BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-84/06

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

by

Paul P. Biemer and S. Lynne Stokes

U.S. Bureau of the Census
Washington, D.C. 20233

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Recommended by: Kirk M. Wolter

Report completed: June 17, 1983

Report issued: January 24, 1984

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

by

Paul P. Biemer

and

S. Lynne Stokes

ABSTRACT

Fellegi's (1974) improved method for estimating the interviewer component of correlated response variance is extended to $\ell$ groups of k interviewer assignments for general multistage survey designs. Using a linear models approach suggestive of Hartley and Rao (1978), the independence of the two estimators of interviewer variance is established and the forms of the variances of the estimators are derived. Then, using 1980 Census data to compute terms in the estimator variances, (1) the optimal design of interviewer variance studies is considered, (2) the improvement of the composite estimator is demonstrated, and (3) some principles of efficient study design are developed.



KEY WORDS: Response or interviewer variance; survey nonsampling error; optimal survey design; survey evaluation.

Paul P. Biemer is a Principal Statistical Researcher in the Center for Survey Methods Research, the Bureau of the Census. S. Lynne Stokes is a Mathematical Statistician in the Center for Survey Methods Research, the Bureau of the Census. The authors would like to thank the referees for their helpful comments and suggestions.

## 1. INTRODUCTION

There are numerous references in the literature which document the importance of response, or measurement, error in sample surveys (see, e.g., (Dalenius 1977), for a comprehensive bibliography). Models have been used extensively to measure the impact of the various types of response errors on survey estimators. They typically assume that a survey response differs from a true response by an error term which is contributed by the various operations and other error sources in a survey. These models indicate that when the errors made by a specified error source (say, a particular interviewer) are correlated, the variance of a survey estimator is increased by additive components called the correlated components of response variance (e.g., (Cochran 1977)). As a result of these correlated components, the usual unbiased estimators of the variances of means and totals are negatively biased, thus overstating the reliability of estimates (cf., (Sukhatme and Sukhatme 1970)). The models indicate that these biases can be eliminated or reduced if estimates of the correlated components are available. Further, it has long been recognized that the estimated components contain considerable information regarding the quality of the particular operation(s) under investigation (cf., (Hansen, Hurwitz, Bershad 1961), (Hansen and Marks 1951)). Therefore, a considerable number of studies have been conducted to estimate the components due to correlated response variance, especially for interviewers. These studies usually require the randomization or interpenetration of interviewer (or operator) assignments as a condition for the estimability of the response variance components.

Fellegi (1974) introduced an improved estimator of the interviewer component which is a convex combination of two unbiased estimators formed as follows:

1) The usual components of variance estimator is formed based upon interpenetrated pairs of interviewer assignments.

2) A second estimator is formed based on all interviewer assignments including those not interpenetrated.

Fellegi speculated on the relative magnitude of the variances of the two estimators and conjectured that they are uncorrelated and, hence, might be combined to provide an estimator having smaller variance than either taken separately. At least two applications of his procedure have been reported: Krotki (1978) and Bailey, Moore, and Bailar (1978).

Because of the costs and complexities of interpenetration studies, estimates of interviewer variance are often based upon a relatively small number of interviewers and are notoriously unstable. Exacerbating this problem is a void in the literature of methods for optimizing the design of interviewer (or operator) variance studies in complex surveys to take advantage of the existing resources.

This paper provides a general methodology for the estimation of the correlated components of response variance. Estimators (1) and (2), are generalized by allowing interpenetration of $\ell$ groups of k interviewer assignments for general multistage survey designs. Their independence is established for the special case of normally distributed observations. Then a generalization of the Fellegi composite estimator is constructed and formulas for the variances of the three estimators are derived. Finally, the optimal design of interviewer variance studies (e.g., the choice of $\ell$ and k) subject to survey budgetary contraints is considered using examples of population parameter configurations computed from 1980 Census data.

## 2. SURVEY ERROR MODEL

The survey design considered is a general stratified multistage survey in which the units of the last stage are drawn with equal probability without replacement within the units of the next-to-last stage. The concepts will first be described in terms of a two-stage survey and then outlined for the general multistage situation.

Let $\Pi$ denote the sample of n primary units selected from all strata according to some sample design. Consider the selection of the sample of secondaries and the formation of initial interviewer work assignments within primaries which we label $p = 1,\ldots n$. (Here, the primary index p is actually a double subscript denoting the primary unit within its stratum.) Most surveys implement a procedure for forming assignments which is equivalent to the following: Each sample primary area, p, is partitioned into $I_p$ ($I_p > 1$) enumeration areas (EA's) which are roughly equal with respect to the number of population units, where $I_p$ is the number of interviewers required to handle the primary workload. A random sample of units denoted by $\xi_{pa}$, $a = 1,\ldots,I_p$, is then drawn independently without replacement from each EA. Each $\xi_{pa}$ is then assigned to one interviewer for enumeration.

