RATIO-AND REGRESSION-TYPE ESTIMATORS EMPLOYING A PRIORI KNOWLEDGE

Paul H. Tomlin, Bureau of the Census

1. Introduction

Ratio- and regression-type sample estimators have often been used instead of the "unbiased" sample mean $\bar{y}$ to estimate the population mean $\bar{Y}$. Both types of estimators make use of a concomitant variable $\bar{x}$ having known mean $\bar{X}$. The simplest ratio-type estimator is given by $\bar{y}_C$, and the simplest regression-type estimator by $\bar{y}_G$. The formulas for these estimates are given below:

$$\bar{y}_C = r_C\bar{X}, \text{ where } r_C = \bar{y}/\bar{x}$$
$$\bar{y}_G = r_G\bar{X}, \text{ where } r_G = \{\bar{X}\}^{-1}\{\bar{y}-\hat{\beta}(\bar{x}-\bar{X})\}$$

Here $\bar{x}$ is the sample mean for the variable x, and $\beta = s_{xy}/s_x^2$, the sample regression coefficient. Note that a vague, "global" form of a priori knowledge is used—the existence of a moderate or high correlation between x and y over the entire population. In such a case $\bar{y}_C$ and $\bar{y}_G$ will have lower mean square errors (MSE) than $\bar{y}$, and so almost all statisticians will choose $\bar{y}_C$ or $\bar{y}_G$, even if biased in place of $\bar{y}$.

The estimators r_C and r_G are biased (with respect to simple random sampling (SRS)[1] unless certain conditions on $E(y|x)$ are met. As a result, a number of statisticians have proposed alternatives to r_C which are designed to reduce the bias over SRS, perhaps to zero. Studies usually, but not always, Monte Carlo) have been performed in which the alternatives have been compared with r_C and each other with respect to MSE - on "real" populations and over a superpopulation model. In the latter situation the usual model is the following:

$$\left.\begin{aligned} E(y|x) &= \alpha + \beta x \\ Var(y|x) &= \varphi(x), \text{ a known function (up to scalar factor)} \end{aligned}\right\} \quad (1)$$

Two of the better papers in this area are [3] and [7]. Most tests under (1) have been performed with $\varphi(x) = \delta x^t$, where $0 \leq t \leq 2$ and $\delta > 0$.[2]

It turns out that the shape of the population—the distribution of x and the parameters α, $\bar{X}$, and $\varphi(x)$ (but not β)—affects the performance of the estimators. This suggests that these parameters be used, or approximated, in the creation of an estimator which is in a class containing r_C, r_G, and the proposed alternatives and has smaller MSE than any of them. This a priori knowledge is more "local" than that of the assumed high correlation. Obtaining such knowledge may involve much work, but if the knowledge is reasonably accurate, great gains in precision of estimation will result. However, a poor choice of φ, or a nonlinear function $E(y|x)$ may result in an estimator less efficient than r_C or r_G. See [11], pp. 45-48.

2. Conditional a priori Mean Square Error (C.a.p. MSE) for Linear Estimators

Let the symbol κ denote "a priori knowledge." In our situation, κ consits of the model (1) - not so much the actual parameters but the general shape of the functions $E(y|x)$ and $Var(y|x)$.

We assume a universe U of size N. Let u be a sample indicator function—an N x 1 vector of integers, u_k representing the number of times universe element k ($1 \leq k \leq N$) appears in the sample. Let $p(\underset{\sim}{u}|D)$ be the probability that $\underset{\sim}{u}$ could be selected via a design D. Let us for simplicity assume $u_k = 0$ or 1. We then define, for a function f defined on samples u.

$$MSE(f|\kappa;D) = \Sigma\, p(u|D) MSE(f|\kappa;u) \quad (2)$$

where $MSE(f|\kappa;u)$—yet to be defined— is that contribution to $MSE(f|\kappa;D)$ associated with sample u. We refer to $MSE(f|\kappa;u)$ as the conditional a priori mean square error of f, given κ and u, and denote it by C.a.p. MSE. We will also use the initials "C.a.p." for other conditional a priori functions such as means and variances.

We now restrict our interest to linear estimators r of the ratio $R = \bar{Y}/\bar{X}$. Our a priori knowledge κ consists of the a priori mean m of the (Nx1) vector y and the (NxN) a priori covariance matrix V. The universe mean $\bar{Y}$ then has an a priori mean $\bar{M} = N^{-1}\Sigma M_k$ and an a priori variance $N^{-2}\underset{\sim}{1}'V\underset{\sim}{1}$. The ratio R also has an a priori mean, $R_o = \bar{M}/\bar{X}$. Note $\bar{M} = N^{-1}(\underset{\sim}{1}'\underset{\sim}{m})$.

