AN IMPROVED PROCEDURE FOR ESTIMATING THE COMPONENTS OF RESPONSE VARIANCE IN COMPLEX SURVEYS

Paul P. Biemer, U.S. Bureau of the Census

1. INTRODUCTION

A general procedure for estimating the total variance of linear estimators in complex surveys was proposed by Hartley and Rao (1978). Based on a linear additive model, their procedure utilizes recent developments in component of variance estimation methodology to form estimates of all relevent sampling and nonsampling variance components directly from the current survey data.

The main objective of the present paper is to provide a more efficient estimator of the interviewer variance component obtained in their procedure without incurring any extra field costs. Furthermore, the estimability conditions of the Hartley-Rao method require that the interviewer work assignments be "interpenetrated" within some last-but-one stage units of a multistage survey. For many surveys of the U.S. Bureau of the Census, this may require that an interviewer travel over an area many times the size of the usual enumeration area thereby increasing field costs. The present method extends their procedure to allow the interviewer assignments to be interpenetrated within one or more areas the size of at least two enumeration areas (interviewer assignment areas).

The present technique is similar in approach to the Fellegi (1974) method for improving the estimator of the correlated response variance but retains the generality of the Hartley-Rao procedure. Two unbiased and (assuming normally distributed observations) uncorrelated estimators of the interviewer variance are formed as follows:

(a) The usual component of variance estimates of the interviewer and coder variances are obtained using the Hartley-Rao technique in enumeration areas satisfying specified estimability conditions.

(b) A second estimate of the interviewer variance is obtained based on all interviewers assignments including those not satisfying the estimability conditions for (a).

Hence, an appropriate weighted average of the estimators of interviewer variance obtained in (a) and (b) may provide an estimator having smaller variance than either (a) or (b) taken separately.

2. SURVEY DESIGN

The survey design considered is a general stratified multistage survey in which the units of the last-but-one stage are drawn with equal probability while any probability sampling may be specified for the higher stages. To fix the ideas, the concepts are described in terms of a stratified two stage survey; however, the formulas and results may still be applied for any multistage survey by interpreting the term "secondary" to mean the last-stage unit and the term "primary" to mean the composite of the last-but-one stage unit within next higher stage unit ... within primary unit within stratum.

Let $P_1, \ldots, P_n$ denote the n primary units selected from all strata according to some sample design. Let the union of these sampled primary units,

$$\bigcup_{p=1}^{n} P_p ,$$

be divided into I mutually exclusive and exhaustive enumeration areas or EAs, denoted by E_i , i=1, ..., I. Let the sample of units within each EA be assigned originally to only one interviewer.

Many times in practice, the area assigned to an interviewer for enumeration is usually the most accessible area to the interviewer and little or no attention is paid to the primary boundaries. Therefore, provision is made for EAs which overlap into two or more primary units. For this purpose let E_{pa}, a = 1, ..., α_p, denote the α_p elements of the set

$$\{P_p \cap E_i \mid P_p \cap E_i \neq \phi, i=1,..,I\} .$$

That is, E_{pa} is the a-th EA - or partial EA for areas which straddle two or more primaries - lying within the boundaries of primary p.

Let $P = \{P_1, \ldots, P_n\}$ denote the set of sampled primary units. It is assumed that

(i) for any set P, the design specifies in advance the number, m_p, of secondaries to be drawn with equal probability from the M_p units in primary p,

(ii) the secondary sampling procedure is performed independently within each E_{pa} - the sampling fraction being equal to the primary sampling fraction, m_p/M_p, and

(iii) given the set of E_{pa} 's and for any sample of secondaries, the number of secondaries to be interviewed by interviewer i and coded by coder j in each E_{pa} is prespecified.

These assumptions are similar to those made by Hartley and Rao for the sake of subsequent conditioning arguments. Assumption (ii) ensures that the number of last-stage units (secondaries) in the sample that fall within each E_{pa} is specified prior to sampling within primaries. This specification is merely conceptual since in actual practice the above sampling procedure may only be approximated; for example, the units may be sorted geographically and a systematic sample taken within primary units. Assumption (iii) is also conceptual since the interviewer and coder work assignments are usually not known until after the complete sample has been drawn.

