A MONTE CARLO STUDY OF SOME RATIO ESTIMATORS

J. N. K. Rao and LeNelle D. Beegle
Texas A & M University

Summary

Small-sample efficiencies of eight ratio estimators of the population ratio $\bar{Y}/\bar{X}$ are investigated by Monte Carlo methods, assuming a linear regression of y on x and x normally distributed. Efficiencies of three variance estimators are also investigated. Sample criteria for which expected values exist are used. From this study, Mickey's unbiased ratio estimator and the approximately unbiased ratio estimator obtained by Quenouille's method appear promising.

1. Introduction

Ratio estimators are often employed in sample surveys for estimating the population mean $\bar{Y}$ of a characteristic of interest 'y' or the population ratio $\bar{Y}/\bar{X}$ utilizing a supplementary variate 'x' that is positively correlated with 'y'. It is well-known that the classical ratio estimator is biased and often, in practice, the bias may be negligible compared to standard error and can be neglected. However, the bias may become considerable in surveys with many strata and small or moderate size samples within strata if it is considered appropriate to use "seperate ratio estimators". In these situations, the use of unbiased or approximately unbiased (i.e., estimators with a smaller bias than the classical ratio estimator) ratio estimators may be of great advantage. Therefore, in recent years, considerable attention has been given to the development of unbiased and approximately unbiased ratio estimators.

In this paper we shall, without loss of generality, confine ourselves to estimation of ratios (assuming the population mean $\bar{X}$ is known). Further, to simplify the discussion we shall confine ourselves to simple random sampling and assume the population size N is infinite.

If a simple random sample of n pairs (y_i, x_i) is drawn, the classical ratio estimator of $R=\bar{Y}/\bar{X}$ is given by

$$r = \frac{\bar{y}}{\bar{x}} \tag{1}$$

where $\bar{y}$ and $\bar{x}$ are the sample means of y_i and x_i respectively. The usual estimator of the variance of r is:

$$v(r) = \frac{1}{n\bar{X}^2}\left(s_y^2 - 2rs_{xy} + r^2 s_x^2\right) \tag{2}$$

where s_x^2 and s_y^2 are the sample mean squares of x_i and y_i respectively and s_{xy} is the sample mean product of x_i and y_i. It is well known that the bias of v(r) is of order 1/n. Kokan (1963) has investigated the large-sample stabilities of v(r) and the unbiased variance estimator:

$$v\left(\frac{\bar{y}}{\bar{X}}\right) = \frac{s_y^2}{\bar{X}^2}, \tag{3}$$

where $\bar{y}/\bar{X}$ is the estimator of R not using the sample x-information. He has shown that the coefficient of variation of v(r) is always larger than that of $v(\bar{y}/\bar{X})$ for a bivariate normal distribution, and this property also holds for a bivariate log normal distribution for certain ranges of the parameters.

Hartley and Ross (1954) were the first to give an exact upper bound for the bias of r and the unbiased ratio estimator:

$$t_1 = \bar{r} + \frac{n}{(n-1)\bar{X}}(\bar{y} - \bar{r}\,\bar{x}) \tag{4}$$

where $\bar{r}$ is the sample mean of $r_i = y_i/x_i$. Goodman and Hartley (1958) have shown that the variance of t_1 will often be larger than the variance of r, for large n.

We now consider two ratio estimators based on dividing the sample at random into g groups, each of size m, where n = mg. Following Mickey (1959), an unbiased estimator of R is given by

$$t_2 = \bar{r}_g + \frac{g}{\bar{X}}(\bar{y} - \bar{r}_g\bar{x}) \tag{5}$$

where $\bar{r}_g = \sum_1^g r_j'/g$ and r_j' is the classical ratio estimator computed from the sample after omitting the j-th group, i.e., $r_j' = (n\bar{y}-m\bar{y}_j)/(n\bar{x}-m\bar{x}_j)$ where $\bar{y}_j$ and $\bar{x}_j$ are the sample means computed from the j-th group. It may be noted that t_2 reduces to t_1 for the important case of n=2.