To describe the error of interviewers and respondents, we adopt an additive model and confine ourselves to only one content item. Let $\eta_{pas}$ denote the true value for the s-th secondary in $\xi_{pa}$ and let $y_{pas}$ denote the corresponding recorded content item. Assume

$$y_{pas} = \eta_{pas} + b_i + \delta_{pas} + r_{pas} \tag{2.1}$$

where $b_i$ is the systematic interviewer error common to all units interviewed by the ith interviewer, $\delta_{pas}$ is the elementary interviewer error and $r_{pas}$ is the respondent error associated with unit (p,a,s). (Generalizations to additional sources of error are feasible in the same manner as Hartley and Rao (1978)). Following Hartley and Rao, it is assumed that (1) $\{b_i, i=1,\ldots,I\}$ where $I = \sum_p I_p$ constitutes a random sample from the infinite population of interviewer errors with mean zero and variance $\sigma_b^2$; (2) the $\delta_{pas}$ are i.i.d. random variables with zero means; and (3) the $r_{pas}$ are zero mean random variables sampled from the finite population of respondents by the survey design implemented. For simplicity, we also assume that finite population corrections are negligible as in their procedure. Denote the variance of the pooled terms $(\eta_{pas} - \bar{\eta}_{pa}) + \delta_{pas} + r_{pas} = e_{pas}$, say, by $\sigma_e^2(p,a)$. We must assume that the $b_i$ are independent from the $e_{pas}$; however, no assumptions need be made about the independence of the three individual terms comprising $e_{pas}$. Model (2.1) may now be written as

$$y_{pas} = \bar{\eta}_{pa} + b_i + e_{pas}. \qquad (2.2)$$

To simplify the subsequent components of variance estimation formulas, it is convenient to rewrite (2.2) in matrix form as

$$\underset{\sim}{y} = \underset{\sim}{X}\,\underset{\sim}{\bar{\eta}} + \underset{\sim}{U}_b\underset{\sim}{b} + \sum_{p}\sum_{a} \underset{\sim}{W}_{pa}\, \underset{\sim}{e}_{pa} \qquad (2.3)$$

where $\underset{\sim}{\bar{\eta}}$ , $\underset{\sim}{b}$ and $\underset{\sim}{e}_{pa}$ are vectors of the components in (2.2) and $\underset{\sim}{X}$, $\underset{\sim}{U}_b$ and $\underset{\sim}{W}_{pa}$ are the corresponding design matrices.

## 3. DESIGN FOR INTERVIEWER ALLOCATION

In this section, general guidelines are given for specifying the design matrices $\underset{\sim}{X}$, $\underset{\sim}{U}_b$, and $\underset{\sim}{W}_{pa}$ in (2.3) so that (1) the variance components $\sigma_b^2$ and $\sigma_e^2(p,a)$ are estimable by the synthesis-based variance component estimation technique (Hartley, Rao, and La Motte 1978), (2) a second unbiased estimator $\sigma_b^2$ may be computed, and (3) the two estimators of $\sigma_b^2$ obtained for (1) and (2) are uncorrelated.

The components of variance estimation procedure of Hartley, Rao, and La Motte provides a general necessary and sufficient condition for the estimability of the components $\sigma_b^2$ and $\sigma_e^2(p,a)$. Simpler sufficient conditions which ensure that this general condition is satisfied and which pertain to the work assignments of survey personnel have been derived for general survey designs (Biemer 1978). These conditions require that

(1) within each $\xi_{pa}$ there are at least two last-stage units interviewed by the same interviewer, and

(2) within at least one $\xi_{pa}$ there are at least two last-stage units interviewed by different interviewers.

Condition (1) is always satisfied by the usual assignment of $\xi_{pa}$'s to interviewers; however, condition (2) specifies that for at least one $\xi_{pa}$, two or more interviewers should share the workload. This is essentially the requirement of interpenetration. We should stress that condition (2) is merely a sufficient condition which provides for only a single interviewer contrast. Without claiming any optimality properties for our procedure, we provide designs for maximizing the number of interviewer contrasts where feasible. We should further note that these conditions only provide for the Hartley, Rao, and La Motte estimator of $\sigma_b^2$. A scheme for assigning secondaries to interviewers is now described which allows a second estimator of $\sigma_b^2$ to be computed.

Let the I EA's formed in the previous section for the sample of primary units, $\Pi$, initially be assigned to one enumerator. Now, group the EA's together into L non-overlapping blocks each containing k EA's. This may be done by any convenient grouping criterion. (We assume for simplicity that L = I/k is an integer.) Let a random sample of $\ell$ blocks be chosen from the L and denote the sample by $\xi$. To each block of EA's in $\xi$ apply the Pattern I interviewer allocation scheme and to the remaining $L - \ell$ blocks apply the Pattern II scheme below.

Pattern I

For this pattern of interviewer allocation, the k interviewers associated with each EA of a block in $\mathcal{G}$ split the workload in each of the k EA's. Let $f_{tt'}(t,t' = 1,\ldots,k)$ denote the fraction of the total sample of units in the t-th EA of a block that is randomly assigned to the interviewer originally allocated for the t′-th EA of the block. Pattern I then specifies $f_{tt'} > 0$, for $t,t'=1,\ldots,k$.

This design may be regarded as a generalization of the concept of interpenetrated interviewer assignments for which each interviewer is assigned the same fraction, 1/k, of the sampled units in each EA. Our procedure only requires that the rank of the matrix $\underset{\sim}{F} = [f_{tt'}]$ be less than k (justification is provided in Appendix C). For example, the classical interpenetration design for k interviewers is a special case where $\underset{\sim}{F} = k^{-1}\underset{\sim}{1}_{kk}$ where $\underset{\sim}{1}_{kk}$ is the k×k matrix of ones.