The linear estimator r is defined by the vector $\underset{\sim}{a} = \underset{\sim}{a}(\underset{\sim}{u})$:

$$r = \underset{\sim}{a}' \cdot \underset{\sim}{y} \quad (3)$$

where $\tilde{a}_k$ is a function of $\underset{\sim}{u}$ only, not of $\underset{\sim}{y}$, such that $a_k = 0$ whenever $u_k = \tilde{0}$.[1]

Then the conditional a priori mean and variance or r, given κ and $\underset{\sim}{u}$, are given by

$$\left.\begin{aligned} E(r|\kappa;\underset{\sim}{u}) &= a'm \\ Var(r|\kappa;\underset{\sim}{u}) &= a'Va \end{aligned}\right\} \quad (4)$$

We now define C.a.p. $MSE(r|\kappa;\underset{\sim}{u})$ with respect to the a priori ratio R_o:

$$\text{C.a.p. } MSE(r|\kappa;\underset{\sim}{u}) = E[(r-R_o)^2|\kappa;\underset{\sim}{u}] \quad (5)$$

We can decompose (5) into the a priori variance of r and the squared difference between the a priori mean of the sample estimate and the a priori ratio R_o:

$$\left.\begin{aligned} &E[(r-R_o)^2|\kappa;u] \\ &= E[(\underset{\sim}{a}'\underset{\sim}{y} - \underset{\sim}{a}'\underset{\sim}{m} + \underset{\sim}{a}'\underset{\sim}{m} - R_o)^2|\kappa;u] \\ &= \underset{\sim}{a}'V\underset{\sim}{a} + (\underset{\sim}{a}'\underset{\sim}{m} - R_o)^2 \end{aligned}\right\} \quad (6)$$

When averaged over the design D, the two components of (6) are within-sample C.a.p. variance and between-sample C.a.p. MSE, given κ. The first term contains only V, and the second, only m (and $R_o = \bar{M}/\bar{X}$). By a "consistent" estimator (not previously defined) we mean, for finite N, an estimator $\underset{\sim}{a}'\underset{\sim}{y}$ such that $\underset{\sim}{a}'\underset{\sim}{m} = R_o$ when $n=N(\underset{\sim}{u} = \underset{\sim}{1}_N)$. Clearly this is important for a meaningful set of estimators. However in (6) the first term goes to an a priori within-census variance which will not be zero, in general. We can think of κ as a superpopulation structure, and $Var(r|\kappa;\underset{\sim}{1}_N)$ as the variance due to that structure.

From (2) it is clear that we can minimize $MSE(r|\kappa;D)$ by minimizing $MSE(r|\kappa;\underset{\sim}{u})$ for any fixed u. Hence, we reduce the problem from N to n dimentions, n being the number of distinct units in the sample $\underset{\sim}{u}$. Hence, from here on, $\underset{\sim}{m}$ and V are assumed to be related to the n-dimensional subset of nonzero elements of the Nxl vector $\underset{\sim}{u}$. The sample size n may be variable over D, but this need not concern us. However u has been determined, we do the best we can with it. It is the only sample available.

3. Reflexive, Parareflexive, and Hyperreflexive Estimators

We now assume that $\underset{\sim}{a}$ is a function only of $\underset{\sim}{x}$ and not of any other properties of $\underset{\sim}{u}$, as in the following examples:

Let r_U denote $\bar{y}/\bar{X}$, the estimate of R derived from the unbiased sample mean $\bar{y}$, and let a_{Uk}, a_{Ck}, and a_{Gk} denote the k^{th} element of the vectors $\underset{\sim}{a}_U$, $\underset{\sim}{a}_C$, and $\underset{\sim}{a}_G$ as given for estimators r_U, r_C, and r_G, respectively. If the sample size is n, we have

$$a_{Uk} = 1/(n\bar{X}) \quad , \quad a_{Ck} = 1/(n\bar{x}) \quad \text{and } a_{Gk} = \frac{1}{\bar{X}}\left[\frac{1}{n} + \frac{(\bar{X}-\bar{x})(x_k-\bar{x})}{(n-1)\ s_x^{\ 2}}\right] \qquad (7\ a,b,c)$$

where $s_x^{\ 2} = (n-1)^{-1}\Sigma(x_k-\bar{x})^2$.