Quenouille (1956) proposed a method of reducing estimation bias from order 1/n to $1/n^2$, based on random division of the sample into groups. Durbin (1959) applied this method to ratio estimators and has shown that the estimator:

$$t_3 = gr - (g-1)\bar{r}_g = \frac{1}{g}\sum_1^g r_{Qj} \tag{6}$$

where $r_{Qj} = gr - (g-1)r_j'$ (called pseudo-values by by Tukey), has bias of order n^{-2} at most. Durbin has further shown that if the regression of y on x is linear (i.e., $y = \alpha + \beta x + e$ where e is the error term) and x normally distributed (model 1, say), t_3 with g = 2 has a smaller asymptotic variance than r. J. Rao (1965) has shown, for the above model, that <u>both</u> asymptotic bias and asymptotic variance of t_3 are decreasing functions of g, so that g=n would be the optimum choice for large or moderately large n. Durbin (1959) has also considered the case where the regression of y on x is linear, but x has a gamma distribution (model 2, say). He has shown that, although the variance t_3 with g=2 compared to the variance of r is slightly increased, the reduction in bias is such that the mean square error of t_3 is reduced. Recently, J. Rao and Webster (1966) have shown, assuming model 2, that <u>both</u> bias and variance t_3 are decreasing functions of g, so that g=n would be the optimum choice. They have also shown that t_3 has a smaller variance than r for g > 2. It may be noted that the results for model 2 are exact for any sample size, n.

Following Tukey (1958) we can use the simple estimator:

$$v(t_3) = g^{-1}(g-1)^{-1}\sum_1^g(r_{Qj} - t_3)^2 \qquad (7)$$

as the variance estimator of t_3, since the g estimators r_{Qj} may be treated as approximately independent and $t_3 = \Sigma\, r_{Qj}/g$. Lauh and Williams (1963) have made a Monte Carlo study of the stabilities of v(r) and $v(t_3)$ with g=n, for small samples (n=2 to 9), using model 1 and model 2 (with exponential distribution for x) and assuming α=0 (i.e., regression through the origin). It was shown that the Monte Carlo variances of v(r) and $v(t_3)$ are about the same for model 1, whereas, for model 2, the variance of $v(t_3)$ is considerably smaller than that of v(r). [3]Incidentally, Lauh and Williams have used $v(t_3)$ as the estimator of the variance of r rather than of the variance of t_3.

Tin (1965) investigated the large-sample bias, variance and approach to normality of the following estimators: r, t_3 with g=2, Beale's estimator

$$t_4 = r\,\frac{\left(1 + \frac{1}{n}\frac{s_{xy}}{\bar{x}\,\bar{y}}\right)}{\left(1 + \frac{1}{n}\frac{s_x^2}{\bar{x}^2}\right)} \qquad (8)$$

and the modified estimator

$$t_5 = r\left[1 + \frac{1}{n}\left(\frac{s_{xy}}{\bar{x}\,\bar{y}} - \frac{s_x^2}{\bar{x}^2}\right)\right] \qquad (9)$$

His comparison shows that t_5 is slightly better than t_4 which in turn is better than t_3 with g=2. Tin has also made a Monte Carlo study for large and moderately large samples (n=50, 200, and 1000), using model 1 and the results are in agreement with his mathematical results. J. Rao and Webster (1966) made an exact comparison of t_5 and t_3 assuming model 2. Their comparison shows that the precisions of t_5 and t_3 with g=n are about the same.