Pattern II

The allocation pattern for blocks in $\mathcal{G}'$(the complement of $\mathcal{G}$) is simply the original allocation of only one interviewer for each EA in the block. This corresponds to $\underset{\sim}{F} = \underset{\sim}{I}$, using the notation above.

Specifying a particular interviewer allocation design according to the above principles is equivalent to specifying the design parameters $k,\ell$, and $\underset{\sim}{F}$. The optimization of k and $\ell$ is considered in Section 6. The optimal choice of $\underset{\sim}{F}$ is not considered here; however, in most cases, the choice is governed more by operational considerations than by precision concerns. Historically, $\underset{\sim}{F} = k^{-1}\underset{\sim}{1}_{kk}$, equal allocation, has been the most administratively convenient interpenetration method since it simply specifies that the secondaries be dealt out to each interviewer in turn until the list of units for the EA is exhausted. This can be handled clerically in the field office. Further, the classical design is similar to the completely balanced one-way components of variance design which have certain optimal properties. Our formulas provide for general $\underset{\sim}{F}$, however, for situations where completely balanced interpenetration is not possible.

Certain generalizations are feasible. The role of the primary stage can be taken over by any lower stage. The procedures and conditions in the foregoing are then restated substituting "primary unit" by "next-to-last stage unit" and "secondary unit" by "last-stage unit." No restrictions on the survey design will be made for stages above the next-to-last stage so that now, unequal probability sampling is permitted for every stage except the last-stage.

## 4. TWO ESTIMATORS OF INTERVIEWER VARIANCE

In this section, we will derive two estimators of the interviewer component, $\sigma_b^2$, which are estimable from the interviewer allocation design of the previous section. The independence of the estimators, proved formally in Appendix C, guarantees an estimator having smaller variance can be formed as an appropriate linear combination of the two. The variance formulas for the estimators which are used in the study of their performance in Section 6, are also given.

To ensure the unbiasedness of the estimators under model (2.3), the following additional assumptions must be made:

1. For a given sample of primary units, $\Pi$, and set of Pattern I blocks $\zeta$, the design matrices $\underset{\sim}{X}$, $\underset{\sim}{U}_b$, and $\underset{\sim}{W}_{pa}$ are constant. This is equivalent to assuming that the secondary sample size, $m_p$, for primary p is predetermined for every primary in the population,

that the design vector $(k,\ell,\underset{\sim}{F})$ does not depend upon $\Pi$ or $\wp$, and the delineation of EA boundaries and block groupings determined for a sample $\Pi$ is independent of the set $\wp$.

2. The selection of secondaries and their assignment to interviewers within primaries (Pattern I blocks) and the assignment of EA's to interviewers (Pattern II blocks) is done by implementing simple random sampling. This requirement may only be approximated in practice. For example, systematic sampling from geographically sorted lists may be used rather than simple random sampling for Pattern I blocks. For Pattern II, interviewers may be assigned to EA's purposively based on administrative convenience. These deviations are tolerated in order to reduce the costs of studies since their effects are believed to be small.

In Biemer (1978), a general method for estimating $\sigma_b^2$ based on the synthesis or MIVQUE0 procedure (Hartley, Rao, and La Motte) was developed for multistage sampling without replacement. This method is briefly reviewed here under the simplifying assumption of negligible sampling fractions within EA's. Biemer has shown that even when sampling fractions are not ignored, the formulas for estimating $\sigma_b^2$ are the same as below.

Given the samples $\Pi$ and $\wp$, the model (2.3) represents a mixed analysis of variance model of the form

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \sum_{q=1}^{I+1} \underset{\sim}{U}_q \underset{\sim}{b}_q \tag{4.1}$$

where $\underset{\sim}{X},\underset{\sim}{U}_1,\ldots,\underset{\sim}{U}_{I+1}$ are design matrices, $\underset{\sim}{\beta}$ is a vector of constants, $\underset{\sim}{b}_q$ is a vector of random variables with $E(\underset{\sim}{b}_q) = \underset{\sim}{0}$ and $V(\underset{\sim}{b}_q) = \sigma_q^2\underset{\sim}{I}$. Then $\underset{\sim}{U}_1$ plays the role of $\underset{\sim}{U}_b$ from model (2.3), $\underset{\sim}{U}_2,\ldots,\underset{\sim}{U}_{I+1}$ play the role of the $\underset{\sim}{W}_{pa}$'s, and $\underset{\sim}{b}_1$, $\underset{\sim}{b}_2,\ldots,\underset{\sim}{b}_{I+1}$ represent $\underset{\sim}{b}$ and $\underset{\sim}{e}_{pa}$, $p=1,\ldots,n$; $a=1,\ldots,I_p$, respectively.

