RATIO ESTIMATION IN MULTI-STAGE DESIGN

by

DR. BURTON T. OÑATE*

It must be admitted that many, if not most of the current, timely and precise statistics useful for socio-economic development programs and indispensable to sound administrative decisions and policy formulation, emanate from national sample surveys. **Current** means up-to-date; **timely** means that the interval between the date of actual collection and the release of the results is as short as possible, if necessary within a few weeks; and **precise** means the estimates or data released carry the required level of statistical reliability. Data from our censuses do not possess the properties of being current and timely.

Many of the current national sample surveys may be termed as "panel surveys" as far as certain sampling units (primaries or secondaries) are concerned and the survey once put into operation commits the estimation procedures to the permanent design. In such case, the choice of technical devices is limited to methods such as sampling within permanently selected sampling units, a redefinition of the sampling units, alternative choice of the estimation procedures and other techniques which will result in higher statiscal efficiency and lower cost of the survey.

This paper will consider the theoretical approaches to the use of an alternative choice of the estimation procdure which may be termned as ratio estimation in a multi-stage sampling with complete replacement of primaries (Oñate, 1960). The estimation procedures are simple and the additional increase in statistical efficiency will more than compensate for the small cost due to the additional work needed for compilation and

* Acting Director, Office of Statistical Coordination and Standards, National Economic Council.

computation relevant to the use of an auxiliary variable. Empirical results will be presented for Metropolitan Manila as a sector in the Philippine Statistical Survey of Households (PSSH Bulletin No. 1, 1957).

Another technique which utilizes an alternative approach to the estimation procedure is through a systematic rotation and the concomittant use of composite estimator (Oñate, 1960; 1961). This particular technique will not be discussed in this paper.

Theoretical Aspects.

In multi-stage sampling with complete replacement of primaries, we have at time α from the ith primary sampling unit (psu) of the hth stratum, an independent unbiased estimate, ${}_{\alpha}\hat{X}_{hi}$, of the stratum total, ${}_{\alpha}X_h$. The sample mean of the ${}_{\alpha}\hat{X}_{hi}$'s is also an unbiased estimate of ${}_{\alpha}X_h$. Each ${}_{\alpha}\hat{X}_{hi}$ is considered as an independent observation from an infinite population of ${}_{\alpha}\hat{X}_{hi}$ at the α visit. Depending on the stage(s)of sampling and availability of an auxiliarry variable, Z, which is highly correlated with X, we can also generate a random sample of the ${}_{\alpha}\hat{Z}_{hi}$'s which may be paired with the corresponding ${}_{\alpha}\hat{X}_{hi}$. With this consideration, standard techniques on ratio estimation for simple random sampling can be utilized. The regular unbiased estimate, ${}_{\alpha}\hat{X}_{hi}$ or ${}_{\alpha}\hat{Z}_{hi}$ of the stratum totals from the ith psu will take different forms depending upon the nature of the design employed for the different stages.

Only pertinent theoretical results and selected formulas will be given since the basic philosophy of the theoretical model are described in the author's paper (Oñate, 1960) and the details of results for simple random sampling are given by Cochran (1953) and Hartley (1959).

The following types of estimators will be studied, namely:

1. regular unbiased (the x- only) estimate.
   a. separate (or standard)
   b. combined
2. biased ratio of means estimate.
   a. separate
   b. combined
3. unbiased ratio estimate.
   a. separate
   b. combined

Regular unbiased estimate (the x- only) estimate. This estimate refers to ${}_a\hat{X}_{hi}$ (or ${}_a\hat{Z}_{hi}$) or the sample mean ${}_a\hat{X}_h$ which when summed through h will give en estimate ${}_a\hat{X}$. The estimate of the variance of ${}_a\hat{X}_h$ or ${}_a\hat{X}$ is rather simple and this has the form:

$$s_h^2 = \sum_i{}' \left(\hat{X}_{hi} - \hat{X}_h\right)^2 \Big/ m\,(m-1)\ ,$$

where $\hat{X}_{hi}$ and $\hat{X}_h$ refer to hth stratum,

$\sum'$ is the sample summation, and

m is the number of independent psu's drawn with replacement (the a subscript has been dropped for simplicity).