Tin (1965) has also considered the estimator:

$$t_6 = \frac{g}{g-1}\,r - \frac{1}{g(g-1)}\sum_1^g \frac{\bar{y}_j}{\bar{x}_j}\,. \qquad (10)$$

It may be noted that t_6 is identical to t_3 when g=2. Also, t_6 is based on the group means $\bar{y}_j$ and $\bar{x}_j$, whereas, t_3 is based on the complementary means $(n\bar{y} - m\bar{y}_j)/(n-m)$ and $(n\bar{x} - m\bar{x}_j)/(n-m)$. Hartley (see Pascual, 1961 and Sastry, 1965) has earlier proposed t_6 with g=n and Murthy and Nanjamma (1959) have used t_6 when g independent and interpenetrating sub-samples each of size m are available.

Pascual (1961) proposed the following estimator obtained by estimating the bias of r approximately:

$$t_7 = r + \frac{1}{(n-1)\bar{X}}\,(\bar{y} - \bar{r}\,\bar{x})\ . \qquad (11)$$

It may be noted that the investigations by Pascual (1961) and Sastry (1965) regarding t_6 (with g=n) and t_7 are not very satisfactory, since they assume the higher order population moments of $\delta x_i = (x_i - \bar{X})/\bar{X}$ and $\delta y_i = (y_i - \bar{Y})/\bar{Y}$ are negligible and $|\delta x_i| < 1$, whereas, to develop asymptotic theories for r, t_2, t_3, t_4 and t_5 we need only assume that n is large or moderately large and $|\bar{x}-\bar{X}|/\bar{X}<1$ or $|\bar{x}_j'-\bar{X}|/\bar{X}<1$ where $\bar{x}_j' = (n\bar{x}-m\bar{x}_j)/(n-m)$. Similarly, for the estimator t_6 ($g \neq n$) an asymptotic theory would not be satisfactory if m is small.

It is clear that efficiency comparisons in small or moderate size samples would be more valuable since these are the cases in which freedom from bias may be important. Therefore, in the present paper, we make a Monte Carlo study of the efficiencies of the eight ratio estimators r, t_1, ..., t_7 and the three variance estimators $v(\bar{y}/\bar{X})$, v(r) and $v(t_3)$ for small and moderate size samples, using model 1. It may be noted that, unlike under model 1, exact analytical comparisons of the estimators can be made under model 2 for any sample size -- some exact results, under model 2, have already been given by J. Rao and Webster

(1966) and further work is in progress.

We have used Lauh and Williams' model as well as that of Tin for our study. The model of Lauh and Williams is ideal for ratio estimators since it is assumed that the regression is through the origin and the coefficient of variation of x, C_x, is small. On the other hand, Tin's model is not so favorable since the regression is not through the origin and C_x is not small. It would be interesting, therefore, to study the performances of the estimators under both the models.

2. Monte Carlo Study

In Lauh and Williams' model, x_i is N(10, 4) and y_i is defined as $5(x_i + e_i)$ where e_i is N(0, 1) and independent of x_i. Therefore, the correlation, ρ, between y_i and x_i is 0.89 and $C_x = 0.2$. In Tin's model, x_i and y_i have a bivariate normal distribution with $\overline{X} = 5$, $\sigma_x^2 = 45$, $\overline{Y} = 15$, $\sigma_y^2 = 500$ and $\rho = 0.4$, 0.6 or 0.8. To carry out the experiment on the computer, the model $y_i = \alpha + \beta x_i + e_i$ was used, where $\alpha = \overline{Y} - \beta\overline{X}$, $\beta = \rho\sigma_y/\sigma_x$, and e_i is $N[0, \sigma_y^2(1-\rho^2)]$ and distributed independently of x_i.