Assume that $\underset{\sim}{\varepsilon}$ has been reparameterized so that $\underset{\sim}{X}'\underset{\sim}{X} = \underset{\sim}{I}$ and define

$$Q_q(\underset{\sim}{y}) = \underset{\sim}{y}'\underset{\sim}{A}_q\underset{\sim}{y} \tag{4.2}$$

where, for q=1,...,I+1,

$$\underset{\sim}{A}_q = (\underset{\sim}{U}_q - \underset{\sim}{X}\underset{\sim}{X}'\underset{\sim}{U}_q)(\underset{\sim}{U}_q - \underset{\sim}{X}\underset{\sim}{X}'\underset{\sim}{U}_q)'. \tag{4.3}$$

Biemer (1978) has shown that an unbiased estimator of $\underset{\sim}{\sigma}^2 = [\sigma_q^2]$ is given by

$$\hat{\underset{\sim}{\sigma}}^2 = \underset{\sim}{\Lambda}^{-1}\underset{\sim}{Q} \tag{4.4}$$

where

$$\underset{\sim}{\Lambda} = [\lambda_{qs}] \text{ for } q,s = 1,\ldots,I+1 \tag{4.5}$$

with

$$\lambda_{qs} = \text{tr } \underset{\sim}{U}_s'\underset{\sim}{A}_q\underset{\sim}{U}_s \tag{4.6}$$

and $\underset{\sim}{Q} = [Q_q]$. Using the estimability conditions of Biemer, it can be verified that $\underset{\sim}{\Lambda}^{-1}$ exists for survey designs and interviewer allocation schemes considered above. Finally, we see from (4.4) that

$$\sigma_1^2 = \underset{\sim}{\lambda}_o'\underset{\sim}{Q}, \tag{4.7}$$

with $\underset{\sim}{\lambda}_o$ the first row of $\underset{\sim}{\Lambda}^{-1}$, is an estimator of $\sigma_b^2$. Let $\sigma_b^2(1)$ denote this estimator.

Note that the quadratic forms (4.2) appear to involve all the observations $\underset{\sim}{y}$, i.e., those in interpenetrated EA's as well as non-interpenetrated EA's. It can be shown, however, that the observations in Pattern II blocks contribute no information for the estimator $\hat{\sigma}_b^2(1)$, a fact we shall make use of in Appendix B. To see this, let

$$\dot{\underset{\sim}{y}} = \ddot{\underset{\sim}{X}}\dot{\underset{\sim}{\beta}} + \sum_{q=1}^{\cdot+1} \dot{\underset{\sim}{U}}_q\dot{\underset{\sim}{b}}_q \tag{4.8}$$

denote the analagous model to (4.1) where the design vectors and matrices correspond only to observations of the $\mathcal{J} = \ell k$ interviewers in Pattern I EA's. It can be shown that $\Lambda$ defined in (4.5) can be written as $\Lambda = \mathrm{diag}[\dot{\Lambda}, I]$ with $\dot{\Lambda}$ for the model (4.8) defined in analogy to (4.5). This leads to the result that the estimator of $\hat{\sigma}_b^2(1)$ is the same regardless if (4.1) or (4.8) is used.

Let $E_3$ and $V_3$ denote the conditional expectation and variance operators given particular choices of $\Pi$ and $\xi$, and let $E_2$ and $V_2$ denote the operators over all possible selections of $\xi$ holding $\Pi$ fixed, and let $E_1$ and $V_1$ denote the operators over all possible samples $\Pi$. If the $y$ are assumed to be normally distributed, the variance of the estimator $\hat{\sigma}_b^2(1)$ is

$$V(1) = \mathrm{Var}\ \hat{\sigma}_b^2(1) = E_1 E_2(\lambda_o' V_3(Q)\lambda_o), \tag{4.9}$$

where $V_3(Q) = [2\mathrm{tr}\ A_q \Sigma A_s \Sigma]$, and $\Sigma = V_3(y) = \sum_{q=1}^{I+1} U_q U_q' \sigma_q^2$, which follows since $E_3(y')A_q E_3(y) = 0$.

Now consider a second estimator of $\sigma_b^2$, which will be denoted by $\hat{\sigma}_b^2(2)$. Let the EA blocks formed in Section 3 be labeled by $\gamma = 1,\ldots,L$ and let $\bar{y}_\gamma' = (\bar{y}_{\gamma 1},\ldots,\bar{y}_{\gamma k})$ denote the $k \times 1$ vector of EA sample means for block $\gamma$. Define the quadratic forms

$$B_\gamma(y) = \bar{y}_\gamma' \Phi \bar{y}_\gamma \tag{4.10}$$

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

$$\Phi = I - FF^- \tag{4.11}$$

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

Let

$$\bar{B}_I = \frac{1}{\ell} \sum_{\gamma \epsilon :} B\gamma(y) \tag{4.12}$$

and

$$\bar{B}_{II} = \frac{1}{L-\ell} \sum_{\gamma \in \mathcal{C}'} B_\gamma(\underset{\sim}{y}). \qquad (4.13)$$

From (4.1), the following relationship holds:

$$\bar{\underset{\sim}{y}}_\gamma = \bar{\underset{\sim}{\eta}}_\gamma + \underset{\sim}{F}\underset{\sim}{b}_\gamma + \bar{\underset{\sim}{e}}_\gamma \text{ for every } \gamma \in \mathcal{C}, \qquad (4.14)$$

and

$$\bar{\underset{\sim}{y}}_\gamma = \bar{\underset{\sim}{\eta}}_\gamma + \underset{\sim}{I}\underset{\sim}{b}_\gamma + \bar{\underset{\sim}{e}}_\gamma \text{ for every } \gamma \in \mathcal{C}', \qquad (4.15)$$

where $\bar{\underset{\sim}{\eta}}_\gamma$ is the k-vector of EA population means in block $\gamma$, $\underset{\sim}{b}_\gamma$ is the k-vector of interviewer variables, $b_i$, for the k interviewers of block $\gamma$, and $\underset{\sim}{e}_\gamma' = (e_{\gamma 1},\ldots,e_{\gamma k})$ where $e_{\gamma t}$ is sample mean of the composite error terms, $e_{pas}$, for the t-th EA in block $\gamma$.

