THE ESTIMATION OF NON-SAMPLING VARIANCES IN CURRENT SURVEYS

H. O. Hartley and Paul P. Biemer, Texas A&M University

1. Introduction

The importance of non-sampling, or measurement errors has long been recognized (for the numerous references see e.g., the comprehensive papers by Hansen, Hurvitz and Bershad (1961) and Bailar and Dalenius (1970)). Briefly the various models suggested for such errors assume that a survey record (recorded content item) differs from its "true value" by a systematic bias and additive error contributions associated with various sources of errors such as, interviewers, coders, etc. The important feature of these models is that the errors made by a specified error source (say a particular interviewer) are usually 'correlated'. These correlated errors contribute additive components to the total mean square error of a survey estimate which do not decrease inversely proportional to the overall sample size but only inversely proportional to the number of interviewers, coders, etc. Consequently, the application of standard textbook formulas for the estimation of the variances of survey estimates may lead to serious underestimates of the real variability which should incorporate the non-sampling errors.

Attempts have, therefore, been made to estimate the components due to non-sampling errors. The early work in this area has concentrated on surveys specifically designed to incorporate features facilitating the estimation of non-sampling components such as reinterviewes and/or interpenetrating samples (see e.g. Sukhatme and Seth (1952)). However, the more recent literature (see e.g. Cochran (1968), Fellegi (1969), Nisselson and Bailar (1976), Battese, Fuller and Hickman (1976)) has also treated surveys in which such features are either lacking or limited, but these results are restricted to simple surveys permitting the use of analysis of variance techniques.

In this paper we provide a general methodology applicable to essentially any survey to estimate the total variance of estimators of target parameters (such as the population total). Our variance formulas include all relevant sampling and non-sampling variance components and these are estimated directly from the survey data and do not assume that estimates of non-sampling errors for "similar" content items made in special studies can be transferred to the current survey estimates. This paper does not address non-sampling biasses and represents a generalization of that by Hartley and Rao (1978) in that it extends the approach to cover situations of interviewer and coder assignments more commonly used in present survey practice. Below we summarize the assumptions made and the conditions under which variances can be estimated.

At the present time the technique has been applied to artificial data generated from special cases of models (1) and (2) below. It is not possible to apply it to survey data acquired in the past since

(a) the required information on interviewer and coder assignments is usually not available and
(b) even if (a) is available, interviewer and coder assignments have to satisfy certain estimability conditions (see below) which were not recognized in the past.

However, both, the Bureau of the Census and the Triangle Research Institute have expressed an interest in trying the method in future survey operations.

2. Model Assumptions for Non-Sampling Errors

This study is confined to non-sampling errors of quantitative content items. It is hoped to cover categorical items in subsequent studies. We adopt "additive error models" (also used in the more recent literature) in which the error made by (say) a particular interviewer are correlated through an additive error term. Confining ourselves to one particular content item, it is assumed that the true content item of the t-th respondent interviewed by interviewer i and coded by coder c has the following additive non-sampling errors.

$$\begin{aligned}\text{Interviewer error} &= b_i + \delta b_t \\ \text{Coder error} &= c_c + \delta c_t \\ \text{Respondent error} &= \delta r_t \end{aligned} \tag{1}$$

where

b_i = error variable contributed by i-th interviewer common to all units, t, interviewed by i-th interviewer.
c_c = error variable contributed by c-th coder common to all units, t, coded by c-th coder.
δb_t, δc_t, and δr_t = elementary interviewer, coder, and respondent errors afflicting the content item of unit t (respondent t).

We assume that the b_i, c_c, δb_t, δc_t are random samples from infinite populations with zero mean and variances σ_b^2, σ_c^2, $\sigma_{\delta b}^2$, $\sigma_{\delta c}^2$ and δr_t a (nonobserved) error with zero mean sampled from the finite population of respondents by the survey design implemented.

Our method can easily be extended to cover other and/or additional sources of non-sampling errors.