Using the IBM 7094 pairs of random numbers (u_{1i}, u_{2i}) were generated from a rectangular distribution with mean 1/2 and range 1 and were transfromed into standard normal variates w_{1i} and by the transformation:

$$w_{1i} = (-2 \log_e u_{1i})^{1/2} \sin 2\pi u_{2i}$$

$$w_{2i} = (-2 \log_e u_{1i})^{1/2} \cos 2\pi u_{2i}$$

(see Box and Muller, 1958). Then the pairs (x_i, y_i) were computed as follows: For Lauh and Williams' model, $x_i = \overline{X} + w_{1i}\sigma_x$ and $y_i = 5(x_i + w_{2i})$; for Tin's model, $x_i = \overline{X} + w_{1i}\sigma_x$ and $y_i = \alpha + \beta x_i + e_i$ where $e_i = w_{2i}\sigma_y(1-\rho^2)^{1/2}$. For each selected n, 1000 samples of n pairs (x_i, y_i) were generated and the eight ratio estimators and the three variance estimators were computed from each sample. Thus we have 1000 values of each estimator for each selected n.

The ratio estimators all have Cauchy distributions when the distribution of (x_i, y_i) is bivariate normal so that the population moments do not exist. It may be meaningless, therefore, to use the variance (or mean square error) of the 1000 values as a sample criterion in comparing the estimators. However, the variance would be a satisfactory criterion for Lauh and Williams' model since $\overline{X}$ is so large compared with σ_x that the range of x is effectively positive. On the other hand, for Tin's model the probability of getting a negative x-value is not negligible and the criterion of variance can lead to meaningless results, particularly for small n. Tin's results based on the variance may, however, be fairly satisfactory since he considered only large or moderately large n(n = 50, 200 and 1000). Therefore, we have used, for Tin's model, the following criteria for which expected values exist, and employed the variance criterion for Lauh and Williams' model:

(1) Concentration. Proportion of the value of an estimator in some *a priori* neighborhood around the population ratio R. In this study, this interval is chosen as (R - 0.1, R + 0.1). An estimator T is more efficient than another estimator S if its concentration is larger than that of S.

(2) Interquartile range. Distance between the upper and lower quartile points. Thus it is a range which contains one-half of the 1000 values of an estimator. An estimator T is more efficient than S if its interquartile range is smaller than that of S.

2.1. Results for Lauh and Williams' Model

The variances of the eight ratio estimators r, $t_1, \ldots, t_7$ (obtained from 1000 samples) are given in Table 1 for n=4, 6, 8 and 12. It may be noted that all the estimators are unbiased under this model. It is clear, from Table 1, that there are very little differences in the variances of the eight estimators, even for small n, and hence it does not matter which ratio estimator is used. We may, however, still make the following observations: (1) the optimum number of groups, g, for the estimators t_2, t_3 and t_6 is n, (2) for n>4, there are virtually no differences in the variances of t_1, t_2 (g=n), t_3(g=n), t_4, t_5, t_6(g=n) and t_7; for n=4, t_1 has a slightly larger variance, (3) the variances of $t_1, \ldots, t_7$ are slightly smaller than that of r for n<12.

Turning to the three variance estimators $v(\overline{y}/\overline{X})$, v(r) and $v(t_3)$, Table 2 gives the coefficients of variation of these estimators for n=4, 6, 8 and 12 -- the criterion of coefficient of variation is more appropriate here since the variance estimators do not have the same mean.

The following tentative conclusions may be drawn from Table 2: (1) Coefficient of variation of $v(t_3)$ decreases considerably as g increases. (2) For any n, the coefficients of variation of v(r) and $v(t_3)$ with g=n are essentially equal. (3) Coefficient of variation of v(r) is slightly larger than that of $v(\overline{y}/\overline{X})$ (for n>4) -- this is in agreement with Kokan's (1963) asymptotic result that the increase in the coefficient of variation of v(r) over that of $v(\overline{y}/\overline{X})$ would be small if C_x is small. We may conclude that the variance estimators v(r) and $v(t_3)$ with g=n are quite stable, for any n, compared to $v(\overline{y}/\overline{X})$, if the regression is approximately through the origin and C_x is small.