3. DEFINITION OF THE MODEL

Let η_{pas} denote the true item response for the s-th secondary sampled in E_{pa}. (Recall that p is actually a double subscript denoting the primary and its stratum). Denote by y_{pas} the corresponding recorded item response. The following additive model is assumed for survey content items subject to respondent, interviewer, and coder errors. Generalizations to additional sources of nonsampling error do not afford any difficulties.

$$y_{pas} = \eta_{pas} + r_{pas} + b_i + \delta b_{pas} + c_j + \delta c_{pas} \qquad (3.1)$$

where

- r_{pas} denotes the response error of the s-th respondent in E_{pa},
- b_i denotes the systematic interviewer error common to all units interviewed by the i-th interviewer,
- c_j denotes the systematic coder error common to all units coded by the j-th coder, and
- δb_{pas}, δc_{pas} denote the elementary errors of the interviewer and coder of unit (p, a, s).

Let I denote the number of interviewers and J denote the number of coders available for the survey. It is assumed that $\{b_1, \ldots, b_I\}$ and $\{c_1, \ldots, c_J\}$ are random samples from infinite populations of interviewer errors and coder errors with mean zero and variances σ_b^2 and σ_c^2, respectively. It is assumed that r_{pas} is an error with zero mean sampled from the finite population of respondents by the survey design implemented. The elementary errors are assumed to be uncorrelated random components with zero mean.

Letting $\bar{\eta}_{pa}$ denote the population mean of E_{pa} and $\bar{\eta}_p$ denote the population mean of primary p, the model (3.1) may be written alternatively as

$$y_{pas} = \bar{\bar{\eta}}_p + \delta_{pa} + b_i + c_j + e_{pas} \qquad (3.2)$$

where

$$\delta_{pa} = (\bar{\eta}_{pa} - \bar{\bar{\eta}}_p)$$

and

$$e_{pas} = (\eta_{pas} - \bar{\eta}_{pa}) + r_{pas} + \delta b_{pas} + \delta c_{pas}$$

Since sampling of secondaries within primaries is with equal probability and neglible sampling fractions, the pooled terms e_{pas} may be regarded as random samples from infinite populations with zero means and variances $\sigma_e^2(p, a)$, say. Moreover, e_{pas}, b_i, and c_j are assumed to be mutually uncorrelated.

The model (3.2) may be finally rewritten incorporating design matrices as

$$\underset{\sim}{y} = X\underset{\sim}{\bar{\bar{\eta}}} + \sum_{p=1}^{n} W_p \underset{\sim}{\delta}_p + U_b \underset{\sim}{b} + U_c \underset{\sim}{c} + \sum_{p=1}^{n} \sum_{a=1}^{\alpha_p} W_{pa} \underset{\sim}{e}_{pa} \qquad (3.3)$$

where $\underset{\sim}{\bar{\bar{\eta}}}$, $\underset{\sim}{\delta}_p$, $\underset{\sim}{b}$, $\underset{\sim}{c}$, and e_{pa} are the vectors of the components in (3.2) and X, W_p, U_b, U_c, and W_{pa} are the corresponding design matrices.

4. DESIGN FOR INTERVIEWER AND CODER ASSIGNMENT

In this section, general guidelines are given for specifying the design matrices U_b, U_c, and W_{pa} in (3.3) so that:

(a) the variance components σ_b^2, σ_c^2, and $\sigma_e^2(p, a)$ are estimable by the synthesis-based variance component estimation procedure (Hartley, Rao, and LaMotte (1978)).

(b) a second unbiased estimator of σ_b^2 may be computed, and

(c) the two estimators of σ_b^2 obtained for (a) and (b) are uncorrelated under certain specified conditions.

The synthesis-based procedure referred to in (a) provides a very general necessary and sufficient condition of estimability in terms of the matrices U_b, U_c, and W_{pa} (see Section 6). However, simpler sufficient conditions which ensure that this general condition is satisfied and which pertain directly to the allocation of interviewers and coders have been derived for general survey designs (Biemer (1978)). These conditions require that

(i) within each E_{pa} there are at least two last-stage units interviewed by the same interviewer and coded by the same coder,

(ii) within at least one E_{pa} there are at least two last-stage units interviewed by different interviewers but coded by the same coder,

(iii) within at least one E_{pa}, there are at least two last-stage units coded by different coders.