3. The Type of Survey Covered

The type of survey here covered is essentially a general stratified multistage survey with restrictions delineated in 4 below. To fix the ideas we first describe the concepts in terms of a three stage survey and then outline the general multistage situation. Denote by η_{pst} the true content item for tertiary, t, of secondary, s, of "primary", p, where the "primary index p" is a double subscript denoting the actual primary within its stratum. Denote by y_{pst} the corresponding recorded content item. Then we clearly have $y_{pst} = \eta_{pst}$ + total error in (1) and if we replace the units label t by the triple subscript pst, this equation can be written in the form

$$\begin{aligned} y_{pst} = \bar{\eta}_{p..} + (\bar{\eta}_{ps.} - \bar{\eta}_{p..}) + b_i + c_c \\ + (\eta_{pst} - \bar{\eta}_{ps.}) + \delta r_{pst} \\ + \delta b_{pst} + \delta c_{pst} \end{aligned} \qquad (2)$$

where the $\bar{\eta}_{ps.}$, $\bar{\eta}_{p..}$ are respectively the secondary and primary population means of the true content items. We now combine the terms in the second and third lines of (2) and write $\varepsilon_{pst} = (\eta_{pst} - \bar{\eta}_{ps.}) + \delta r_{pst}$ and $e_{pst} = \delta b_{pst} + \delta c_{pst}$. The characteristic feature of our approach (similar to that of Hartley and Rao (1978)) is that we recognize that for most survey designs the variance of the target parameter estimates only depends on the variances of the pooled terms ε_{pst} and e_{pst} so that such estimates can be computed without estimating the variances of the individual terms in (2). Moreover, the special case (also considered by Hartley and Rao (1978)) in which

(a) the last stage (tertiary) units are drawn with equal probability and
(b) the last stage (tertiary) populations are essentially infinite (fpc negligible)

the pooled terms $\varepsilon_{pst} + e_{pst}$ are random samples from infinite populations with variances $\sigma_e^2(p, s)$ (say) and the variance of the target parameter estimates only depends on $\sigma_e^2(p, s)$. No assumptions need therefore be made about the independence of the five individual terms comprising $\varepsilon_{pst} + e_{pst}$ and randomizations (of say respondents) ensuring this independence are not required (note, however, the randomization of interviewers and coders ensuring the independence of the b_i and c_c from $\varepsilon_{pst} + e_{pst}$ discussed in Section 5). Denoting the between secondary within primary variation $\bar{\eta}_{ps.} - \bar{\eta}_{p..}$ by δ_{ps}, the variance components S_δ^2, σ_b^2, σ_c^2 and $\sigma_e^2(p, s)$ can be estimated by the method of synthesis (Hartley, Rao and LaMotte (1978)), briefly summarized in Appendix 1 provided that certain estimability conditions are satisfied. However, the estimates of the S_δ^2 are not used in the formulas for variance estimation of target parameters as will be seen in Appendix 2. In this paper we retain assumption (a) but are able to eliminate (b) although the formulas given in Appendix 1 below retain both (a) and (b). Moreover the estimability conditions implied by the Hartley-Rao variance estimator are somewhat restrictive and are here generalized to bring them closer to customary practices of interviewer (and coder) work assignments.

One of our estimability conditions (Section 4) will stipulate that all tertiaries in a particular secondary are handled by one interviewer and one coder. Moreover, in order to invoke the components of variance estimation procedure, it is convenient to average (2) over the tertiary units and obtain

$$\bar{y}_{ps.} = \bar{\eta}_{p..} + b_i + c_c + \delta_{ps} + \bar{e}_{ps.} \qquad (3)$$

where $\delta_{ps} = (\bar{\eta}_{ps.} - \bar{\eta}_{p..})$ and $\bar{e}_{ps.}$ is the average of the pooled terms $\varepsilon_{pst} + e_{pst}$. Finally, in order to give a concise description of the components of variance estimation techniques (see Appendix 1) it is convenient to rewrite (3) in analysis of variance notation using design matrices as shown below.

$$y = X\bar{\eta} + U_b b + U_c c + \sum_p W_p(\delta_p + e_p) \qquad (4)$$

where y, $\bar{\eta}$, b, c, δ_p and e_p are the vectors of the terms in (3) and X, U_b, U_c and W_p are the corresponding design matrices.