Table 1. **Variances** of the eight ratio estimators r, $t_1, \ldots, t_7$ (obtained from 1000 samples) for selected values of n (Lauh and Williams' model)

Estimator \ n		4	6	8	12
Classical: r		0.0665	0.0398	0.0322	0.0219
Hartley-Ross: t_1		0.0661	0.0394	0.0318	0.0218
Mickey: t_2	$g = 2$		0.0397	0.0321	0.0218
	$g = \frac{n}{2}$		0.0397	0.0319	0.0218
	$g = n$	0.0659	0.0394	0.0318	0.0218
Quenouille: t_3	$g = 2$		0.0397	0.0321	0.0218
	$g = \frac{n}{2}$		0.0397	0.0318	0.0218
	$g = n$	0.0658	0.0394	0.0318	0.0218
Beale: t_4		0.0659	0.0394	0.0318	0.0218
Modified: t_5		0.0659	0.0394	0.0318	0.0218
Tin: t_6	$g = 2$		0.0397	0.0321	0.0218
	$g = \frac{n}{2}$		0.0397	0.0318	0.0218
Hartley:	$g = n$	0.0659	0.0394	0.0317	0.0218
Pascual: t_7		0.0659	0.0394	0.0317	0.0218

Table 2

Coefficients of Variation of the Variance Estimators $v(\bar{y}/\bar{X})$, $v(r)$ and $v(t_3)$ for Selected Values of n (Lauh and Williams' Model)

Variance Estimator \ n		4	6	8	12
$v(\bar{y}/\bar{X})$		0.85	0.61	0.52	0.42
$v(r)$		0.86	0.67	0.55	0.43
$v(t_3)$:	$g = 2$		1.45	1.38	1.34
	$g = \frac{n}{2}$		1.03	0.81	0.64
	$g = n$	0.87	0.68	0.54	0.44

2.2 Results for Tin's Model

The interquartile ranges (for n=4, 6, 10, 20 and 50) and the concentrations (for n=10, 20, and 50) of the eight ratio estimators (obtained from 1000 samples) are given in Table 3 for ρ=0.6, and in Table 4 for ρ=0.8. We have also computed the concentrations for n=4 and 6, but the values are erratic and, hence, unreliable -- this may be probably due to the narrowness of the interval (R-0.1, R+0.1). The variances are also given in Table 3 for ρ=0.6 and n=10 to show that the criterion of variance may lead to meaningless results for small samples. The criterion of variance may, however, become satifactory with large or moderately large n, excepting for those estimators based on the individual ratios y_i/x_i or the group ratios $\bar{y}_j/\bar{x}_j$ for small m. Therefore we have included the variances of r, t_2, t_3, t_4, t_5 and t_6 (g=2) for n=50 --- our values of the variance of r(0.313 with ρ=0.6 and 0.150 with ρ=0.8) are fairly close to those obtained from the usual asymptotic variance formula (0.293 with ρ=0.6 and 0.148 with ρ=0.8), whereas Tin's values (0.376 with ρ=0.6 and 0.238 with ρ=0.8) are markedly different.

The following tentative conclusions may be drawn from Tables 3 and 4: (1) Unlike under Lauh and Williams' model, the differences in the efficiences of the estimators are quite significant. (2) The interquartile ranges (for all n) and the variances (for n=50) of t_2 and t_3 appear to decrease as g increases, but, for n > 10, the values for g = n/2 and g=n are essentially equal. The concentrations of t_2 and t_3 seem to increase as g increases, excepting that in one case (n=10, ρ=0.6) the concentration of t_2 is 7.1% for g=2, 7.3% for g = n/2 and 7.0% for g=n. In any case, the combined evidence of the three criteria seems to indicate that the optimum value of g for t_2 and t_3 may be taken as n. Moreover, with g=n there is no random splitting involved and the ratio $\bar{r}_g$ is simply given by $n^{-1} \sum^{n} (n\bar{y} - y_j)/(n\bar{x} - x_j)$. (3) Efficiency of t_6^1 is maximum at g=2 unlike under Lauh and Williams' model. This may be because t_6 is biased under Tin's model and as Tin pointed out, the bias increases as g increases. (4) All three criteria indicate that the efficiencies of t_2(g=n), t_3(g=n), t_4 and t_5 are about the same for n $\geq$ 10; for n=4 and 6, t_4 appears to be more efficient than the others. All the above four estimators are more efficient than r, expecting that in one case (n=50, ρ=0.8), the concentrations are essentially equal. (5) r is more efficient than t_6(g=2) for n=4 and 6; for n $\geq$ 10 the efficiencies are about the same, excepting that in one case (n=10, ρ=0.6) the concentration of t_6(g=2) is somewhat low. (6) r is more efficient than Hartley's estimator t_6 (g=n), Pascual's estimator t_7 and the Hartley-Ross unbiased estimator t_1. The efficiencies of t_6(g=n), t_7 and t_1 are about the same for n=4 and 6; for n $\geq$ 10, t_1 is less efficient than t_6(g=n) and t_7.

