 shows that the distribution of y is not Gaussian, especially populations 1 and 1*M* are not. We emphasize that a *logarithmic* transformation does not help: although log(x) and log(y) are more normally distributed, the $1-\alpha$ confidence interval for $E\{\log(x)\}$ does not yield a $1-\alpha$ confidence interval for $E(x)=\Theta'$; see PATTERSON (1966).

7.2. *Other Monte Carlo inputs*

We study sample sizes n equal to 200 (the minimum size required by KRIENS and PETERSE, 1986) and 100. (FROST and TAMURA, 1982, study $n=$ 50, 100, 200; BECK, 1980, p. 20, takes n = 200, 600.)

TABLE 1. Characteristics of Error Distribution in Study Populations

Accounting Population	Error Rate	[θ] ($1.000)	Distribution of Errors: Mean	σ	Skewness	Kurtosis	Heteroscedas.: $\hat{\gamma}$	$s(\hat{\gamma})$	cv(y) n = 100
1	30%	380	-.09	4.1	-1.1	72.6	.86	.19	.291
1	10%	380	-.05	2.7	-1.8	186.1	.96	.23	.291
1	5%	379	.01	1.9	6.2	466.1	.90	.32	.291
1	1%	379	-.02	.9	-30.9	1442.6	.94	.47	.291
1	½%	379	-.01	.5	-31.0	2287.3	-.40	.53	.291
2	70%	3.565	-14.2	194.3	-1.5	41.3	.94	.08	.182
2	10%	3.491	-.8	80.9	-1.4	360.2	1.11	.12	.182
2	5%	3.491	-.8	63.9	-7.4	748.8	1.48	.16	.182
2	1%	3.486	.2	25.7	48.4	3267.7	.97	.37	.182
2	½%	3.487	-.1	8.6	-9.7	912.0	.22	.34	.182
3	30%	13.510	23.0	72.7	4.5	22.4	.28	.31	.361
3	10%	13.623	6.9	39.6	8.4	82.4	.20	.30	.361
3	5%	13.648	3.4	27.0	12.1	176.1	.26	.30	.361
3	1%	13.666	.8	13.5	26.5	827.0	-.01	.36	.361
3	½%	13.669	.4	9.7	42.0	2029.4	-.30	.38	.361
4	30%	6.442	263.0	1529.7	9.9	113.8	1.92	.02	.208
4	10%	7.237	65.9	662.9	19.6	471.5	1.83	.04	.208
4	5%	7.402	24.9	302.5	25.3	840.3	1.74	.06	.208
4	1%	7.469	8.5	222.9	46.1	2448.6	1.60	.40	.208
4	½%	7.478	6.2	216.5	49.9	2751.5	1.44	.41	.208

Reproduced from Frost and Tamura (1982, p. 110).

TABLE 2. Neter-Loebbecke Populations of Account Book Values

Population	1	2	3	4
Total book value	$379,131	$3,486,530	$13,671,500	$7,502,957
Mean	45.63	636	1,946	1,860
Standard deviation	132.61	1,156	7,022	3,865
Skewness	22.0	3.5	7.9	3.2
Kurtosis	906.4	15.2	78.1	11.4
Largest account	6,869.70	9,989	98,163	24,928.6
Smallest account	.50	1.00	.10	.10
Error pattern	small, offsetting	moderate, primarily understate-ment	small overstatement	large overstatement (hetero-scedastic)

Reproduced from Beck (1980, p. 23).

Though the auditor takes samples without replacement, we sample with replacement, because the latter procedure is simpler; both procedures are equivalent from a practical viewpoint, since the factor n/N is smaller than 0.05. We repeat this sampling 600 times (as NETER and LOEBBECKE, 1975, do; BECK (1980, p. 20) takes only 300 replications).

The pseudorandom numbers are created by a multiplicative congruential generator with multiplier 13^{13} and modulus 2^{59}, developed by NAG (Numerical Algorithms Group, United Kingdom). We never reset the seed, so that all results in the following tables may be considered independent, except for results on the same line (obtained by applying different estimators to the same sample data $\mathbf{x}_j, \mathbf{y}_j$).