Then, for $\gamma \in \mathcal{C}$,

$$E_3(B_\gamma(\underset{\sim}{y})) = \bar{\underset{\sim}{\eta}}_\gamma' \underset{\sim}{\Phi} \bar{\underset{\sim}{\eta}}_\gamma + \text{tr } \underset{\sim}{\Phi}\underset{\sim}{F}\underset{\sim}{F}^- \sigma_b^2$$

$$+ \text{tr } \underset{\sim}{\Phi} V_3(\bar{\underset{\sim}{e}}_\gamma) \qquad (4.16)$$

or, since from (4.11) tr $\underset{\sim}{\Phi}\underset{\sim}{F}\underset{\sim}{F}^- = 0$

$$= \bar{\underset{\sim}{\eta}}_\gamma' \underset{\sim}{\Phi} \bar{\underset{\sim}{\eta}}_\gamma + \text{tr } \underset{\sim}{\Phi} V(\bar{\underset{\sim}{e}}_\gamma) \qquad (4.17)$$

$$= H_\gamma(\Pi, \mathcal{C}), \text{ say.} \qquad (4.18)$$

Since tr $\Phi$ = k-r where r = rank $\underset{\sim}{F}$, we have for $\gamma \in \mathcal{C}'$

$$E_3(B_\gamma(\underset{\sim}{y})) = H_\gamma(\Pi, \mathcal{C}) + (k-r)\sigma_b^2. \qquad (4.19)$$

Thus, from (4.18) and (4.19),

$$E_2(\bar{B}_{II} - B_I) = (k-r)\sigma_b^2. \qquad (4.20)$$

It follows that, for k > r,

$$\hat{\sigma}_b^2(2) = \frac{1}{k-r}(\bar{B}_{II} - \bar{B}_I) \qquad (4.21)$$

is an unbiased estimator of $\sigma_b^2$ given $\Pi$ and is, therefore, unconditionally unbiased.

The variance of $\hat{\sigma}_b^2(2)$ can also be derived under the assumption of normality of $\underset{\sim}{y}$. It is shown in Appendix A that

$$\begin{aligned} V(2) &= \text{Var } \hat{\sigma}_b^2(2) \\ &= E_1 \left\{ \frac{1}{(k-r)(L-\ell)} \left\{ \frac{1}{(k-r)} \left[ \frac{L}{\ell} \left(S_H^2(\Pi)+\overline{V}(\Pi)\right) \right.\right.\right. \\ &\quad \left.\left.\left. + \; 4\sigma_b^2 \, \overline{H}(\Pi) \right] + 2\sigma_b^4 \right\}\right\} \end{aligned} \tag{4.22}$$

where $S_H^2(\Pi)$, $\overline{H}(\Pi)$, and $\overline{V}(\Pi)$ are given by (A.3) through (A.5) and (A.10).

In Appendix C, the independence of the estimators $\hat{\sigma}_b^2(1)$ and $\hat{\sigma}_b^2(2)$ is established for normally distributed $\underset{\sim}{y}$. Thus, for our model, it follows that the unbiased estimator

$$\hat{\sigma}_b^2 = \alpha \, \hat{\sigma}_b^2(1) + (1-\alpha) \, \hat{\sigma}_b^2(2) \tag{4.23}$$

for

$$\alpha = V(2)/(V(1) + V(2)) \tag{4.24}$$

has variance

$$\mathcal{V} = V(1)V(2)/(V(1) + V(2)). \tag{4.25}$$

$\mathcal{V}$ is never greater than the smaller of V(1) and V(2).

## 5. SPECIAL CASE: SIMPLE RANDOM SAMPLING

In this section we show the form of the estimators $\hat{\sigma}_b^2(1)$ and $\hat{\sigma}_b^2(2)$ and their variances for the special case where $\underset{\sim}{F} = k^{-1} \underset{\sim}{1}_{kk}$, the balanced interpenetration scheme, and where each of the $\hat{\wp}_{pa}$'s initially assigned to one interviewer constitutes a simple random sample of m units with negligible sampling fraction.

Let $y_{\gamma tjs}$ denote the s-th observation recorded by the j-th interviewer in the t-th EA of the $\gamma$-th block. A dot in place of a subscript denotes summing over that subscript; e.g., $y_{\gamma\cdot j\cdot} = \sum_{t=1}^{k} \sum_{s=1}^{m/k} y_{\gamma tjs}$.

It is shown in Appendix B that

$$\hat{\sigma}_b^2(1) = \frac{1}{m\ell[k(m-2)+1]} \left[\frac{k(m-1)}{(k-1)m} \sum_{\gamma=1}^{\ell} \sum_{j=1}^{k} (y_{\gamma\cdot j\cdot} - \bar{y}_{\gamma\cdots})^2 - \sum_{\gamma=1}^{\ell} \sum_{t=1}^{k} \sum_{j=1}^{k} \sum_{s=1}^{f} (y_{\gamma tjs} - \bar{y}_{\gamma t\cdot\cdot})^2\right] \tag{5.1}$$

where $\bar{y}_{\gamma\cdots} = y_{\gamma\cdots}/k$, $\bar{y}_{\gamma t\cdot\cdot} = y_{\gamma t\cdot\cdot}/m$, and $f = m/k$ is assumed to be integral for simplicity. It is also shown there that

$$V(1) = \frac{2}{\ell(k-1)} \{[\sigma_b^4 + \frac{2}{n} \sigma_b^2 \sum_{\gamma t}\sum \sigma_e^2(\gamma,t)] + \frac{2(m-1)}{k\ell} \left(\frac{k}{m}\right)^2 [\frac{m-1}{k-1} \sum_t \{[\Sigma\sigma_e^2(\gamma,t)]^2 - \sum_t \sigma_e^4(\gamma,t)\}]\}. \tag{5.2}$$