We are not interested in the set of all linear estimators; we are, however, interested in certain subsets. Consider expressions (3) and (7). For r_U and r_G we have, for all $\underset{\sim}{x}$:

$$\underset{\sim}{a}'\ \underset{\sim}{1} = 1/\bar{X} \qquad (8)$$

whereas for r_C and the proposed ratio-type alternatives in the literature along with $r_G = \bar{y}_G/\bar{X}$, we have the following:

$$\underset{\sim}{a}'\ \underset{\sim}{x} = 1 \qquad (9)$$

for all $\underset{\sim}{x}$. $\bar{y}$ does not satisfy (9), nor does r_C satisfy (8).

We call property (9) the reflexive property, and denote by R the class of reflexive linear estimators. Linear estimators which satisfy (8) will be called parareflexive, and their class will be denoted by P. Estimators in $P \cap R$ will be called hyperreflexive.

Under (1), $\underset{\sim}{m} = \alpha\underset{\sim}{1} + \beta\underset{\sim}{x}$ and $R_o = \alpha/\bar{X} + \beta$, so that formula (6) becomes

$$\text{MSE}(r|\kappa,\ \underset{\sim}{u}) = \underset{\sim}{a}'\ V\underset{\sim}{a} + \alpha^2\ (\underset{\sim}{a}'\underset{\sim}{1} - 1/\bar{X})^2 \qquad (10)$$

if r is reflexive and

$$\text{MSE}(r|\kappa,\ \underset{\sim}{u}) = \underset{\sim}{a}'\ V\underset{\sim}{a} \qquad (11)$$

if r is hyperreflexive.

M. C. Hutchison [3] noted the relationship (9) upon investigation of the properties of six specific ratio-type estimators under (1) with $\alpha = 0$. Observe that if there is some constant R for which $y_k = Rx_k$ for all k, then a reflexive estimator $\underset{\sim}{r}$ will always yield R as the ratio estimate and $\bar{Y}$ as the mean estimate. Thus, there is some intuitive appeal for property (9). Reflexive linear estimators can also be called ratio-type estimators.

Property (8) indicates a kind of unbiasedness—the "weights" $\bar{X} \cdot a_k$ add to 1. Either (9) or (8) is a useful constraint to place upon a linear estimator, as are both (9) and (8) together. Hyperreflexive estimators can also be called regression-type estimators. They are "ξ-unbiased" under (1).[1]

Expression (10) does not contain β, and (11) contains neither α nor β. Thus the a priori knowledge (except for the basic form) need not be specific as (1) indicates, when one restricts the class of linear estimates somewhat. Of course, $\bar{X}$ and V are still quite crucial.

4. Optimization

Let $r^* = (\underset{\sim}{a}^*)'\underset{\sim}{y}$ and $r_H = \underset{\sim}{a}_H'\underset{\sim}{y}$ denote the optimal (with respect to C.a.p. MSE) reflexive and hyperreflexive estimators, respectively, of R. The formulas for $\underset{\sim}{a}^*$ and r^* are:

$$\underset{\sim}{a}^* = \frac{W\underset{\sim}{x} + [\alpha^2\Phi][Q_{x1}W\underset{\sim}{x} - Q_{xx}W\underset{\sim}{1}]}{Q_{xx}} \quad \text{and} \quad r^* = \frac{Q_{xy} + [\alpha^2\Phi][Q_{x1}Q_{xy} - Q_{xx}Q_{1y}]}{Q_{xx}} \qquad (12)$$

where $W = V^{-1}$;

$$Q_{zt} = \underset{\sim}{z}'W\underset{\sim}{t} \text{ for vectors } \underset{\sim}{z},\underset{\sim}{t} \text{ (e.g., } Q_{x1} = \underset{\sim}{x}'W\underset{\sim}{1}) ;\quad \Delta = Q_{xx}Q_{11} - Q_{x1}^{\ 2}; \text{ and } \quad \Phi = (Q_{x1} - Q_{xx}/\bar{X} \div (Q_{xx} + \alpha^2\Delta). \qquad (13)$$

The formulas for $\underset{\sim}{a}_H$ and r_H are

$$\underset{\sim}{a}_H = \frac{(Q_{11}\bar{X} - Q_{x1})W\underset{\sim}{x} + (Q_{xx} - Q_{x1}\bar{W})W\underset{\sim}{1}}{\bar{X}\Delta} \quad \text{and} \quad r_H = \frac{(Q_{11}\bar{X} - Q_{x1})Q_{xy} + (Q_{xx} - Q_{x1}\bar{X})Q_{1y}}{\bar{X}\Delta} \qquad (14)$$

Formulas (12)-(14) are derived by means of the method of Lagrange multipliers, using the appropriate C.a.p. MSE equation—(10) or (11)—with the appropriate set of side conditions - (9) for r* or both (8) and (9) for r_H, respectively.