Turning to the variance estimators, the interquartile ranges (for n=10, 20 and 50) and coefficients of variation (for n=50) of $v(\bar{y}/\bar{X})$, v(r) and $v(t_3)$ are given in Table 5. We have not computed the concentrations around the variances because, as pointed out earlier, the Monte Carlo variances of r and t_3 would be meaningless for small samples. First, considering the criterion of coefficient of variation (for n=50) it is clear from Table 5 that the coefficient of variation of $v(t_3)$ decreases considerably as g increases, so that the optimum value of g is n. The coefficients of variation of $v(t_3)$ (with g=n) and v(r) are essentially equal, but both are considerably larger than the coefficient of variation of $v(\bar{y}/\bar{X})$-- this is in agreement with Kokan's (1963) result because C_x is not small. Turning to the criterion of interquartile range, it may be noted that, in comparing the variance estimators, it would be more appropriate to take the interquartile ranges given in Table 5 relative to the interquartile ranges of the corresponding estimators of R. Now since the interquartile range of t_3 (for any n) is smaller than that of r which in turn is considerably smaller (for ρ=0.8) than that of $\bar{y}/\bar{X}$, it follows from Table 5 that v(r) is slightly more efficient than $v(t_3)$ with g=n (particularly for n=10), but $v(\bar{y}/\bar{X})$ is considerably more efficient than v(r) and $v(t_3)$ with g=n.

3. Concluding Remarks

The approximately unbiased ratio estimators cannot be expected to help when very small samples are taken within strata. In such situations, Mickey's unbiased estimator (with g=n) may be promising since it behaves well under ideal conditions and, under non-ideal conditions, it is considerably more efficient than the Hartley-Ross unbiased estimator and also slightly more efficient than the classical estimator. However, for the important case of n=2 in each stratum, Mickey's estimator is identical to the Hartley-Ross estimator so that no improvement can be achieved. If an approximately unbiased estimator serves the purpose, then Quenouille's estimator (with g=n), Beale's estimator and the modified estimator look favorable, and it does not matter which one is used (although Beale's estimator looks slightly more efficient for n=4 and 6). However, Quenouille's estimator has the added advantage of having a simple variance estimator (Tukey's variance estimator). The approximately unbiased estimators of Hartley and Pascual appear to be unsatisfactory compared to the above approximately unbiased estimators.

Since both Tukey's variance estimator and the usual variance estimator of the classical ratio estimator are considerably less efficient, under non-ideal conditions, than the variance estimator

TABLE 3

Interquartile ranges (I.R.), concentrations (C) and variances (V) of the eight ratio estimators $r, t_1, .., t_7$ (obtained from 1000 samples) for selected values of n and $\rho = 0.6$ (Tin's model)