By appropriate random numbering of the ith psu in the hth stratum, we can generate a separate and a combined estimate of total or variance for the universe in question.

Biased ratio estimate. The estimate in this case will have the simple form:

$$X^* = q\ Z$$

where q and Z may be for the hth stratum or for the universe,

q is the ratio of two multi-stage estimators $\hat{X}$ to $\hat{Z}$, each of which is the mean of m estimates, and

Z is the actual count of the auxiliary variable in the stratum or the universe.

(Note that the appropriate subscripts have been dropped for simplicity in presentation.)

The estimate of the variance of X^* is

$$s^2(X^*) = \left[s^2(\hat{X}_i) + q^2\ s^2(\hat{Z}_i) - 2qs(\hat{X}_i, \hat{Z}_i) \right] / m$$

As in the regular unbiased estimate, we can generate a separate or combined biased ratio of means estimate. Eq. (3) indicates that theoretically we achieve a gain in precision if

$$\rho(\hat{X}_i, \hat{Z}_i) > \tfrac{1}{2}\, C.V.(\hat{Z}_i) / C.V.(\hat{X}_i)$$

where p is the correlation coefficient, and
C.V. is the coefficient of variation.

Unbiased ratio estimate. An unbiased ratio estimate for each stratum total or universe total has the form:

$$X' = r\,Z + (\hat{X} - r\,\hat{Z})m/m - 1$$

where r is the sample of ratios generated by the ith psu,

$$\text{i.e. } r_i = \hat{X}_i / \hat{Z}_i, \quad \text{and}$$

Z, m, $\hat{X}$, $\hat{Z}$ are as explained or defined earlier.

The appropriate estimate of the variance of X' is

$$s^2(X') = \left[s^2(\hat{X}_i) + r^2\, s^2(\hat{Z}_i) - 2rs(\hat{X}_i, \hat{Z}_i) \right] / m$$

and as in the two types of estimates, we can develop a separate or a combined unbiased ratio estimate of the universe total, X.

Note that $\hat{X}$ and $\hat{Z}$ are obtained through application of the appropriate multi-stage estimation procedures wherein the psu's are drawn with complete replacement. With the corresponding subscript on $\hat{X}$ or $\hat{Z}$, the estimate may refer to the ith psu estimate for the stratum or universe, or to the sample mean for the stratum or universe.

Empirical Results.

In order that empirical results have some national significance, the PSSH design for Metropolitan Manila will be used to indicate the applications of the theory. Since population count within sampling areas is usually highly correlated with most socio-economic characteristics observed in the area, listed population and/or population in sample hhs will be used as the indicator variable, x or X. The number of registered voters by precinct, by stratum and by the universe is made available every two years through the usual administrative channels with little or no cost to the actual survey. We will refer to this characteristic as the auxiliary variable, Z.

The six estimates, their estimated variances and the efficiency of each estimate relative to the regular unbiased (separate or combined) estimate were derived and these are presented in Tables I and II for October 1957. The significance in the use of these simple ratio estimators becomes doubly evident in view of the following:

a. with complete replacement of psu's, the reduction in variance due to the application of the finite population correction (1 — f) is lost, and

b. the psu's were drawn with equal probability and not with probability proportional to some measure of size (pps).

The application of ratio estimation will generally recover the losses in statistical efficiency with little or no cost to the survey proper and these facts are borne out by the relative efficiencies observed with the use of ratio estimators.

For population listed, the separate ratio estimators showed a gain in precision of about 5 percent while the combined ratio estimators were 48 percent more efficient than the combined regular unbiased (x- only) estimate (Table I). In 1960, the author using seven different strata reported a gain of 25 per cent for the separate estimators and 36 percent for the combined estimators. For population in sample households, the gains in precision were 25 percent and 37 percent for separate and combined estimators, respectively (see Table II).