REMARKS:

1. $\underset{\sim}{a}_H$ can be shown to be the limit of $\underset{\sim}{a}^*$ as $\alpha^2 \to \infty$.[3]
2. The estimator r_H is due to A. A. Hasel (1942) [2]. It is invariant when the matrix V is multiplied by a constant.
3. There can be developed optimal estimators $\hat{\alpha}$ and $\hat{\beta}$ satisfying certain conditions analogous to (8) and (9). The generalized Gauss-Markov theorem shows, under (1) that

$$\hat{\alpha} = (Q_{xx}Q_{1y} - Q_{x1}Q_{xy})/\Delta \quad \text{and} \quad \hat{\beta} = (Q_{11}Q_{xy} - Q_{x1}Q_{1y})/\Delta. \quad \text{Note } r_H = \hat{\alpha}/\bar{X}+\hat{\beta}. \qquad (14a)$$

4. If $V = \sigma^2 I$, then $r_H = r_G$.
5. If V is diagonal, with $\varphi(x) = d \cdot x$ for some constant d, and if $\alpha = 0$, then $r^* = r_C$.

(well-known results)

The special case 5 is interesting for two reasons: a) The resulting estimator $r_C = Q_{xy}/Q_{xx}$ goes to R as $n \to N$ even if in fact $\alpha \neq 0$, whereas for other functions $\varphi(x)$, Q_{xy}/Q_{xx} does not converge to R when $\alpha = 0$. b) The situation can occur very frequently in practice. For example, let x be an integer-valued variable denoting the number of units of interest (e.g., persons) in a sample unit (e.g., housing unit) and let y denote some aggregate with respect to the units of interest. If the y's for units of interest are independent and identically distributed, then $E(y|x = s) = s \cdot \beta$, where $\beta = E(y|x = 1)$, and $\text{Var}(y|x = s) = s \cdot \sigma^2$ where $\sigma^2 = \text{Var}(y|x = 1)$, and $r^* = r_C$ indeed.

5. Comparisons of Reflexive Estimators

We compare r* and r_H with the estimators r_C, r_G, and r_Q (due to Quenouille) to note the improvements in MSE of the former estimators over their more classical counterparts for several sample sizes. We use formula (10), useful in practice as well as theory. The formula for r_Q is:

$$\left.\begin{aligned} r_Q &= n\cdot r_C - (n-1)r_D \\ \text{where } r_D &= \frac{1}{n}\sum_{k=1}^{n}\frac{n\bar{y} - y_k}{n\bar{x} - x_k}\ . \end{aligned}\right\} \qquad (15)$$

The values given in Table 1 are estimates of $MSE(r|D)$ for each of the five estimators, where D = SRS. The x-distributions are discrete with finitely - many (s, say) x-values but we take $N = \infty$. All populations are assumed to satisfy (1), and the parameter α and the function $\varphi(x)=Var(y|x)$ are provided, thus making V a diagonal matrix.

The x-distributions for all populations except D contain x=0; thus there is a positive, though small, probability that a sample type will contain all zeros (in which case no estimator will be mathematically defined) or only one nonzero (in which case r_Q is still undefined.

For population A, x is distributed as Binomial (ν,p) with $\nu = 12$ and $p = .5$ the conditional distribution of y given x has $\alpha = 100$, and $\varphi(x) = .0001 + 1000x^2$.

Populations B-F are defined in the following charts. The same random start was used in these populations in order to better compare the results between them.

Populations B-F:
Probability Distribution for x Values

x-Set	.15	.35	.25	.15	.05	.025	.015	.01
1	0	1	2	3	4	6	10	20
2	0	1	2	3	4	6	10	70
3	2	3	4	5	6	8	12	22

Populations B-F: Parameters

Population Name	x-Set	α	$\varphi(x)$
B	1	1	$1 + x^2$
C	2	1	$1 + x^2$
D	3	1	$1 + (x-2)^2$
E	1	0.3	$1 + x^2$
F	1	1	$1 + x$