Further simplification results when we assume $\sigma_e^2(\gamma,t) = \sigma_e^2$. Then

$$V(1) = \frac{2}{\ell(k-1)} [\sigma_b^4 + \frac{2}{m} \sigma_b^2 \sigma_e^2 + \frac{k(m-1)}{m^2(km-2k+1)} \sigma_e^4]. \tag{5.3}$$

Using $\underset{\sim}{F}^{-} = k^{-1} \underset{\sim}{1}_{kk}$, we have from (4.10) through (4.13) and (4.19) that

$$\hat{\sigma}_b^2(2) = (k-1)^{-1} [\ell^{-1} \sum_{\gamma\in\mathcal{S}} \sum_{t=1}^{k} (\bar{y}_{\gamma t\cdot\cdot} - \bar{\bar{y}}_{\gamma})^2 - (L-\ell)^{-1} \sum_{\gamma\in\bar{\mathcal{S}}} \sum_{t=1}^{k} (\bar{y}_{\gamma t\cdot\cdot} - \bar{\bar{y}}_{\gamma})^2] \tag{5.4}$$

where $\bar{\bar{y}}_{\gamma} = y_{\gamma\cdots}/km$. Then from (4.22)

$$V(2) = [(k-1)(L-\ell)]^{-1} \{(k-1)^{-1}[(L\ell)(S_H^2+\overline{V}) + 4\sigma_b^2\overline{H}] + 2\sigma_b^4\} \tag{5.5}$$

where from (A.10), (A.4), and (A.3) respectively

$$\overline{V} = E_1\overline{V}(\Pi) = L^{-1} \sum_{\gamma=1}^{L} \{ \frac{2}{k^2m^2} [(k-2) \sum_{t=1}^{k} \sigma_e^4(\gamma,t) + (\sum_{t=1}^{k} \sigma_e^2(\gamma,t))^2]$$

$$+ m^{-1} \sum_{t=1}^{k} (n_{\gamma t} - \overline{n}_\gamma)^2 \sigma_e^2(\gamma,t)\};$$

$$\overline{H} = E_1\overline{H}(\Pi) = L^{-1} \sum_{\gamma=1}^{L} \sum_{t=1}^{k} [(n_{\gamma t} - \overline{n}_\gamma)^2 + (k-1)\, \sigma_e^2(\gamma,t)/km];$$

$$S_H^2 = E_1 S_H^2(\Pi).$$

Assuming $\sigma_e^2(\gamma,t) = \sigma_e^2$ we have

$$V(2) = \frac{1}{(k-1)(L-\ell)} \{\frac{1}{(k-1)} [(L/\ell)(S_H^2 + 4(k-1)\sigma_e^4/km^2 + \sigma_e^2\, \overline{v}/m)$$

$$+ 4\sigma_b^2(\overline{v} + (k-1)\sigma_e^2/m)] + 2\sigma_b^4\} \tag{5.6}$$

where $\overline{v} = \sum_{\gamma=1}^{L} \sum_{t=1}^{k} (n_{\gamma t}-\overline{n}_\gamma)^2/L = \sum_{\gamma=1}^{L} v_\gamma/L$ and (5.7)

$$S_H^2 = \sum_{\gamma=1}^{L} (v_\gamma - \overline{v})^2/(L-1). \tag{5.8}$$

For the important case of k = 2, or interpenetration of pairs of assignments, also considered by Fellegi (1974), we have from (5.1)

$$\hat{\sigma}_b^2(1) = \sum_{\gamma\epsilon\ell} [(m-1)(y_{\gamma\cdot 1\cdot} - y_{\gamma\cdot 2\cdot})^2/m$$

$$- \sum_{t=1}^{2} \sum_{j=1}^{2} \sum_{s=1}^{m/2} (y_{\gamma tjs} - \overline{y}_{\gamma t\cdot\cdot}]/\ell m(2m-3)$$

and from (5.4)

$$\hat{\sigma}_b^2(2) = \ell^{-1} \sum_{\gamma \in \mathcal{P}} (\bar{y}_{\gamma 1 \cdot\cdot} - \bar{y}_{\gamma 2 \cdot\cdot})^2/2 - (L-\ell)^{-1} \sum_{\gamma \in \bar{\mathcal{P}}} (\bar{y}_{\gamma 1 \cdot\cdot} - \bar{y}_{\gamma 2 \cdot\cdot})^2/2.$$

These estimators are precisely those given by Fellegi (1974, eqs. (3.8) and (3.12)), when his weights, $W_k$, were all identical and equal to their average value. In his paper, Fellegi speculates on the independence of the two estimators and their relative variances. Their independence under the assumption of normality is proved in Appendix C and their relative efficiency can be explored using the variance formulas developed in this section.