The separate estimate of variance will have L(m — 1) degrees of freedom (D.F.) where L is the number of strata and (m — 1) is the number of D.F. for each stratum. On the other hand, the combined estimate of variance will have only (m — 1) D.F. It is therefore of low precision but simple

TABLE I

COMPARISON OF THE ESTIMATE, ESTIMATE OF VARIANCE AND STATISTICAL EFFICIENCY OF SIX TYPES OF ESTIMATORS FOR POPULATION IN LISTED HOUSEHOLDS

Type of Estimator	Estimate (thousand)	Estimate of Variance (million)	Relative Efficiency* (percent)	
			Separate	Combined
Regular unbiased				
(x- only)				
Separate	391.57	735.52	100	—
Combined	391.57	1,470.63	—	100
Biased ratio of means				
Separate	383.09	699.78	105	—
Combined	380.04	995.95	—	148
Unbiased ratio				
Separate	389.40	701.52	105	—
Combined	380.47	997.02	—	148

* Estimated variance of the x - only estimate as numerator.

COMPARISON OF THE ESTIMATE, ESTIMATE OF VARIANCE AND STATISTICAL EFFICIENCY OF SIX TYPES OF ESTIMATORS FOR POPULATION IN SAMPLE HOUSEHOLDS

Type of Estimator	Estimate (thousand)	Estimate of Variance (million)	Relative Efficiency* (percent)	
			Separate	Combined
Regular unbiased (x- only)				
Separate	408.07	1,016.29	100	—
Combined	408.07	1,755.02	—	100
Biased ratio of means				
Separate	399.64	813.16	125	—
Combined	396.07	1,278.35	—	137
Unbiased ratio				
Separate	400.65	820.89	124	—
Combined	395.98	1,279.26	—	137

* Estimated variance of the x - only estimate as numerator.

to compute. Also, the numbering of the psu's (i = 1, 2, . . . , m) must be independent and the allocation of number to the psu's must be done in a random manner.

The listing operation is carried a few days before actual enumeration of sample households. Because of this situation and the more relevant bias associated with listing, the universe under study may have changed at least within the sampling area. This will require a process of adjustment in the population of listed households, i.e.

$$\frac{\text{population enumerated in sample housholds}}{\text{population listed in sample households}} \times \text{population listed}$$

in sampling area = adjusted population listed in sampling area, which is a form of a ratio estimator itself. This problem is worth looking into.

Summary and Conclusion.

The theoretical framework in the use of ratio estimation in multi-stage sampling with complete replacement of primaries is described and empirical results are presented to indicate the level of precision attained in the use of ratio estimation.

Population in listed households and population in sample households for seven strata in the PSSH Metropolitan Manila were correlated with number of registered voters. Estimates of correlation from covariance analysis ranged between 0.4 to 0.5 and these correlations brought about gains in precision from 5 to 25 percent for the separate ratio estimators and from 36 to 48 percent for the combined ratio estimators as compared with the regular unbiased (x- only) estimates.

It is evident that the use of ratio estimation in the PSSH Metropolitan Manila with the given auxiliary variable will increase precision at no cost or very little additional cost to the survey operations. With the use of this estimation procedure, one may reduce the cost of the survey by about 20 percent, i.e., reduce the number of psu's from 5 to 4 and still maintain the present level of precision with 5 psu's and the use of the regular unbiased (x- only) estimate.

Literature Cited

Cochran, W. G. Sampling Techniques. New York, John Wiley and Sons, Inc. 1953.

Hartley, H. O. Theory of Advanced Design of Survey. (Mimeographed lecture notes). Ames, Iowa, Department of Statistics, Iowa State University, 1959.

National Economic Council and Bureau of Census and Statistics. Labor Force, Philippine Statistical Survey of Households Bulletin 1(1) 1957.

Oñate, B. T. Development of Multi-stage Designs for Statistical Surveys in the Philippines. Multi-lith series No. 3, Stat. Laboratory, Iowa state University. February 1960 or OSCAS Statistical Reporter 4(4), October 1960.

————. Rotation of Sample in Multi-stage Designs. The Philippine Statistician 10(1). March 1961.