4. Estimability Conditions and Interviewer-Coder-Work Load Assignments

A general necessary and sufficient condition for the estimability of the components of variance σ_b^2, σ_c^2, $\sigma_e^2(p, s)$ is given in Appendix 1. We confine ourselves here to giving simple sufficient conditions for work assignments of survey personnel which ensure that these conditions are satisfied.

(i) The sample contains at least two primaries per stratum, two secondaries per primary, two tertiaries per secondary, etc.
(ii) In each primary there are at least two secondaries entirely interviewed by the same interviewer and coded by the same coder.
(iii) In at least one primary there are at least two secondaries entirely interviewed by different interviewers but coded by the same coder.
(iv) In at least one primary there are at least two secondaries entirely coded by different coders.

Condition (i) will almost always be satisfied for classical variance estimation. Surveys with one primary per stratum do not even permit unbiased estimation of sampling variances and are nowadays rare. Condition (ii) will almost always be satisfied since normally the total work load in a particular primary is assigned to one interviewer and one coder. Conditions (iii) and (iv) specify that exceptions to rule (ii) should occur with at least one pair of secondaries in at least one primary. Condition (iv) is easily satisfied by coder allocation but condition (iii) requires a special interviewer assignment.

At the present stage of development we must further stipulate that

(v) the sampling of secondaries within primaries and tertiaries within secondaries must be with equal probability and without replacement.

For this case of a three stage survey, formulas for the estimated variance of target parameter estimates are given in Appendix 2.

Certain generalizations are feasible. The role of the primary sampling stage can be taken over by any lower stage which we will term the "critical stage". The conditions (i) to (v) above are then restated by calling "primary units", "critical units" and "secondary units", "subcritical units". No restrictions on the survey design will be made for stages above the critical stage so that now condition (v) permits unequal probability sampling for all stages from the primary down to and including the critical stage. However, condition (iii) now stipulates that the work load in a (lower order) critical unit must be shared by two interviewers and this may not be feasible in all practical situations.

In terms of the above concepts, the method of Hartley and Rao (1978) uses the last-but-one stage as the critical sampling stage and hence would require that two different interviewers share the load in at least one last-but-one stage unit. If the critical stage is the last stage, then the subcritical stage would require repeat observations on a set of some last stage units. This case is not addressed in this paper.

In many survey operations it will be very costly to satisfy condition (iii) by assigning different interviewers to two secondaries in the same primary (or, indeed, to two different tertiaries in the same secondary, etc.), while for coders such assignments are quite feasible. In such cases it may be well to use a modification of conditions (ii), (iii) and (iv) which specify strata as the critical stage for interviewers and secondaries as the critical stage for coders. These modifications result in the following conditions for estimability.

(ii)' In each stratum there are at least two primaries entirely interviewed by the same interviewer and in each secondary there are at least two tertiaries coded by the same coder.

(iii)' In at least one stratum there are at least two primaries entirely interviewed by different interviewers.

(iv)' In at least one secondary there are at least two tertiaries coded by different coders.

Under the above assumptions the component of variance estimation requires modification which is described in Appendices 3 and 4.

5. A Design for Interviewer and Coder Allocation

In this section we confine ourselves to the implementation of personnel allocation satisfying conditions (ii)', (iii)', and (iv)'. Other cases will be treated in a more extensive follow up paper.