Estimator		n = 4	n = 6	n = 10			n = 20		n = 50		
		I.R.	I.R.	I.R.	% C	V	I.R.	% C	I.R.	% C	V
Classical: r		2.8	2.3	1.8	6.5	51	1.07	9.1	0.72	17.1	0.31
Hartley-Ross: t_1		4.0	3.2	2.5	4.5	1988	1.63	6.8	1.05	12.8	
Mickey: t_2	$g = 2$	3.7	2.5	1.9	7.1	90	1.09	9.0	0.71	16.3	0.31
	$g = \frac{n}{2}$	3.7	2.3	1.5	7.3	466	0.96	10.7	0.68	18.5	0.29
	$g = n$	2.6	1.7	1.4	7.0	637	0.96	10.7	0.68	18.9	0.28
Quenouille: t_3	$g = 2$	3.5	2.6	1.9	5.0	404	1.06	9.1	0.71	16.4	0.31
	$g = \frac{n}{2}$	3.5	2.4	1.5	6.1	6407	0.93	9.5	0.67	18.6	0.28
	$g = n$	2.7	1.9	1.4	7.7	11508	0.93	11.4	0.67	19.2	0.28
Beale: t_4		1.7	1.6	1.4	6.8	1	0.96	10.9	0.68	19.2	0.28
Modified: t_5		2.1	1.7	1.4	6.7	66544792	0.94	11.6	0.68	19.1	0.28
	$g = 2$	3.5	2.6	1.9	5.0	404	1.06	9.1	0.71	16.4	0.31
Tin: t_6	$g = \frac{n}{2}$	3.5	3.1	2.3	5.0	2494	1.29	8.5	0.78	14.6	
Hartley:	$g = n$	4.0	3.0	2.2	4.9	119	1.24	7.6	0.77	15.3	
Pascual: t_7		4.0	3.1	2.2	5.7	166	1.26	7.2	0.77	14.9	

of the estimator which does not use the supplementary information, caution is needed in the indiscriminate use of ratio estimators.

It may be noted that the main reason for using Mickey's unbiased estimator or an approximately unbiased estimator over the classical estimator is to eliminate or reduce bias in situations where freedom from bias is important; however, it is gratifying that these estimators may, in fact, lead to small or moderate gains in efficiency over the classical estimator.

The conclusions from this study are not necessarily applicable to all populations, since we have considered only two particular models; however, these models reflect many situations that are encountered in practice. Also, our study is of necessity empirical in nature and not mathematical. Clearly therefore, further work, mathematical as well as empirical, using other models and actual data is needed. To this end, we are at the present time investigating the following problems:

(1) Asymptotic results for Mickey's estimator along the lines of J. Rao (1965) using Durbin's model 1.
(2) Exact mathematical results for the eight ratio estimators and the three variance estimators considered in this paper, using Durbin's model 2.
(3) Mathematical and Monte Carlo results when x has a log normal distribution.
(4) Empirical results using several sets of real data.
(5) Results for other models.

References

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates, Ann. Math. Statist., 29, 610-1.

Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios, Biometrika, 46, 477-80.

Goodman, L. A. and Hartley, H. O. (1958). The precision of unbiased ratio-type estimators, J. Amer. Statist. Assoc., 53-491-508.

Hartley, H. O. and Ross, A. (1954). Unbiased ratio estimators. Nature, 174, 270-1.

TABLE 4

Interquartile ranges (I.R.), concentrations (C) and variances (V) of the eight ratio estimators $r, t_1, .., t_7$ (obtained from 1000 samples) for selected values of n and $\rho = 0.8$ (Tin's model)