(4.1)

Satisfying these minimal conditions for estimability will ensure that (a) above is true without regard to (b) or (c). In order that a second estimator of σ_b^2 may be computed, the following procedure for interviewer allocation is proposed.

Let the I EAs formed in Section 2 be grouped together by any arbitrary criterion into L non-overlapping "blocks" each containing k EAs. It has been assumed for convenience that the numbers of interviewers, I, is evenly divisible by k.

Let a random sample of ℓ blocks be drawn from these L blocks of the EAs and let this sample be denoted by C. Then to each block of EAs in C, apply the Pattern I interviewer allocation scheme and to the remaining L - ℓ blocks apply the Pattern II scheme as follows.

<u>Pattern I</u> - For this pattern of interviewer allocation, the k interviewers associated with each EA of a block in C split the workload in

each of the k EAs. This design may be regarded as a generalization of the concept of interpenetrating interviewer assignment areas in which each interviewer is assigned the same fraction, 1/k, of the sample of units in each EA.

Let $f_{tt'}$ (t, t' = 1, ..., k) denote the fraction of the total sample of units in the t-th EA of a block that is interviewed by the interviewer originally assigned to the t'-th EA of the block. Pattern I then specifies

$$f_{tt'} > 0 \quad , \quad \text{for } t,t' = 1,\ldots,k. \tag{4.2}$$

It is a further requirement of the procedure (justified in Appendix B) that the rank of matrix $F = [f_{tt'}]$ be less than k. For example, the classical interpenetration design for k interviewers may be specified by $F = (1/k)1_{kk}$ where 1_{kk} is the kxk matrix of ones.

Pattern II - This pattern simply specifies the original allocation with one interviewer working in only one EA in the block. Using the notation defined above, this may be specified simply by F = I.

The cost of interpenetrating interviewers' assignments within EAs can add considerably to the cost of a survey. On the other hand, interpenetration of coders' assignments is more easily accomplished with few additional costs. Hence, only one estimator of the coder variance component, σ_c^2, is provided based on an allocation pattern satisfying the conditions in (4.1).

As long as (4.1) is satisfied, the particular pattern of coder assignments is immaterial to the estimability of the interviewer variance, σ_b^2, by either of the two methods utilized. However, the independence of the estimators of σ_b^2 may be destroyed unless careful consideration is given to the allocation of coders (see Appendix B). Hence, sufficient conditions have been derived which will guarantee this independence.

Define G_γ (γ = 1, ..., L) as the kxJ matrix whose (t,j) element is the fraction of the total sample of units in the t-th EA of block γ coded by coder j. The estimators of σ_b^2 are independent under certain specified conditions if either

(i) $G_\gamma = FC_\gamma$, for some kxJ matrix C_γ, or
(ii) the elements of G_γ are either 0 or 1.

(4.3)

These conditions imply that each EA of a block must be handled by the same number of coders. Furthermore, blocks in which the EAs are assigned to two or more coders must also satisfy (i) above.

The matrices F and C_γ are assumed to be known for a given set P of sampled primaries and for any set $\mathcal{C}$. That is, F and the coder work assignments do not depend upon the selection of Pattern I blocks.

5. THE VARIANCE OF LINEAR ESTIMATORS

The estimators of population parameters and their variances are now considered in terms of the model (3.3).

The majority of the estimators of target parameters which are computed from sample survey data are linear functions of the y_{pas}. Since sampling within primaries is with equal probability, the discussion is confined to estimators of the form

$$\hat{Y} = \underset{\sim}{w}' \bar{\bar{\underset{\sim}{y}}} \tag{5.1}$$

where $\bar{\bar{y}}$ is the nx1 vector of primary means

$$\bar{\bar{y}}_p = \frac{1}{m_p} \sum_{a=1}^{\alpha_p} \sum_{s=1}^{\alpha_{pa}} y_{pas} , \tag{5.2}$$

with α_{pa} being the number of secondaries sampled in E_{pa}, and where $\underset{\sim}{w}$ is a coefficient vector which may depend upon the set P. Clearly, (5.1) is unbiased if $\underset{\sim}{w}$ is chosen so that $\underset{\sim}{w}'\bar{\eta}$ is unbiased.