8. MONTE CARLO OUTPUT

We are interested in the probability that the constructed "nominal" $1-\alpha$ confidence interval indeed covers the true mean Θ. ROBERTS (1978) requires a number of non-zero differences higher than two ($\mathbf{m}>2$) (he uses t with $m-2$ degrees of freedom). Therefore we do not compute a confidence interval if $\mathbf{m}\leqslant 2$. If $\mathbf{m}=0$ then we do not reject H_0. Let $\mathbf{K}$ denote the number of iterations with either $\mathbf{m}>2$ or $\mathbf{m}=0$ (§7 implies that $0 \leqslant \mathbf{K} \leqslant 600$). Let $\mathbf{v}_k$ be 0 if the $(1-\alpha)$ confidence interval does not cover the true mean; else $\mathbf{v}_k$ is 1 $(k=1,...,\mathbf{K})$. Then the estimator for the *coverage* equals $\sum_1^{\mathbf{K}} \mathbf{v}_k/\mathbf{K}$. A "conservative" confidence interval has expected coverage higher than $1-\alpha$; for example, the interval from zero to infinity is certainly conservative. Therefore we also estimate the expected *length* of the confidence interval.

We emphasize that we also compute a confidence interval in case (nearly) all differences $\mathbf{x}_j - \mathbf{y}_j$ have the same sign; see populations 3 and 4. If in practice one does not compute a confidence interval for such cases, then the coverages will be higher than our estimated coverages (which will turn out to be low).

Table 3 shows that ROBERTS (1978) gives too high coverages; table 4 shows that such coverages imply very long confidence intervals. An exception is population 4 (large overstatement; see table 2). However, for population 4 the crude estimator behaves somewhat better than Roberts's estimator (table 3); its confidence interval length remains inferior (table 4). KRIENS and PETERSE (1986) give results very close to LAVENBERG et al. (1982), as §5 suggested. *Jackknifing* gives coverages which in most cases are closest to the nominal value $1-\alpha$ $(=0.95)$; consequently its confidence intervals are slightly longer than KRIENS and PETERSE (1986) and LAVENBERG et al. (1982); see tables 3 and 4. The main conclusion is that confidence intervals with practical lengths give inferior coverage for the non-normal populations investigated.

Comparing tables 3 and 5 shows how the coverages change when the *sample size* increases from 100 to 200. The *crude* estimator $\bar{\mathbf{x}}$ gives higher estimated coverages (central limit theorem); of course the lengths of its confidence intervals decrease (table 4 versus table 6); its intervals remain much longer than the other estimators' intervals (table 6). ROBERTS (1978) remains too conservative. Kriens and Peterse - and hence Lavenberg et al. - and jackknifing have coverages that do not clearly benefit from sample size increase; these methods do clearly result in shorter confidence intervals as n increases (tables 4 and 6).

KRIENS and PETERSE (1986) conjecture that the normal approximation holds, if the *number of non-zero differences* is at least 20 $(\mathbf{m} \geqslant 20)$ (and if some more conditions hold; see §3). Therefore we also *stratify* the Monte Carlo results on $\mathbf{m}$, which gives table 7 (the last column aggregates the preceding two columns). We aggregate the error percentages per population. Let $\mathbf{K}_s$ denote the number of Monte Carlo observations in stratum $s (s=1,2,3,4)$. An asterisk (*) means that the estimated coverage in the stratum is *not* significantly different from the nominal $1-\alpha=0.95$; this difference is tested at $\alpha'=0.05$; unfortunately, a small $\mathbf{K}_s$, tends to create an asterisk (small power of test; we estimate the variance by $\alpha(1-\alpha)/K_s$, and approximate the binomial distribution's quantile by the normal distribution quantile). We do not display the results for LAVENBERG et al. (1982) because these results are very similar to KRIENS and PETERSE (1986).

Table 7 suggests that condition (ii) of KRIENS and PETERSE (1986) is too strong; their method also gives good results for $15 < m \leqslant 20$. Note that for populations 3 and 4 Kriens and Peterse do not compute a confidence interval because the sign test rejects the null hypothesis (only positive errors). For populations 1 and 2 the point estimates $1-\hat{\alpha}$ exceed the nominal value $1-\alpha(=0.95)$ so that-whatever the size of the standard errors - the hypothesis of correct coverage is not rejected. For population 1M the point estimate $1-\hat{\alpha}$ is not significantly low; however, the standard error is rather high because $\mathbf{K}_s$ is small. For $n=200$ the results are similar; for details see KLEIJNEN et al. (1988, p. 29).