First we should stress that conditions (i), (ii)', (iii)', and (iv)' are merely sufficient conditions ensuring the estimability of all variance components. However, they only provide for a single interviewer contrast and a single coder contrast which would result in variance estimates of poor precision. The mere principle of unbiased estimates is in line with the general tendency in the sample survey literature which places emphasis on designs yielding small variances for the estimates of target parameters while the precision of variance estimates have only received limited attention. The inclusion of non-sampling variance components may well raise the question of how to design the personnel allocation to obtain variance estimates of reasonable precision. Without claiming any optimality properties for our procedure, we aim at a roughly equal number of interviewer contrasts within strata.

We confine ourselves here to a three-stage stratified design and discuss the allocation of

(a) interviewers to primaries within strata and

(b) coders to tertiaries within secondaries.

The procedure employs modified incomplete block designs and will be explained below for the special case of 5 interviewers to be allocated to a set of strata containing 4 or 3 or 2 primaries. Below are shown two modified incomplete block designs which will be used as a basis for the interviewer allocation.

Design A; Block #

	1	2	3	4	5	6	7	8	9	10
	1	1	2	1	2	1	2	1	3	1
Interviewer	2	4	4	3	3	2	3	2	4	3
#	3	5	5	4	5	5	4	4	5	5
	1	4	5	3	2	5	3	2	4	1

Design B; Block #

	1	2	3	4	5	6	7	8	9	10
	1	1	1	1	2	2	2	3	3	4
Interviewer	2	3	4	5	3	4	5	4	5	5
#	2	1	4	1	3	2	5	3	5	4

Denote by n_p the number of strata with p primaries (p = 2, 3, 4). If $n_4 = 0$, Design A will not be used. If $1 \leq n_4 \leq 10$, select the first n_4 blocks of Design A, note the interviewer contrasts represented and "even up" the contrasts by selecting n_3 blocks from Design B which have contrasts which are under represented. Finally, even up the interviewer load by using the n_2 strata with two primaries per stratum.

Example: $n_4 = 2$, $n_3 = 4$, $n_2 = 6$.

We select blocks 1 and 2 from Design A providing the interviewer contrasts 12, 13, 23 and 14, 15, 45 leaving the contrasts 24, 25, 34, and 35 to be provided by Design B. Hence we must select blocks 6, 7, 8 and 9 to provide the full complement of ten contrasts. This leaves two more blocks to be selected from Design B. If we select blocks 10 and 1, we have the following work load for interviewers.

Interviewer	1	2	3	4	5
# of Primaries Allocated	4	6	4	6	6

Each of the $n_2 = 6$ strata with 2 primaries must be covered by a single interviewer. To even up the above work-load to 8 or 6 primaries per interviewer, we allocate two of these strata to interviewer 1 or 3 (say 1) and one stratum to interviewers 2, 3, 4 and 5. This gives interviewer #3 the lower work load of 6 primaries. The allocation of interviewers to the numbers 1, 2, 3, 4 and 5 and that of actual strata to blocks should be random.

For the allocation of coders to secondaries we wish to retain a large number of degrees of freedom for the estimation of each between tertiary within secondary variance component since this contains also the four components of elementary errors. Hence it is suggested that two coders split the work load in each secondary. Since the number of secondaries is likely to be large, it is suggested that all possible pairs of coders be allocated to secondaries in random sequences as long as secondaries are available. For example, for 4 coders and 31 secondaries, we would allocate the sequence of pairs in the order 12, 13, 14, 23, 24, 34; ... 34; 12.

Appendix 1. Formulas for Component of Variance Estimation

A modified version of the synthesis-based procedure for estimating the components of variance in mixed ANOVA models (Hartley, Rao and LaMotte (1978)) is now described for obtaining unbiased estimates of σ_b^2, σ_c^2 and $\sigma_e^2(p, s)$.

Without loss of generality we may assume that $X'X = I$ in (4) since this can always be achieved with a suitable reparameterization of the vector $\bar{\eta}$.