1. MEAN SQUARE ERROR OF RATIO ESTIMATORS UNDER SRS ($N = \infty$)

Population	Sample Size	No. of Monte Carlo Samples	$MSE(r_C)$	$MSE(r_Q)$	$MSE(r_G)$	$MSE(r^*)$	$MSE(r_H)$
A	2	EXACT	542.0	611.4	2059.	524.9	2059.
	4	300	273.5	285.0	476.1	261.3	466.8
	8	150	137.3	139.8	152.7	130.5	147.4
	16	75	68.73	69.30	70.59	65.20	68.26
	32	37	34.45	34.59	34.58	32.75	33.56
	64	18	17.28	17.32	17.11	16.43	16.65
	128	9	8.64	8.723	8.502	8.214	8.260
B	2	EXACT	1.397	1.052	1.813	.9907	1.813
	4	EXACT	.7283	.7519	.6240	.4743	.5689
	8	400	.3399	.4067	.2843	.2250	.2366
	16	200	.1708	.2133	.1465	.1087	.1100
	32	100	.08829	.1022	.07712	.05323	.05343
	64	50	.04622	.04991	.04488	.02648	.02652
	128	25	.02343	.02430	.02435	.01319	.01320
C	2	EXACT	1.459	1.071	1.507	1.017	1.507
	4	EXACT	.7806	.8083	.6895	.4912	.6252
	8	400	.3982	.5654	.3376	.2365	.2634
	16	200	.2501	.4643	.1866	.1152	.1195
	32	100	.1788	.3340	.1150	.05667	.05750
	64	50	.1214	.1946	.08868	.02827	.02846
	128	25	.07229	.09480	.07410	.01409	.01414
D	2	EXACT	.1936	.2520	.4533	.1557	.4533
	4	EXACT	.1148	.1497	.1560	.07789	.1422
	8	400	.06656	.08485	.07109	.03983	.05916
	16	200	.03774	.04497	.03663	.02097	.02750
	32	100	.02018	.02233	.01928	.01114	.01336
	64	50	.01044	.01101	.01122	.00591	.00663
	128	25	.00532	.00546	.00609	.00309	.00330
E	2	EXACT	1.089	.9635	1.813	.9320	1.813
	4	EXACT	.5412	.6473	.6240	.4515	.5689
	8	400	.2722	.3462	.2843	.2165	.2366
	16	200	.1470	.1858	.1465	.1057	.1100
	32	100	.07888	.09191	.07712	.05221	.05343
	64	50	.04133	.04512	.04488	.02604	.02652
	128	25	.02127	.02227	.02435	.01302	.01320
F	2	EXACT	1.089	.5931*	.9827	.7023	.9827
	4	EXACT	.4202	.3915	.3292	.2791	.3228
	8	400	.1735	.1676	.1300	.1180	.1256
	16	200	.08132	.08031	.05670	.05390	.05487
	32	100	.03807	.03756	.02618	.02529	.02550
	64	50	.01981	.01957	.01265	.01233	.01239
	128	25	.009024	.008910	.006110	.006024	.006037

*r_Q is not really better than r^* on F for samples of size 2; it just appears that way because r_Q is defined on fewer samples than either r_C or r^*, and its average MSE over those samples is less than that of r^* on its larger set of samples.

For populations A, G, and H, the components of C.a.p. MSE($r|\kappa\cdot$SRS) were computed for all samples of size 2 and averaged appropriately. Samples of size 4 were selected by a Monte Carlo procedure. Two successive samples of size 4 were combined into one sample of size 8, and so on to size 128. For populations B through F, the exact values (subject only to rounding) of MSE($r|\kappa$,SRS) were computed for both n=2 and n=4, and the Monte Carlo procedure began at n=8. Results for sample sizes 2^k, k = 1 to 7 are presented.[4]

The MSE's are conditional given that the estimators in question are defined. The estimators r_G and r_H are undefined if all x's are equal.

Quadratic functions $\varphi(x)$ were chosen because the reductions in MSE were sizable; linear functions such as $\varphi(x) = c + dx$ or functions such as $\varphi(x) = c + \delta x^t$, $t \doteq 0$ or $t \doteq 1$, would have shown little or no reduction in MSE. In practice, the variation of $\varphi(x)$ may be highly irregular (populations G and H, table 2).

Although there is sampling error in the estimator of MSE($r|\kappa$,D), the following assertions can be made with the usual caveats related to small numbers of Monte Carlo samples for $n \geq 4$ ($n \geq 8$):

(1) For populations A and F, little is gained by using r^* or r_H in place of r_G. For population F, the reason is that $\varphi(x) = 1 + x$ -- a situation in which r_H is approximately the same as r_G. For population A, the reason is that the x-distribution has a low relvariance - .0833 - and is symmetric and approximately normal rather than skewed, with mean far away from the origin.

(2) Comparing populations B and C, we find that increasing greatly the maximum value of x adversely affects all the classical estimators under SRS but does not affect r^* or r_H much at all. This is the clearest indication that under SRS use of a priori knowledge can improve the estimation procedure even when the distribution of x is unusual.