TABLE 3. Estimated Coverages for Nominal $1-\alpha=$ 0.95 and Sample Size $n=$ 100

Population	Error Rate (%)	Crude Est.	Kriens-Peterse	Lavenberg et al.	Jackknife	Roberts
1	0.5	.82	1.00	1.00	1.00	1.00
1	1.0	.84	.82	.83	.83	1.00
1	5.0	.79	.95	.96	.99	1.00
1	10.0	.79	.89	.90	.93	1.00
1	30.0	.80	.87	.89	.94	1.00
1M	1.0	.87	.84	.86	.84	1.00
1M	10.0	.88	.90	.91	.94	1.00
2	0.5	1.00	1.00	1.00	1.00	1.00
2	1.0	.94	.96	.96	.96	1.00
2	5.0	.91	.85	.86	.87	.99
2	10.0	.90	.97	.98	.98	1.00
2	70.0	.92	.93	.93	.97	1.00
3	0.5	.74	1.00	1.00	1.00	1.00
3	1.0	.82	.91	.92	.92	1.00
3	5.0	.83	.77	.78	.78	.94
3	10.0	.83	.78	.79	.79	.96
3	30.0	.83	.89	.90	.90	1.00
4	0.5	.92	.75	.75	.75	.83
4	1.0	.86	.86	.87	.86	.98
4	5.0	.93	.59	.60	.60	.88
4	10.0	.93	.60	.61	.62	.84
4	30.0	.91	.71	.71	.75	.97

TABLE 4. Average Lengths (in thousands) of Estimated Confidence Intervals for Sample Size $n=100$ and Nominal $1-\alpha=0.95$

Population	Error Rate(%)	Crude Est.	Kriens Peterse	Lavenberg et al.	Jackknife	Roberts
1	0.5	385	5	5	5	48
1	1.0	344	3	3	3	30
1	5.0	323	4	5	5	21
1	10.0	323	6	7	7	26
1	30.0	317	11	12	14	70
1M	1.0	220	3	3	3	25
1M	10.0	221	7	7	7	27
2	0.5	2,609	27	28	28	294
2	1.0	2,407	64	66	67	538
2	5.0	2,421	93	96	99	402
2	10.0	2,373	134	138	143	520
2	70.0	2,246	385	393	420	3,350
3	0.5	15,904	74	78	74	797
3	1.0	16,538	52	53	52	470
3	5.0	17,064	62	64	63	275
3	10.0	17,188	96	99	97	372
3	30.0	17,399	190	198	194	1,142
4	0.5	6,269	617	638	691	4,606
4	1.0	6,015	292	301	324	2,590
4	5.0	5,947	301	311	321	1,306
4	10.0	5,861	660	680	742	2,658
4	30.0	5,484	1,751	1,806	2,096	11,341

TABLE 5. Estimated Coverages for Nominal $1-\alpha = 0.95$ and Sample Size $n = 200$

Population	Error Rate (%)	Crude Est.	Kriens-Peterse	Lavenberg et al.	Jackknife	Roberts
1	0.5	.85	.84	1.00	.84	1.00
1	1.0	.87	.76	.76	.78	1.00
1	5.0	.84	.95	.96	.98	1.00
1	10.0	.82	.91	.92	.95	1.00
1	30.0	.85	.89	.89	.95	1.00
1M	1.0	.89	.80	.81	.83	1.00
1M	10.0	.89	.90	.91	.94	1.00
2	0.5	.90	.96	.96	.96	1.00
2	1.0	.94	.84	.84	.85	1.00
2	5.0	.94	.91	.91	.92	.99
2	10.0	.95	.96	.96	.97	1.00
2	70.0	.94	.94	.95	.96	1.00
3	0.5	.90	.92	.92	.92	.96
3	1.0	.88	.87	.87	.87	.98
3	5.0	.87	.81	.82	.81	.97
3	10.0	.88	.88	.89	.89	1.00
3	30.0	.87	.91	.92	.91	1.00
4	0.5	.91	.83	.83	.83	.95
4	1.0	.92	.74	.74	.73	.99
4	5.0	.93	.66	.66	.66	.90
4	10.0	.94	.68	.69	.69	.95
4	30.0	.93	.81	.82	.84	1.00

TABLE 6. Average Lengths (in thousands) of Estimated Confidence Intervals for Sample Size $n = 200$ and Nominal $1-\alpha = 0.95$