Define

$$Q_b(y) = y'V_bV_b'y \quad \text{where} \quad V_b = U_b - XX'U_b ,$$

$$Q_c(y) = u'V_cV_c'y \quad \text{where} \quad V_c = U_c - XX'U_c$$

and, for all p,

$$Q_p(y) = y'V_pV_p'y \quad \text{where} \quad V_p = W_p - XX'W_p .$$

The essence of our procedure is to

(a) compute $\hat{D}_p = \text{diag}\left\{\frac{\hat{\sigma}_e^2(p,s)}{m(p,s)}\right\}$ where

$$\hat{\sigma}_e^2(p, s) = \sum_{t=1}^{m(p,s)} \frac{(y_{pst}-\bar{y}_{ps})^2}{m(p,s)-1}$$

and m(p, s) is the number of tertiaries in the secondary labeled (p, s).

(b) Compute

$$Q_b^*(y) = Q_b(y) - \text{tr } W_p'V_bV_b'W_p\hat{D}_p$$

$$Q_c^*(y) = Q_c(y) - \text{tr } W_p'V_cV_c'W_p\hat{D}_p$$

and, for all p,

$$Q_p^*(y) = Q_p(y) - \text{tr } W_p'V_pV_p'W_p\hat{D}_p.$$

(c) Apply the usual synthesis-based procedure using the modified quadratic forms computed in (b) in place of Q(y) to obtain unbiased estimates of σ_b^2 and σ_c^2.

Using a result in Hartley, Rao and LaMotte (1978), a necessary and sufficient condition for the components σ_b^2 and σ_c^2 to be estimable is that the matrices V_bV_b', V_cV_c' and V_pV_p', for all p, be linearly independent.

Appendix 2. Formulas for Estimated Variances of Target Parameter Estimates

The estimators of target parameters and their variances are now considered in terms of the model (4). The majority of target parameters - including population totals and means - which are computed from sample survey data are linear functions of the y_{pst}. Since sampling within secondaries is with equal probabilities, the discussion is confined to estimators of the form

$$\hat{Y} = \gamma'\bar{y} \tag{5}$$

where $\bar{y}$ is the vector of secondary means and γ is a coefficient vector which may depend upon the set S of secondaries in the sample. It is easily shown that $\hat{Y}$ is unbiased for the target parameter if $\gamma'\dot{\eta}$ is unbiased where $\dot{\eta}$ is the vector of true secondary means, $\bar{\eta}_{ps}$.

Let G be the set of sampled secondaries S and the interviewer-coder work assignments. The variance of $\hat{Y}$ is composed of two components as follows:

$$\text{Var } \gamma'\bar{y} = \underset{G}{\text{Var}}\ E|_G\ \gamma'\bar{y} + \underset{G}{E}\ \text{Var}|_G\ \gamma'\bar{y} \tag{6}$$

where $\text{Var}|_G$ and $E|_G$ denote the variance and expectation, respectively, given the set G and Var_G and E_G denote the variance and expectation over all possible sets G.

An unbiased estimator of (6) is derived in Biemer (1978) by estimating each of the two components separately. The resulting estimation formula is

$$\text{var } \gamma'\bar{y} = \bar{y}'\Omega\bar{y} - \text{tr } \Omega\hat{\Sigma} + \gamma'U_bU_b'\gamma\hat{\sigma}_b^2$$
$$\gamma'U_cU_c'\gamma\hat{\sigma}_c^2 + \sum_p \gamma'W_p\hat{D}_pW_p'\gamma \tag{7}$$

where

Ω = a constant matrix for given set S which is directly provided by standard finite population sampling theory without non-sampling errors and satisfies $E(\dot{\eta}'\Omega\dot{\eta}) = \mathrm{Var}\ \gamma'\dot{\eta}$,

$$\hat{\Sigma} = U_b U_b' \hat{\sigma}_b^2 + U_c U_c' \hat{\sigma}_c^2 + \overset{G}{\Sigma} W_p \hat{D}_p W_p' ,$$

and $\hat{\sigma}_b^2$, $\hat{\sigma}_c^2$ and $\hat{D}_p$ are computed as described in Appendix 1.