Estimator		n = 4	n = 6	n = 10		n = 20		n = 50		
		I.R.	I.R.	I.R.	% C	I.R.	% C	I.R.	% C	V
Classical: r		1.9	1.6	1.3	7.7	0.81	15.4	0.50	23	0.15
Hartley-Ross: t_1		2.8	2.2	1.9	5.6	1.20	9.2	0.73	15	
Mickey: t_2	$g = 2$	2.3	1.8	1.3	8.0	0.80	15.4	0.51	22	0.15
	$g = \frac{n}{2}$	2.3	1.6	1.1	9.1	0.72	16.0	0.49	23	0.14
	$g = n$	1.9	1.2	1.0	10.5	0.72	16.0	0.49	23	0.14
Quenouille: t_3	$g = 2$	2.5	1.9	1.3	8.3	0.81	14.8	0.51	22	0.15
	$g = \frac{n}{2}$	2.5	1.7	1.1	8.5	0.70	16.0	0.49	23	0.14
	$g = n$	2.0	1.4	1.0	10.0	0.70	16.2	0.48	23	0.14
Beale: t_4		1.3	1.2	1.0	9.9	0.71	16.2	0.48	23	0.14
Modified: t_5		1.7	1.3	1.0	10.3	0.69	16.0	0.48	23	0.14
Tin: t_6	$g = 2$	2.5	1.9	1.3	8.3	0.81	14.8	0.51	22	0.15
	$g = \frac{n}{2}$	2.5	2.1	1.5	7.7	0.93	11.1	0.56	22	
Hartley:	$g = n$	2.9	2.1	1.5	6.1	0.88	12.7	0.55	20	
Pascual: t_7		2.8	2.2	1.6	7.1	0.89	13.3	0.55	20	

Table 5

Interquartile ranges and coefficients of variation of the variance estimators $v(\bar{y}/\bar{X})$, $v(r)$ and $v(t_3)$ for selected values of n and ρ (Tin's model)

Variance Estimator		$\rho = 0.6$				$\rho = 0.8$			
		Interquartile range			Coeff. of Variation	Interquartile range			Coeff. of Variation
		n=10	n=20	n=50	n=50	n=10	n=20	n=50	n=50
$v(\bar{y}/\bar{X})$		1.3	0.43	0.11	0.20	1.21	0.42	0.10	0.20
$v(r)$		3.1	0.97	0.20	0.61	1.60	0.45	0.09	0.51
$v(t_3)$:	$g=2$	4.2	1.55	0.47	2.21	2.26	0.81	0.22	2.22
	$g=\frac{n}{2}$	4.4	1.08	0.22	0.66	2.24	0.52	0.10	0.58
	$g=n$	3.7	0.97	0.20	0.61	1.88	0.46	0.10	0.51

Kokan, A. R. (1963). A note on the stability of the estimates of standard errors of the ordinary mean estimate and the ratio estimate, Cal. Statist. Assoc. Bull., 12, 149-58.

Lauh, E. and Williams, W. H. (1963). Some small sample results for the variance of a ratio, Proc. Amer. Statist. Assoc., (Social Statistics Section), 273-83.

Mickey, M. R. (1959). Some finite population unbiased ratio and regression estimators, J. Amer. Statist. Assoc., 54, 594-612.

Murthy, M. N. and Nanjamma, N. S. (1959). Almost unbiased ratio estimators based on interpenetrating subsamples, Sankhya, 21, 299-314.

Pascual, J. N. (1961). Unbiased ratio estimators in stratified sampling, J. Amer. Statis. Assoc., 56, 70-87.

Quenouille, M. H. (1956). Notes on bias in estimation, Biometrika, 43, 353-60.

Rao, J. N. K. (1965). A note on the estimation of ratios by Quenouille's method, Biometrika, 52, 647-9.

Rao, J. N. K. and Webster, J. T. (1966). On two methods of bias reduction in the estimation of ratios, Biometrika, 53 (in press).

Sastry, K. V. R. (1965). Unbiased ratio estimators, J. Ind. Soc. Agric. Statist., 17, 19-29.

Tin, M. (1965). Comparison of some ratio estimators, J. Amer. Statist. Assoc., 60, 294-307.

Tukey, J. W. (1958). Bias and confidence in not-quite large samples, Ann. Math. Statist. (Abstract), 29, 614.

This rearch was supported by the National Science Foundation, under Grant GP-6400.