Population	Error Rate(%)	Crude Est.	Kriens-Peterse	Lavenberg et al.	Jackknife	Roberts
1	0.5	247	2	2	2	16
1	1.0	248	2	2	2	14
1	5.0	250	3	3	4	13
1	10.0	252	5	5	6	27
1	30.0	255	8	9	10	70
1M	1.0	157	2	2	2	14
1M	10.0	156	5	5	5	25
2	0.5	1,678	15	15	15	142
2	1.0	1,721	30	31	31	219
2	5.0	1,700	65	67	70	250
2	10.0	1,714	97	100	101	473
2	70.0	1,723	280	287	298	3,381
3	0.5	12,882	23	23	23	205
3	1.0	12,897	27	27	27	187
3	5.0	12,795	48	49	49	185
3	10.0	12,916	71	73	73	348
3	30.0	12,934	136	140	139	1,130
4	0.5	4,192	294	304	327	2,809
4	1.0	4,168	171	177	179	1,194
4	5.0	4,143	229	237	244	896
4	10.0	4,136	561	581	609	2,813
4	30.0	3,886	1,414	1,463	1,573	12,460

TABLE 7. Estimated Coverages after Stratification on Number of Non-zero Differences **m**, for Sample Size $n = 100$ (* means *not* significant at $\alpha' = 0.05$)

POPULATION 1 (number of obs.)	$5 < m \leq 10$ (585)	$10 < m \leq 15$ (243)	$15 < m \leq 20$ (42)	$10 < m \leq 20$ (285)
Crude $\bar{x}$	0.762	0.815	0.810	0.814
Kriens & Peterse	0.911	0.914	0.952*	0.919
Jackknifing	0.952*	0.951*	0.976*	0.954*
Roberts	1	1	1*	1
POPULATION 1M	$5 < m \leq 10$ (333)	$10 < m \leq 15$ (212)	$15 < m \leq 20$ (24)	$10 < m \leq 20$ (236)
Crude $\bar{x}$	0.844	0.920	0.875*	0.915
Kriens & Peterse	0.901	0.910	0.917*	0.911
Jackknifing	0.931*	0.953*	0.958*	0.953*
Roberts	1	1	1*	1
POPULATION 2	$5 < m \leq 10$ (561)	$10 < m \leq 15$ (240)	$15 < m \leq 20$ (31)	$10 < m \leq 20$ (271)
Crude $\bar{x}$	0.914	0.883	0.936*	0.884
Kriens & Peterse	0.959*	0.992	0.968*	0.989
Jackknifing	0.970	1	0.968*	0.996
Roberts	0.998	1	1*	1
POPULATION 3	$5 < m \leq 10$ (559)	$10 < m \leq 15$ (244)	$15 < m \leq 20$ (34)	$10 < m \leq 20$ (278)
Crude $\bar{x}$	0.832	0.828	0.971*	0.845
Kriens & Peterse	0.775	0.934*	0.912*	0.932*
Jackknifing	0.782	0.943*	0.912*	0.950*
Roberts	0.968*	0.992	1*	0.993
POPULATION 4	$5 < m \leq 10$ (572)	$10 < m \leq 15$ (235)	$15 < m \leq 20$ (22)	$10 < m \leq 20$ (257)
Crude $\bar{x}$	0.928	0.936*	0.773	0.922
Kriens & Peterse	0.633	0.762	0.818	0.767
Jackknifing	0.647	0.775	0.864*	0.782
Roberts	0.855	0.949*	0.909*	0.946*

9. CONCLUSIONS

Monte Carlo studies often use arbitrary experimental designs. In audit sampling, however, the NETER and LOEBBECKE (1975) populations are accepted as representative.

We addressed the question: may the auditor accept the total book value $(\Sigma_1^N y_i)$ as a correct figure for the total audit value $(\Sigma_1^N x_i)$? To save time the auditor starts with a sample and observes both the book values $\mathbf{y}_j$ and the corresponding audit values $\mathbf{x}_j (j=1,\ldots,n)$. Some *preliminary conditions* on these n pairs $(\mathbf{x}_j,\mathbf{y}_j)$ are:

(i) If "nearly" all differences $\mathbf{x}_j - \mathbf{y}_j$ have the *same sign*, then the bookkeeping systematically overestimates or underestimates the true values, and the auditor will not simply accept the book values as a substitute for the audit values; instead additional audit procedures follow to find out whether there are acceptable reasons for this phenomenon. KRIENS and PETERSE (1986, p. 30) use the *sign* statistic to test if $P(\mathbf{x}_j > \mathbf{y}_j) = 0.5$; we recommend such a test.

(ii) If all book values in the sample are *equal* to the audit values, then the auditor needs no mathematical statistics to conclude that the books are extremely accurate and to accept the book values y_i.

If these two preliminary conditions do not hold, then a confidence interval for the audit value is needed.

The *literature* gives several point estimators and confidence intervals. ROBERTS (1970) gives a complicated procedure that turns out to be inferior. FROST and TAMURA (1982) investigate the ratio estimator, possibly combined with jackknifing, but they find several extremely low coverages such as $1-\hat{\alpha} = 0.30$. BECK (1980) examines regression estimators without jackknifing; our study indicates that jackknifing the regression estimators yields a more robust estimator (also see below).