Appendix 3. Component of Variance Estimation for Survey Designs Specified in Section 5

Formulas for estimating the total variance of target parameter estimates for survey designs in which interviewers are allocated to primaries within strata and coders are allocated to tertiaries within secondaries are given in Appendix 4. In this appendix, the procedure for estimating the necessary components in these formulas is outlined.

By replacing the composite index p defined in Section 3 by the double index (h, p), denoting the p-th primary in stratum h, (2) may be rewritten as

$$y_{hpst} = \bar{\eta}_{h...} + \phi_{hp} + \delta_{hps} + b_i + c_j + \varepsilon_{hpst} + e_{hpst} \qquad (8)$$

where $\bar{\eta}_{h..}$ denotes the population mean of true values for stratum h, $\phi_{hp} = (\bar{\eta}_{hp..} - \bar{\eta}_{h..})$ and the remaining terms are defined as in (2). The variances $S_\phi^2(h)$, $S_\delta^2(h, p)$, σ_b^2, σ_c^2 and $\sigma_e^2(h, p, s)$ are estimable by the synthesis-based procedure described below provided the estimability conditions stated in Section 4 are satisfied. Although they are computed, the estimates of S_ϕ^2 and S_δ^2 are not used in the formulas given in Appendix 4.

The estimation of the components σ_b^2, σ_c^2 and $\sigma_e^2(h, p, s)$ is accomplished in three stages. In the first stage, σ_c^2 and $\sigma_e^2(h, p, s)$ are simultaneously estimated for a given sample of secondaries. The next stage provides estimates of $S_\delta^2(h, p)$ for a given sample of primaries. Finally, estimates of σ_b^2 and as a by-product $S_\phi^2(h)$ are obtained in the third stage while utilizing the estimates obtained in the previous two stages.

Stage 1: Estimate σ_c^2 and $\sigma_e^2(h, p, s)$.

For convenience and conciseness of notation, (8) can be written alternatively as

$$y = X\eta + U_b b + U_c c + \sum_{h,p,s} W_{hps} e_{hps} \qquad (9)$$

where η is the vector with elements $\bar{\eta}_{hps.}$, e_{hps} is the vector of pooled terms $\varepsilon_{hpst} + e_{hpst}$, $b = (b_i)$, $c = (c_j)$, and X, U_b, U_c and W_{hps} are design matrices. It is assumed without loss of generality that $X'X = I$.

(a) Compute the quadratic forms

$$Q_c(y) = y'V_cV_c'y \quad \text{where} \quad V_c = U_c - XX'U_c$$

and, for all secondaries (h, p, s)

$$Q_{hps}(y) = y'V_{hps}V_{hps}'y \quad \text{where}$$
$$V_{hps} = W_{hps} - XX'W_{hps} .$$

(b) Using the method of synthesis, obtain the coefficients ℓ_{ck} and ℓ_{hpsk} such that

$$E\ Q_c(y) = \sum_k \ell_{ck}\sigma_k^2$$

and (10)

$$E\ Q_{hps}(y) = \sum_k \ell_{hpsk}\sigma_k^2$$

where σ_k^2 denotes the components σ_c^2 and $\sigma_e^2(h, p, s)$.

(c) Invert the system (10) to obtain unbiased estimates of σ_c^2 and $\sigma_e^2(h, p, s)$.

Stage 2: Estimate $S_\delta^2(h, p)$.