Consider the variance of $\hat{Y}$. Let G denote the set of sampled primary units, P, with the EAs and blocks of EAs specified. Then

$$\text{Var}(\hat{Y}) = \underset{G}{\text{Var}}\, E(\hat{Y}|G) + \underset{G}{E}\, \text{Var}(\hat{Y}|G) \tag{5.3}$$

where $E(\cdot|G)$ and $\text{Var}(\cdot|G)$ denote the expectation and variance given the set G and $\underset{G}{E}$ and $\underset{G}{\text{Var}}$ denote the expectation and variance over all possible sets G.

Let $\nu(p, i; b)$ represent the number of units in primary p interviewed by interviewer i and let $\nu(p, j; c)$ denote the number of units in primary p coded by coder j. It follows from the assumption that the matrices F and C_γ are prespecified for any set $\mathcal{C}$ that $\nu(p, j; c)$ and α_{pa} are fixed given G while $\nu(p, i; b)$ is only fixed given the set $\mathcal{C}$ of Pattern I blocks. Therefore, for the second term on the right of (5.3), one obtains

$$\text{Var}(\underset{\sim}{w}'\bar{\bar{\underset{\sim}{y}}}|G) = \underset{\sim}{w}' \Sigma \underset{\sim}{w} \tag{5.4}$$

where the nxn matrix Σ is the conditional covariance matrix of $\bar{\bar{y}}$ given G whose (p, π) element, denoted $\Sigma_{p,\pi}$, is given by

$$m_p^2 \Sigma_{p,p} = \sigma_b^2 \sum_{i=1}^{I} E(\nu^2(p,i;b)|G) + \sigma_c^2 \sum_{j=1}^{J} \nu^2(p,j;c) + \sum_{a=1}^{\alpha_p} \alpha_{pa} \sigma_e^2(p,a) \tag{5.5}$$

for $p \neq \pi$, and

$$m_p^2 \Sigma_{p,p} = \sigma_b^2 \sum_{i=1} E(\nu^2(p,i;b)|G) + \sigma_c^2 \sum_{j=1}^{J} \nu^2(p,j;c) + \sum_{a=1}^{\alpha_p} \alpha_{pa} \sigma_e^2(p,a)$$

An unbiased estimator of $\underset{G}{E}\, \text{Var}(w'\bar{\bar{y}}|G)$ is therefore given by

$$\text{var}_W = \hat{\sigma}^2_b \sum_i \{\sum_p \nu(p,i;b) \frac{w_p}{m_p}\}^2$$

$$+ \hat{\sigma}^2_c \sum_j \{\sum_p \nu(p,j;c) \frac{w_p}{m_p}\}^2 + \sum_p \frac{w_p^2}{m_p^2} \sum_a \alpha_{pa} \hat{\sigma}^2_e(p,a) \tag{5.6}$$

where $\hat{\sigma}^2_b$, $\hat{\sigma}^2_c$ and $\hat{\sigma}^2_e(p, a)$ are computed as described in the next section.

Now, consider the first term on the right of (5.3), viz.

$$\underset{G}{\text{Var}}\, E(Y|G) = \text{Var}\, \underset{\sim}{w}'\, \bar{\bar{\underset{\sim}{\eta}}} . \tag{5.7}$$

Standard finite population sampling theory, regarding the primary units as last-stage units, provides unbiased estimators of Var $w'\bar{\bar{\eta}}$ of the form

$$\widehat{\text{Var}}\, \underset{\sim}{w}'\bar{\bar{\underset{\sim}{\eta}}} = \bar{\bar{\underset{\sim}{\eta}}}'\Omega\, \bar{\bar{\underset{\sim}{\eta}}} \tag{5.8}$$

where Ω is a nxn coefficient matrix which may depend upon P. Hence,

$$\text{var}_B = \bar{\bar{\underset{\sim}{y}}}'\Omega \bar{\bar{\underset{\sim}{y}}} - \text{tr}\, \Omega\, \hat{\Sigma} . \tag{5.9}$$

is an unbiased estimator of (5.7) where $\hat{\Sigma}$ is the unbiased estimate of Σ computed in (5.6).