The *crude* estimator $\bar{\mathbf{x}}$ does not yield better coverage, except for population 4; then, however, the crude estimator gives confidence intervals so long that they are of no practical value. The regression estimators drastically reduce the lengths of the confidence intervals. The rule of thumb in KRIENS and PETERSE (1986), namely $m \geqslant 20$, can be relaxed to $m > 15$ in their procedure.

Practitioners want simple guidelines. We give the following *rules of thumb:*

(i) The use of *auxiliary information*, namely the population book values y_i and their sampled values $\mathbf{y}_j$, yields a confidence interval much shorter than the *crude* estimator $\bar{\mathbf{x}}$ does.

(ii) *Jackknifing* improves the coverage of the regression estimator; yet that coverage is not acceptable if the error distributions are too skew or the error percentages are very low. Fortunately, the sample $(\mathbf{x},\mathbf{y})$ gives a clear warning; if the error distribution is skew, then most errors have the same sign and the auditor will proceed to additional audit procedures not based on sampling. If there are only a few non-zero differences, then the auditor may accept the book value. In all other situations the auditor uses the jackknife regression estimator.

Note that future research may try to adjust the above procedures such that correct coverages result, and next compare the interval lengths. KLEIJNEN et al. (1986) made a start, showing that better coverage results if the estimated skewness is incorporated into the t test for the classical estimator $\bar{x}$.

ACKNOWLEDGEMENT

We thank two anonymous referees and the Editor (prof. Snijders) for their comments which lead to a better presentation of our research.

REFERENCES

BAKER, R.L. and R.M. COPELAND (1979), Evaluation of the stratified regression estimator for auditing accounting populations, *Journal of Accounting Research 17,* pp. 606-617.

BECK, P.J. (1980), A critical analysis of the regression estimator in audit sampling, *Journal of Accounting Research 18,* pp. 16-37.

COCHRAN, W.G. (1963), *Sampling Techniques,* Second edition, John Wiley & Sons, New York.

DENG, L.Y. and C.F.J. WU (1987), Estimation of variance of the regression estimator, *Journal of the American Statistical Association 82,* pp. 568-576.

FROST, P.A. and H. TAMURA (1982), Jackknifed ratio estimator in statistical auditing, *Journal of Accounting Research 20,* pp. 103-120.

KLEIJNEN, J.P.C. (1987), *Statistical Tools for Simulation Practitioners,* Marcel Dekker, Inc., New York.

KLEIJNEN, J.P.C., G.L.J. KLOPPENBURG and F.L. MEEUWSEN (1986), Testing the mean of an asymmetric population: Johnson's modified t test revisited, *Communications in Statistics, Simulation and Computation 15,* pp. 715-732.

KLEIJNEN, J.P.C., J. KRIENS, H. TIMMERMANS and H. VAN DEN WILDENBERG (1988), Regression sampling in statistical auditing, Research Memorandum, FEW 306, Catholic University Brabant, Tilburg.

KRIENS, J. and J.J.M. PETERSE (1986), Toepassing van de regressieschatter in de accountantscontrole. (Application of the regression estimator in auditing.), *Kwantitatieve Methoden 20,* pp. 21-36.

LAVENBERG, S.S., T.L. MOELLER and P.D. WELCH (1982), Statistical results on control variables with application to queueing network simulation, *Operations Research 30,* pp. 182-202.

MILLER, R.G. (1974), The Jackknife - a review, *Biometrika 61,* pp. 1-15.

NELSON, B.L. (1988), Control-variate remedies, Department of Industrial and Systems Enginering, The Ohio State University, Colombus (Ohio 43210).

NETER, J. and J.K. LOEBBECKE (1975), *Behavior of Major Statistical Estimators in Sampling Accounting Populations: An Empirical Study,* American Institute of Certified Public Accountants, Inc., New York.

PATTERSON, R.L. (1966), Difficulties involved in the estimation of a population mean using transformed sample data, *Technometrics 8,* pp. 535-537.

RAO, P.S.R.S. (1988), Ratio and regression estimators. in: P.R. Krishnaiah and C.R. Rao (eds.), *Handbook of Statistics, Volume 6,* North-Holland, Amsterdam, pp. 449-468.

ROBERTS, D.M. (1978), *Statistical Auditing,* American Institute of Certified Public Accountants, New York.

Received March 1988, Revised February 1989.