For population C, MSE($r^*|\kappa$,D = SRS, size n) and MSE($r_H|\kappa$,D = SRS, size n) are approximately proportional to 1/n, whereas for r_C, r_G the proportionality is probably more like $1/\sqrt{n}$. This may be of interest to Taylor approximation advocates.

(3) Translating the x-distribution to the right appears to help all estimators the same amount (compare populations B and D); moving α toward zero appears to help r_C and r_Q more than r^* and leaves r_G and r_H unchanged, since they do not depend on α (compare populations B and E).

(4) MSE(r_H) does indeed approach MSE(r^*) asymptotically, as stated in remark 1. However, MSE($r_H|n = 2$) = MSE($r_G|n = 2$), which suggests that $r_H = r_G$ when n = 2. This, in fact, can be verified when V is a diagonal matrix. (See [11], p. 16)

(5) r_Q is poorer than r_C for populations A to E but better than r_C for population F. This is because of the shapes of the conditional variance functions; only for population F does $\varphi(x)$ grow slowly enough for an estimator like r_Q to be superior to r_C. (See [11], pp. 45-48.)

6. Populations from the "Real World"

Table 3 contains computations for the two "real life" situations: two distributions from economic data described in Table 1. The population units are firms in a particular kind of business, and, for both populations, x is the number of employees in the firm. In population G, y is the value of taxes paid, and in H, y is the value of payroll. The linearity was approximately true in the original population so for populations G and H linearity is assumed using the true regression coefficients. The functions $\varphi_G(x)$ and $\varphi_H(x)$ denote the conditional variance functions of y, given x, and the regression coefficients α for y on x are -3.177 and 12.06 for populations G and H, respectively.

2. BUSINESS DISTRIBUTIONS

x	Pr[x]	$\varphi_G(x)$	$\varphi_H(x)$
0.	.1743	7.5	601.
1.	.2203	9.9	218.
2.	.1464	13.4	330.
3.	.1091	24.0	474.
4.	.0684	27.9	685.
5.	.0507	52.4	871.
6.	.0419	72.2	1067.
7.	.0322	67.7	1156.
8.	.0297	82.8	1490.
9.	.0195	122.7	1528.
10.	.0084	251.5	3349.
11.	.0066	226.2	2946.
12.	.0077	188.9	2106.
13.	.0057	184.0	1971.
14.	.0057	331.9	2328.
15.	.0064	370.7	2715.
16.	.0050	303.2	2641.
17.	.0042	599.9	4187.
18.	.0042	496.8	3092.
19.	.0034	405.2	3382.
20.5	.0098	931.3	4457.
24.	.0094	1344.4	5607.
29.5	.0105	2929.2	7482.
39.	.0096	10276.5	12434.
80.	.0107	28488.0	82026.

The values of MSE(r) were calculated for $r = r_C$, r_G, r^*, and r_H with the value of $\varphi(x)$ guessed at in the formula. In each case, guess 1 is linear, guess 2 is quadratic, and guesses 3 and 4 are of the form $\varphi(x) = Cx^g + D$, with guess 3 attempting to fit the entire range of x and guess 4 only those x up to 16. The results appear in Table 2.

For population G, the reductions in MSE for the true φ are very impressive when compared with r_G, the better of the two "classical" estimators r_C and r_G, particularly for $n \geq 8$. Quadratic guess 2 is the best of the four, again with noticeable reduction of MSE for $n \geq 8$. Guess 1 is generally poorer than r_G; guesses 3 and 4 are better than r_G for $n \geq 16$, with guess 3 better than 4 (guess 4 is better than guess 3 for small n).[2]

3. MSE RESULTS ON BUSINESS DATA:
COMPARISON OF TRUE OPTIMAL AND OPTIMAL WITH GUESSED $\varphi(x)$

Population	Size	No. of Monte Carlo Samples	$MSE(r_C)$	$MSE(r_G)$	$MSE(r^*)$	$MSE(r_H)$	Guess 1[a] $MSE(r^*)$	1 $MSE(r_H)$
G	2	EXACT	7.085	11.39	3.296	11.39	3.470	11.39
	4	549	4.032	1.675	1.224	1.473	1.570	1.691
	8	275	1.296	.8178	.4872	.5073	.9156	.8969
	16	137	.8034	.5548	.2275	.2304	.6474	.6387
	32	68	.5074	.3965	.1101	.1107	.4329	.4299
	64	34	.2885	.2774	.05433	.05445	.2586	.2579
	128	17	.1555	.1385	.02694	.02697	.1473	.1471