6. PROCEDURE FOR ESTIMATING σ^2_b, σ^2_c, AND $\sigma^2_e(p, a)$

The model in (3.2) may be written alternatively as

$$y_{pas} = \bar{\eta}_{pa} + b_i + c_j + e_{pas} \tag{6.1}$$

or, using design matrices,

$$\underset{\sim}{y} = \dot{X}\bar{\underset{\sim}{\eta}} + U_b\underset{\sim}{b} + U_c\underset{\sim}{c} + \sum_p \sum_a W_{pa}\underset{\sim}{e}_{pa} \tag{6.2}$$

where $\bar{\eta}$ is the

$$\alpha(= \sum_{p=1}^{n} \alpha_p) \times 1$$

vector of the E_{pa} means, $\dot{X}$ is the associated

$$m(= \sum_{p=1}^{n} m_p) \times \alpha$$

design matrix and the remaining terms are defined as in (3.3.).

Because of the foregoing assumptions for a given set C of Pattern I blocks, the model (6.2) represents a mixed analysis of variance model of the form

$$y = X \underset{\sim}{\beta} + \sum_{q=1}^{\alpha+2} U_q \underset{\sim}{b}_q \tag{6.3}$$

where X, U_1, ..., $U_{\alpha+2}$ are design matrices, $\underset{\sim}{\beta}$ is a vector of constants, $\underset{\sim}{b}_q$ is a vector of random variables with $E(\underset{\sim}{b}_q) = 0$ and $\text{Var}(\underset{\sim}{b}_q) = \sigma^2_q I$. Using the synthesis-based procedure of variance component estimation, estimates of the variances σ^2_q (or σ^2_b, σ^2_c, and $\sigma^2_e(p, a)$) can be obtained which are conditionally unbiased given C and, hence, are generally unbiased.

The form of the synthesis estimates will be described in terms of the general model (6.3). Assume that β has been reparameterized so that $X'X = I$ and define

$$Q_q(\underset{\sim}{y}) = \underset{\sim}{y}'\, A_q\, \underset{\sim}{y} \tag{6.4}$$

where

$$A_q = (U_q - XX'U_q)(U_q - XX'U_q)' . \tag{6.5}$$

Using the results of Hartley, Rao, and LaMotte, an unbiased estimator of $\underset{\sim}{\sigma}^2 = [\sigma^2_q]$ is given by

$$\hat{\underset{\sim}{\sigma}}^2 = \Lambda^{-1}\, \underset{\sim}{Q} \tag{6.6}$$

provided this inverse exists, where

$$\Lambda = [\lambda_{qq'}] \text{ for } q,q' = 1,\ldots, \alpha + 2 \tag{6.7}$$

with

$$\lambda_{qq'} = \text{tr}\, U_q'\, A_{q'}\, U_q$$

and $\underset{\sim}{Q} = [Q_q]$.

The general necessary and sufficient condition for estimability provided by the synthesis-based procedure states that Λ^{-1} exists if and only if the matrices A_q are all linearly independent. As mentioned in Section 4, this is true if conditions (4.1) are satisfied.

Let $\hat{\sigma}^2_b(1)$ denote the estimate of σ^2_b obtained by (6.6). A second estimator of σ^2_b, denoted $\hat{\sigma}^2_b(2)$, will now be formed which is approximately uncorrelated with $\hat{\sigma}^2(1)$.

Consider the γ-th EA group formed by the procedure described in Section 4. Let $\bar{\underset{\sim}{y}}_\gamma$ denote the kx1 vector of EA sample means $\bar{y}_{\gamma 1}, \ldots, \bar{y}_{\gamma k}$ for γ-th block. Define the quadratic forms

$$B_\gamma(\underset{\sim}{y}) = \bar{\underset{\sim}{y}}_\gamma'\, \Phi\, \bar{\underset{\sim}{y}}_\gamma \tag{6.8}$$

for $\gamma = 1, \ldots, L$ where

$$\Phi = I\text{-}FF^-$$

with F^- denoting the generalized inverse of F. Let

$$\bar{B}_I = \frac{1}{\ell} \sum_{\gamma \in C} B_\gamma(\underset{\sim}{y}) \qquad (6.9)$$

and

$$\bar{B}_{II} = \frac{1}{L-\ell} \sum_{\gamma \notin C} B_\gamma(\underset{\sim}{y}) . \qquad (6.10)$$