In order to invoke the variance component estimation procedure at this stage, a model for the secondary sample means is needed. Therefore, (8) is averaged over the tertiary units yielding

$$\bar{y}_{hps.} = \bar{\eta}_{hp..} + \delta_{hps} + b_i + \frac{1}{m(h,p,s)} \sum_j \nu[(h, p, s); j]c_j + \bar{e}_{hps.} \qquad (11)$$

where m(h, p, s) is the number of tertiaries sampled in secondary (h, p, s) and $\bar{e}_{hps.}$ is the average of the pooled terms $\varepsilon_{hpst} + e_{hpst}$. This model is expressed using design matrices as

$$\bar{y} = \tilde{X}\bar{\eta} + \tilde{U}_b b + Z_c c + \sum_{h,p} W_{h,p}(\delta_{hp} + \bar{e}_{hp}) \qquad (12)$$

where $\bar{y}$, $\bar{\eta}$, b, c, δ_{hp} and $\bar{e}_{hp}$ are vectors of the terms in (11), X, U_b and W_{hp} are design matrices and $Z_c = \left(\frac{\nu[(h,p,s);j]}{m(h,p,s)}\right)$. It is assumed that $\tilde{X}'\tilde{X} = I$ without loss of generality.

(a) For each primary (h, p), compute the modified quadratic forms

$$Q_{hp}^*(\bar{y}) = \bar{y}'V_{hp}V_{hp}'\bar{y} - \mathrm{tr}\ Z_c'V_{hp}V_{hp}'Z_c\hat{\sigma}_c^2 - \sum_{h,p} \mathrm{tr}\ W_{hp}'V_{hp}V_{hp}'W_{hp}\hat{D}_{hp}$$

where $\hat{D}_{hp} = \mathrm{diag}\left\{\frac{\hat{\sigma}_e^2(h,p,s)}{m(h,p,s)}\right\}$.

(b) Using the method of synthesis, obtain the coefficients ℓ_{hpk} such that

$$E\, Q^*_{hp}(\bar{y}) = \sum_k \ell_{hpk} S^2_\delta(k) \qquad (13)$$

where $S^2_\delta(k)$ denotes the components $S^2_\delta(h, p)$.

(c) Invert the system (13) to obtain unbiased estimates of $S^2_\delta(h, p)$.

Stage 3: Estimate σ^2_b.

For this procedure, the model (8) is averaged over tertiaries and secondaries yielding the model

$$\bar{\bar{y}}_{hp..} = \bar{\eta}_{h..} + \phi_{hp} + \bar{\delta}_{hp.} + b_i + \frac{1}{m(h,p)} \sum_s \frac{1}{m(h,p,s)} \sum_j \nu[(h, p, s);j]c_j + \bar{\bar{e}}_{hp..} \qquad (14)$$

where m(h, p) is the number of secondaries sampled in primary (h, p). The corresponding design model is given by

$$\bar{\bar{y}} = \tilde{\tilde{X}}\bar{\bar{\eta}} + \tilde{\tilde{U}}_b b + \tilde{Z}_c C + \sum_h W_h(\phi_h + \bar{\delta}_h + \bar{\bar{e}}_h) \qquad (15)$$

where $\bar{\bar{y}}$, $\bar{\delta}_h$ and $\bar{\bar{e}}_h$ are vectors of the corresponding elements in (14), $\tilde{\tilde{X}}$, $\tilde{\tilde{U}}_b$ and W_h are design matrices and $\tilde{Z}_c$ is the matrix with [(h, p); j] element equal to $\frac{1}{m(h,p)} \sum_s \frac{1}{m(h,p,s)} \nu[(h, p, s); j]$. As usual, $\tilde{\tilde{X}}'\tilde{\tilde{X}} = I$ is assumed without loss of generality.

Define

$$\hat{\Lambda}_h = \mathrm{diag}\{(1 - \frac{m(h,p)}{M(h,p)}) \frac{\hat{S}^2_\delta(h,p)}{m(h,p)}\}$$

and

$$\hat{\hat{D}}_h = \mathrm{diag}\{\frac{1}{m(h,p)} \sum_s \frac{\hat{\sigma}^2_e(h,p,s)}{m(h,p,s)}\}$$

where M(h, p) is the number of secondaries in primary (h, p).