## 6. OPTIMALITY CRITERIA

In this section, the optimal design of interviewer variance studies is considered analytically using the variance formulas developed in the last section. In addition, data from the 1980 Census of Population and Housing are used to compute the terms in the variance formulas in order to empirically investigate other aspects of optimal design which do not lend themselves to an analytical study. We assume for this section that $\underline{y}$ has a multivariate normal distribution so the variance formulas and the independence of the two estimators apply.

One can show by considering each term of (5.3) separately and holding the number of interpenetrated interviewer assignments, $\mathcal{J} = k\ell$, fixed that V(1) decreases slowly as k increases if $m > 2$. The same can be shown to be true of V(2) from (5.5). In other words, the greater the number of EA's to be grouped per block, the more precise are the estimators. For most applications, however, there is a practical limitation to the size of k. For personal interviewing, the major disadvantage is the increased travel costs as k increases and the interviewers' assignments are dispersed over larger geographic areas.

One of our goals is to determine the values of k and $\ell$ that minimize the variance of $\hat{\sigma}_b^2$ when the additional cost of interpenetration is taken into account. Our second goal is to establish criteria for determining the efficiency of using the composite estimator (4.23) instead of the estimator $\hat{\sigma}_b^2(1)$ alone.

Suppose that costs of an interviewer evaluation study are a function of $\mathcal{J}$, the number of interpenetrated interviewer assignments. Further assume that interpenetration of a block containing k EA's will increase the usual cost of interviewing (or enumerating) in that block by a factor of $\sqrt{k}$. (Justification: The average distance between randomly distributed points in a plane is increased by $\sqrt{k}$ when the density of those points is decreased by a factor of k. Thus if the interviewers' assignments were randomly distributed and if the additional cost of interpenetration were due to within-block travel, this assumption would be reasonable.) Therefore, the added cost of interpenetrating $\ell$ blocks of k EA's each is $c\mathcal{J}\sqrt{k} - c\mathcal{J} = c\mathcal{J}(\sqrt{k}-1)$, where c is the usual per-EA cost of interviewing.

The optimal allocation of k and $\ell$ for an interviewer variance estimator is defined as that which minimizes the variance of the estimator for a fixed evaluation budget, or, equivalently for a fixed proportional increase in cost per EA i.e., $c\mathcal{J}(\sqrt{k}-1)/c\mathcal{J} = (\ell/L)(\sqrt{k}-1) = \Delta_o$.

We denote (5.3) and (5.6) by $V1(k|\Delta_o)$ and $V2(k|\Delta_o)$ respectively, upon substituting $\ell = L\Delta_o/(\sqrt{k}-1)$ and $L = \mathcal{J}/k$. Then it can be shown by minimizing over k that:

1. $V1(k|\Delta_o)$ is monotonically increasing for $k > 2$ if $m > 2$.
2. $V2(k|\Delta_o)$ is monotonically decreasing for $k > 2$.

Thus if $\hat{\sigma}_b^2(1)$ were to be used alone, the optimal choice of k would be 2, whereas if $\hat{\sigma}_b^2(2)$ were to be used alone, one should choose k as large as possible. The optimal allocation for the composite estimator, however, must be addressed empirically since it depends on the configuration of the other parameters in the variance formulas.

The variability of EA means within block $\gamma$ is reflected by $v_\gamma$, and the mean and variance over all blocks of the $v_\gamma$'s is given by $\bar{v}$ and $S_H^2$ in (5.7) and (5.8). The larger is either of these parameters, the worse is the estimator $\hat{\sigma}_b^2(2)$. Therefore, to improve the efficiency of the composite estimator, attempts should be made to keep the blocks as homogeneous as possible (small $\bar{v}$) or at least equally homogeneous (small $S_H^2$).

Fellegi (1974) speculates that the arithmetic mean of the two estimators $\hat{\sigma}_b^2(1)$ and $\hat{\sigma}_b^2(2)$ is likely to be an improvement on either. This will not be the case if $\bar{v}$ and/or $S_H^2$ are large and dominate V(2). In fact, even the optimal linear combination of the two estimators given by (4.23) and (4.24) will sometimes be virtually no improvement over $\hat{\sigma}_b^2(1)$, even when $\ell/L$ is quite small. (This effect is exemplified in the numerical results that follow.) When $S_H^2$ and $\bar{v}$ are dominated by the other terms in the variance expression, the composite estimator can be a substantial improvement. From (4.25) we see that the relative efficiency of $\hat{\sigma}_b^2(1)$ to to $\hat{\sigma}_b^2$, denoted by $RE(\hat{\sigma}_b^2(1), \hat{\sigma}_b^2)$, is $V(2)/(V(1) + V(2))$. In the extreme case where $S_H^2 = \bar{v} = 0$ and $k = 2$, one can show from (5.3) and (5.6) that $RE(\hat{\sigma}_b^2(1), \hat{\sigma}_b^2)$

$$= (\ell/L) \frac{(L/\ell)+2(1-\rho)/\rho+[(1-\rho)/\rho]^2}{[2-(\ell/L)][1+(2m-3)^{-1}]+2[(1-\rho)/\rho]+[(1-\rho)/\rho]^2} \tag{6.1}$$

where

$$\rho = \sigma_b^2/(\sigma_e^2 + \sigma_b^2) \tag{6.2}$$

is the intra-interviewer assignment correlation coefficient defined by Kish (1962).