Let r denote the rank of the matrix F and recall that $r < k$. Then it is verified in Appendix A that

$$\hat{\sigma}_b^2 (2) = \frac{1}{k-r} (\bar{B}_{II} - \bar{B}_I) \qquad (6.11)$$

is an unbiased estimator of σ_b^2 and, under the assumption of normality for y, is uncorrelated with $\hat{\sigma}_b^2(1)$ obtained in (6.6). Hence, the estimator

$$\hat{\sigma}_b^2 = \xi \hat{\sigma}_b^2 (1) + (1-\xi) \hat{\sigma}_b^2 (2) \qquad (6.12)$$

for suitably chosen ξ is a better estimator than either of the estimators $\hat{\sigma}_b^2(1)$ and $\hat{\sigma}_b^2(2)$. Note that the optimal choice of ξ is

$$\xi = \frac{V_2}{V_1 + V_2} \qquad (6.13)$$

where $V_i = Var(\hat{\sigma}_b^2(i))$, i = 1, 2, may be estimated from the survey data.

It is easily established that $\hat{\sigma}_b^2(2)$ is uncorrelated with the quantities $\lambda(p, i; b)$ defined in Section 5. Hence, (6.12) together with the variance component estimates of σ_c^2 and $\sigma_e^2(p, a)$ provide an unbiased estimator of $Var(\hat{Y})$ via (5.6) and (5.9).

7. SUMMARY

The steps involved in the procedure just described are now summarized. Any procedure equivalent to the following will satisfy the specification assumptions made for the survey design:

(1) For every primary in the population, specify m_p, the number of secondaries to be drawn with equal probability from the M_p secondaries in the population.

(2) Implement the primary stage sampling procedure and select the set $\mathcal{P}$ of primary units.

(3) Delineate the EA boundaries for this set of sampled primaries and group these EAs into L blocks each containing k EAs.

(4) Specify the type of interviewer interpentration design to be followed in Pattern I blocks, i.e. specify the matrix F.

(5) Assign coders to the EAs in accordance with condition (4.3).

(6) Select the set C of ℓ Pattern I blocks randomly from the L blocks of EAs.

(7) Select the sample of secondaries in each of the $\mathcal{E}_{pa}$'s, assigning them at random to interviewer-coder teams in accordance with conditions (4.1). As mentioned in Section 2, this sampling can usually be accomplished approximately by sorting the secondary units geographically within a primary unit and drawing a systematic sample.

This procedure for drawing the sample and allocating units to the interviewers and coders will ensure that the components of sampling and nonsampling variance which compose $Var(\hat{Y})$ are estimable and, moreover, that two unbiased and independent estimates of the interviewer variance component, σ_b^2, may be computed. Of course, the estimability conditions that obtain for estimating the sampling variances at each stage when nonsampling errors are ignored are still applicable to the last-but-one and higher stages.

APPENDIX A: DERIVATION OF $\sigma_b^2(2)$

Assume the vector $\underset{\sim}{y}$ is ordered by enumeration areas. Let $m_{\gamma t}$ denote the number of units sampled in the t-th EA of block γ. Define the mxI block diagonal matrix

$$T = \text{diag}\{T_{\gamma t}\}, \gamma = 1,\ldots,L, t = 1,\ldots,k \qquad (A.1)$$

where

$$T_{\gamma t} = \frac{1}{m_{\gamma t}} \underset{\sim}{1}_{m_{\gamma t}}$$

and where $\underset{\sim}{1}_{m_{\gamma t}}$ is the $m_{\gamma t}$x1 vector of 1's. Write $\bar{\underset{\sim}{y}}_\gamma$, the vector of EA means for block γ, as

$$\bar{\underset{\sim}{y}}_\gamma = H_\gamma T' \underset{\sim}{y} \qquad (A.2)$$

where H is the kxI matrix with elements h given by

h_{ti} = 1 if $\bar{y}_{\gamma t}$ is the i-th EA mean in $T'\underset{\sim}{y}$
= 0 if otherwise.

From (A.2) and (3.1), one obtains

$$\bar{\underset{\sim}{y}}_\gamma = \bar{\underset{\sim}{\eta}}_\gamma + [\theta_\gamma F + (1-\theta_\gamma)I]\underset{\sim}{b}_\gamma + G_\gamma \underset{\sim}{c} + \bar{\underset{\sim}{e}}_\gamma \qquad (A.4)$$

where

$\bar{\underset{\sim}{\eta}}_\gamma$ = k-vector of EA population means for block γ,
θ_γ = 1 if $\gamma \in C$, 0 if $\gamma \notin C$,
$\underset{\sim}{b}_\gamma$ = k-vector of interviewer variables, b_j, associated with block γ,
$\underset{\sim}{e}_\gamma$ = $(\bar{e}_{\gamma 1}, \ldots, \bar{e}_{\gamma k})$ where $\bar{e}_{\gamma t}$ is the sample mean of the elementary errors, $r_{pas} + \delta b_{pas} + \delta c_{pas}$, for the t-th EA in block γ,

and F, G_γ, and $\underset{\sim}{c}$ are as previously defined.