Population	Size	No. of Monte Carlo Samples	Guess 2[b] $MSE(r^*)$	2 $MSE(r_H)$	Guess 3[c] $MSE(r^*)$	3 $MSE(r_H)$	Guess 4[d] $MSE(r^*)$	4 $MSE(r_H)$
G	2	EXACT	3.346	11.39	3.704	11.39	3.589	11.39
	4	549	1.299	1.533	1.673	1.789	1.557	1.701
	8	275	.5497	.5716	.8324	.8294	.8183	.8226
	16	137	.2666	.2705	.4400	.4396	.4753	.4757
	32	68	.1323	.1331	.2105	.2106	.2658	.2658
	64	34	.06591	.06611	.09961	.09961	.1426	.1426
	128	17	.03285	.03290	.04864	.04864	.07537	.07537

[a]Guess 1: $\varphi(x) = 7 + 35x$. [b]Guess 2: $\varphi(x) = 7 + 5x + 2x^2$. [c]Guess 3: $\varphi(x) = 10x^{1.8} + .001$.
[d]Guess 4: $\varphi(x) = 10x^{1.25} + .001$.

Population	Size	No. of Monte Carlo Samples	$MSE(r_C)$	$MSE(r_G)$	$MSE(r^*)$	$MSE(r_H)$	Guess 1 $MSE(r^*)$	1 $MSE(r_H)$
H	2	EXACT	162.3	158.1	63.73	158.1	64.22	158.1
	4	527	65.08	34.75	21.53	33.58	22.15	34.47
	8	263	15.36	9.151	7.547	8.267	8.146	8.861
	16	131	5.536	3.741	3.134	3.192	3.711	3.687
	32	65	2.600	1.807	1.445	1.451	1.872	1.834
	64	32	1.349	1.002	.7091	.7098	1.021	1.006
	128	16	.6893	.5601	.3505	.3506	.5425	.5380

Population	Size	No. of Monte Carlo Samples	Guess 2[b] $MSE(r^*)$	2 $MSE(r_H)$	Guess 3[c] $MSE(r^*)$	3 $MSE(r_H)$	Guess 4[d] $MSE(r^*)$	4 $MSE(r_H)$
H	2	EXACT	64.47	158.1	92.52	158.1	91.61	158.1
	4	527	22.70	34.63	35.75	41.05	32.27	37.90
	8	263	8.452	9.104	15.35	15.41	12.41	12.51
	16	131	3.631	3.669	6.107	6.102	5.352	5.352
	32	65	1.707	1.708	2.469	2.469	2.530	2.530
	64	32	.8445	.8441	1.093	1.093	1.328	1.328
	128	16	.4192	.4190	.5072	.5072	.7078	.7078

[a]Guess 1: $\varphi(x) = 400 + 200x$. [b]Guess 2: $\varphi(x) = 608 - 104x + 13x^2$. [c]Guess 3: $\varphi(x) = 220x^{1.3} + .001$.
[d]Guess 4: $\varphi(x) = 200x^{0.9} + .001$.

Population H does not have as large a variation in the function φ as does population G, so that the improvements, if any, are smaller. Except for r^* for n = 2,4, reductions in MSE for r^* and r_H, when compared to r_G, do not exceed 25 percent except for $n \geq 64$ (as compared with $n \geq 8$ for population G). Guess 2 is a marginal improvement over r_G except for $n \geq 64$. Except for r^* ($n \leq 8$), guess 1 (for both r^* and r_H) is about as efficient as r_G. Guesses 3 and 4 are almost always less efficient than r_C or r_G. Again, guess 3 is better than 4 for large n, and worse for small n.[2]

For both populations G and H, there seems to be a turning point: for some critical n, say $\hat{n}$, the *a priori* knowledge (or a suitable guess for it) really makes a difference — for $n \geq \hat{n}$, reductions in MSE are quite sizable. All four guesses yield improvements over r_C (as opposed to r_G) for population G, and also (for most n) for population H.

7. Final Comments

The relative efficiency of r^* and r_H over r_C and r_G depends on the shape of both the distribution of x and of the function $\varphi(x)$, as well as on the validity of the assumption $E(y|x) = \alpha + \beta x$. A low coefficient of variation of x (as in population A), or a linear function $\varphi(x)$ (as in population F) results in little reduction in MSE; one may as well use r_C or r_G. Furthermore, a poor guess of $\varphi(x)$ may result in an increase in MSE. The examples presented, however, show that significant gains in precision are sometimes possible, even when $\varphi(x)$ is guessed. It is felt, though not tested, that r^* and r_H perform well when the linearity of $E(y|x)$ is only approximately true.

The applicability of this theory is toward recurring surveys and toward surveys with a census base for determining guesses of $\underline{m}$ and φ. The theory presented may be extended to the errors-in-variable context, and is more clearly generalizable to stratified sampling (both stratum-by-stratum and over-all-strata estimators) and to a p-dimensional concomitant vector (p>1).