(a) Compute the modified quadratic forms

$$Q^*_b(\bar{\bar{y}}) = \bar{\bar{y}}'V_bV_b'\bar{\bar{y}} - \mathrm{tr}\, \tilde{Z}_c'V_bV_b'\tilde{Z}_c\sigma^2_c - \mathrm{tr}\, W_h'V_bV_b'W_h\hat{\Lambda}_h - \mathrm{tr}\, W_h'V_bV_b'W_h\hat{\hat{D}}_h$$

where $V_b = \tilde{\tilde{U}}_b - \tilde{\tilde{X}}\tilde{\tilde{X}}'\tilde{\tilde{U}}_b$, and for all strata h,

$$Q^*_h(\bar{\bar{y}}) = \bar{\bar{y}}'V_hV_h'\bar{\bar{y}} - \mathrm{tr}\, \tilde{Z}_c'V_hV_h'\tilde{Z}_c\sigma^2_c - \mathrm{tr}\, W_h'V_hV_h'W_h\hat{\Lambda}_h - \mathrm{tr}\, W_h'V_hV_h'W_h\hat{\hat{D}}_h$$

where $V_h = W_h - \tilde{\tilde{X}}\tilde{\tilde{X}}'W_h$.

(b) Using the method of synthesis, obtain the coefficients ℓ_{bb}, ℓ_{bk}, ℓ_{hb} and ℓ_{hk} such that

$$E\, Q^*_b(\bar{\bar{y}}) = \ell_{bb}\sigma^2_b + \sum_k \ell_{bk}S^2_\phi(k)$$

and

$$E\, Q^*_h(\bar{\bar{y}}) = \ell_{hb}\sigma^2_b + \sum_k \ell_{hk}S^2_\phi(k) \, . \qquad (16)$$

(c) Invert the system (16) to obtain unbiased estimates of σ^2_b and $S^2_\phi(h)$.

Appendix 4: Formulas for Estimated Variances of Target Parameters Estimates for Survey Designs Specified in Section 5.

For the survey designs specified in Section 5, the variance estimation formula given in (7) for estimator (5) still applies if the matrix U_c in (7) is replaced by the matrix Z_c defined in Appendix 3. The components σ^2_b, σ^2_c and the diagonal matrix $\hat{D}_p$ are now estimated by the procedure summarized in Appendix 3.

REFERENCES

Bailar, B. A. and Dalenius, T. (1970). "Estimating the response variance components of the U.S. Bureau of the Census survey model." Sankhya Series B, 341-360.

Battese, G. E., Fuller, W. A., and Hickman, R. D. (1976). "Estimation of response variance from interviewer re-interview surveys." Journal Indian Society of Agricultural Statistics, 28, 1-14.

Biemer, P. P. (1978). "The estimation of non-sampling variance components in sample surveys," unpublished Ph.D. dissertation, Institute of Statistics, Texas A&M University.

Cochran, W. G. (1968). "Errors of measurements in statistics." Technometrics 10, 637-666.

Fellegi, I. P. (1974). "An improved method of estimating the correlated response variance." Journal of American Statistical Assn., 1969, 496-501.

Hansen, M. H., Hurwitz, W. N. and Bershad, M. A. (1961). "Measurement errors in censuses and surveys." Bull. International Stat. Inst., 38, 359-374.

Hartley, H. O. and Rao, J. N. K. (1978). "The estimation of non-sampling variance components in sample surveys," in Survey Sampling and Measurement, New York: Academic Press.

Hartley, H. O., Rao, J. N. K., and LaMotte, L. (1978). "A simple 'synthesis'-based method of variance component estimation," Biometrics 34, 233-42.

Nisselson, H. and Bailar, B. A. (1976). "Measurement analysis and reporting of non-sampling errors in surveys." Proceedings of the International Biometric Conference (Boston, 1976), 301-321.

Searle, S. R. (1971). Linear Models, John Wiley & Sons, Inc., New York.

Sukhatme, P. V. and Seth, G. R. (1952). "Non-sampling errors in surveys." Journal Indian Society of Agricultural Statistics, 4, 5-41.