Let $B_\gamma(\underset{\sim}{y})$ be defined by (6.8). From (A.4), one obtains

$$E(B_\gamma(\underset{\sim}{y})|\gamma) = \bar{\underset{\sim}{\eta}}' \Phi \bar{\underset{\sim}{\eta}}_\gamma + \text{tr } \Phi FF' \sigma_b^2 + \text{tr } \Phi G_\gamma G_\gamma' \sigma_c^2 + \text{tr } \Phi \text{ Var}(\bar{\underset{\sim}{e}}|\gamma) \quad (A.5)$$

for $\gamma \varepsilon C$, and

$$E(B_\gamma(\underset{\sim}{y})|\gamma) = \bar{\underset{\sim}{\eta}}_\gamma' \Phi \bar{\underset{\sim}{\eta}}_\gamma + \text{tr } \Phi \sigma^2 + \text{tr } \Phi G_\gamma G_\gamma' \sigma_c^2 + \text{tr } \Phi \text{ Var}(\bar{\underset{\sim}{e}}|\gamma) \quad (A.6)$$

for $\gamma \notin C$. Now, taking the expectation of (A.5) and (A.6) over all possible choices of C but for fixed G, one obtains

$$E(B_\gamma(\underset{\sim}{y})|G) = \frac{1}{L} \sum_{\gamma=1}^{L} E(B_\gamma(\underset{\sim}{y})|\gamma) . \quad (A.7)$$

Clearly from (A.5) to (A.7)

$$E(\bar{B}_{II} - \bar{B}_I|G) = \text{tr } \Phi (I-FF')\sigma_b^2$$

which simplifies to

$$= \text{tr } (I-FF^-)\sigma_b^2 \quad (A.8)$$

if Φ is chosen as in (6.8). (This choice of Φ is discussed in Appendix B.) Since

$$\text{tr } (I-FF^-) = k-r \quad (A.9)$$

(see, for example, Searle (1971), p. 12), it follows that

$$\hat{\sigma}_b^2(2) = \frac{1}{k-r}(\bar{B}_{II} - \bar{B}_I)$$

is an unbiased estimator of σ_b^2.

APPENDIX B: PROOF THAT $\hat{\sigma}_b^2(1)$ AND $\hat{\sigma}_b^2(2)$ ARE INDEPENDENT

The covariance of $\hat{\sigma}_b^2(1)$ and $\hat{\sigma}_b^2(2)$ may be decomposed as follows:

$$\text{Cov}(\hat{\sigma}_b^2(1),\hat{\sigma}_b^2(2)) = \underset{C}{\text{Cov}}(E(\hat{\sigma}_b^2(1)|C), E(\hat{\sigma}_b^2(2)|C)) + \underset{C}{E} \text{ Cov}(\hat{\sigma}_b^2(1),\hat{\sigma}_b^2(2)|C) .$$

Since $E(\hat{\sigma}_b^2(1)|C) = \sigma_b^2$, the first term on the right is zero. Hence, all that remains is to show that

$$\text{Cov}(\hat{\sigma}_b^2(1),\hat{\sigma}_b^2(2)|C) = 0.$$

Due to the lack of space this proof has been omitted. However, the details are available upon request from the author.

REFERENCES

Biemer, P.P. (1978). "The Estimation of Non-sampling Variance Components in Sample Surveys", Unpublished Ph.D. dissertation, Instutite of Statistics, Texas A & M University.

Fellegi, I.P. (1974). "An Improved Method of Estimating the Correlated Response Variance". Journal of the American Statistical Association, 69: 496-501.

Hartley, H.O. and Rao, J.N.K. (1978). "The Estimation of Non-sampling Variance Components in Sample Surveys," Survey Sampling and Measurements, 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: 233-242.

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