Linear superpopulation models are hardly new. R. M. Royall (e.g., in [8]) has advocated use of estimators and sample designs which minimize something akin to C.a.p. MSE. Two of the many other papers which treat this subject are by Cassel, Särndal, and Wretman [1] and by Scott and Smith [9]. Finally, A. A. Hasel presented r_H in 1942!

The major problem is, of course, the determination of α and V — or of V alone (if (8) and r_H are used). Many interpretations of m and V are possible — both sampling and nonsampling error may be included. More testing of the sensitivity of r* and r_H to κ is necessary; alternative κ's need be examined. Naive assumptions such as $V = \sigma^2 I$ or $\varphi(x) = dx$ will often result in classical estimators, which may suffice if no other a priori information is available. However, it is believed that in a recurring survey or special survey with a data base in a census, it is possible to obtain practical guesses for α and V.

---

[1] The development here can be shown to be equivalent to that in Royall and Herson ([8], 881-3), except that they deal with totals rather than ratios and, hence, are restricted to a finite population. Instead of using a sample indicator function u(s) which is zero for all nonsample elements, they make use of the nonsample moments $(j=0,1,\ldots)(\Sigma x_k^j)k\notin s$. They refer to estimators chosen to be unbiased under a model ξ as "ξ-unbiased." They compare $\hat{T}[0,1:x]$ and $\hat{T}[1,1:x]$ under the model $\xi(1,1:x)$, analogous to the comparison of r* and r_H as shown here. In general, however, they are concerned more with sample design than with estimation. Moreover, they do not exhibit their estimators in such a way to show the importance of the matrix V (the conditional variance function $\varphi(x)$.

An estimator which is unbiased under a sample design with function p is called, in the terminology of R. M. Royall [8] and others, "p - unbiased." The alternative ratio estimators of Hartley and Ross (r_{HR}) and M. R. Mickey (r_M) are p - unbiased when p is derived from SRS. Estimators which are chosen to be unbiased under a model such as (1), independent of the parameters in E(y|x), are called "ξ - unbiased." r_G is ξ - unbiased under (1); r_C is ξ - unbiased under (1) if $\alpha = 0$.

[2] This situation is comparable to that of population C: a "big" x - even if rare - will raise the C.a.p. MSE, more so for a large n than for a small one, since when n is large, a "big" x is more likely to appear in the sample. In this situation the low weighting implied by r* and r_H can help immensely. See [11], p. 45.

[3] See [11], pp. 13-15.

[4] See [11], pp. 7-12, and 27-32 for a more detailed discussion of these problems.

REFERENCES

[1] Cassel, Claes M., Särndal, Carl E., and Wretman, Jan H., "Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations," Biometrika (1976), 63, 3, pp. 615-620.

[2] Hasel, A. A. (1942), "Estimation of Volume in Timber Stands by Strip Sampling," Annals of Mathematical Statistics, Vol. XIII, No. 2, pp. 179-206.

[3] Hutchison, M. C. (1970), "A Monte Carlo Comparison of Some Ratio Estimators," Biometrika, pp. 313-321.

[4] Rao, J. N. K. (1969), "Ratio and Regression Estimators," in New Developments in Survey Sampling, eds. N. H. Johnson and H. Smith, pp. 213-234, New York: Wiley.

[5] ____________ (1965). "A note on the estimation of ratios by Quenouille's method," Biometrika, 52, 647-9.

[6] ____________ (1967). "Precision of Mickey's unbiased ratio estimator," Biometrika, 54, 321-4.

[7] Rao, Poduri S. R. S. (January 1977), "Ratio Method of Estimation in Finite Populations," University of Rochester, research paper partially supported by the U. S. Bureau of the Census.

[8] Royall, R. M. and Herson, Jay (1973), "Robust Estimation in Finite Populations I," Journal of the American Statistical Association (68), pp. 880-889.

[9] Scott, Alastair and Smith, T. M. F., "Linear Superpopulation Models in Survey Sampling," (1973), Proceedings of the IASS, Vienna, Austria.

[10] Tepping, B. J. (1968), "Variance Estimation in Complex Surveys," Proceedings of the American Statistical Association, August 20-23, 1968.

[11] Tomlin, P. H. (1972), "Ratio Estimation in a New Light," (unpublished paper). Research Center for Measurement Methods, U. S. Bureau of the Census.

[12] Woodruff, R. S. (1971), "Simple Method of Approximating Variance of a Complicated Estimate," Journal of the American Statistical Association, June 1971, pp. 411-414.