In order to illustrate some principles of optimal interviewer variance study design, data from the 1980 U.S. Census of Population and Housing was used to estimate the parameters $\sigma_e^2$, $S_H^2$ and $\overline{v}$ for a number of demographic and housing variables. About 1400 Census EA's in a random sample of 57 Census Districts in the New York and Boston Census regions were paired together clerically using Census maps to form 700 mutually exclusive pairs of contiguous EA's. Table 1 reports the parameters $\overline{v}/\sigma_e^2$ and $S_H^2/\sigma_e^2$ for several characteristics. As would be expected, the variable "Sex" shows the least amount of heterogeneity within blocks and the housing characteristics show the most.

(Insert Table 1 about here)

The variance formulas developed in the previous section were not intended to be used with variables of this type, since they are categorical. It is not known how robust the variance expressions are to deviations from normality. However the numerical results that follow are useful to describe the behavior of the variances of the estimators and the optimal choice of k for various configurations of the parameters $\sigma_b^2$, $\sigma_e^2$, $S_H^2$, and $\overline{v}$.

Figures A and B show graphs of $V1(k|\Delta_o)$, $V2(k|\Delta_o)$, and $V(k|\Delta_o)$ as functions of k for selected variables and categories from Table 1. A typical value of $\rho$ for the variable shown is used (obtained from Kish (1962)) and $\Delta_o$ is set to .05 which is equivalent to interpenetrating about 12% of the blocks. Several characteristics about the curves are illustrated by these figures:

1. The precision of the composite estimator for fixed cost is virtually unaffected by the size of k. Thus a choice of k = 2 is probably best from the point of view of simplicity in block formation and minimizing within-block heterogeneity.

2. The larger is k, the greater is the improvement in precision of the composite estimator over $\hat{\sigma}_b^2(1)$.

(insert figures A through C about here)

In Table 1, the range of RE($\hat{\sigma}_b^2$, $\hat{\sigma}_b^2(1)$) is shown for each variable category which results when $\rho$ is in the interval (.005,.05) and when k = 2, m = 125, and $\Delta_o$ = .05. It can be shown and is illustrated by these examples that:

1. Characteristics having the smallest within-block heterogeneity show the most gain (i.e., small RE($\hat{\theta}_b^2$, $\hat{\sigma}_b^2(1)$)) from using the composite estimator over $\hat{\sigma}_b^2(1)$.
2. RE($\hat{\sigma}_b^2$, $\hat{\sigma}_b^2(1)$) is a decreasing function of $\rho$.

The behavior of the curves was investigated for values of $\Delta_o$ between .01 and .10 and m between 50 and 500. Large values of $\Delta_o$ will moderate the benefit of using the composite estimator for variables having large within-block variability and/or $\rho$. Large values of m have the effect of increasing RE($\hat{\sigma}_b^2$, $\hat{\sigma}_b^2(1)$) for the opposite set of variables, those having small within-block variability and/or $\rho$. For none of the variables described in Table 1, however, would a change in $\Delta_o$ or m within the ranges considered affect RE($\hat{\sigma}_b^2$, $\hat{\sigma}_b^2(1)$) substantially enough to change the decision of whether to use the composite estimator.

## 7. SUMMARY

This investigation into the properties of the estimators $\hat{\sigma}_b^2(1)$, $\hat{\sigma}_b^2(2)$, and $\hat{\sigma}_b^2$ has led to several observations.

1. The composite estimator can be a substantial improvement over the usual estimator of interviewer variance $\hat{\sigma}_b^2(1)$. The characteristics which show the most benefit are those having large interviewer correlation $\rho$ and/or homogeneous block groupings.

2. For personal interviewing, the number of EA's grouped per block, k, has little effect on the precision of the composite estimator.
3. The precision of $\hat{\sigma}_b^2(2)$ and $\hat{\sigma}_b^2$ is adversely affected by heterogeneity of EA means within blocks (large $\overline{v}$) and by discrepancies in this heterogeneity between blocks (large $S_H^2$). Thus care should be exercised in formation of the blocks if use of the composite estimator is contemplated.

## APPENDIX A

To derive (4.22), note that

$$V(2) = V_1E_2E_3(\hat{\sigma}_b^2(2)) + E_1V_2E_3(\hat{\sigma}_b^2(2)) + E_1E_2V_3(\hat{\sigma}_b^2(2)) \qquad (A.1)$$

$$= E_1[V_2E_3(\hat{\sigma}_b^2(2)) + E_2V_3(\hat{\sigma}_b^2(2))]$$

since the first term of (A.1) is zero from (4.18). From (4.16) and (4.17),

$$E_3(\hat{\sigma}_b^2(2)) = \{[\sum_{\gamma\varepsilon\varphi'} H_\gamma(\Pi,\varphi) + (k-r)\sigma_b^2]/(L-\ell) - \sum_{\gamma\varepsilon\varphi} H_\gamma(\Pi,\varphi)/\ell\}/(k-r).$$

Thus

$$V_2E_3(\hat{\sigma}_b^2(2)) = [(\ell/L)S_H^2(\Pi)/(L-\ell) + ((L-\ell)/L)S_H^2(\Pi)/\ell$$

$$+ S_H^2(\Pi)/L]/(k-r)^2 \qquad (A.2)$$

where

$$S_H^2(\Pi) = \sum_{\gamma=1}^{L} [H_\gamma(\Pi) - \overline{H}(\Pi)]^2/(L-1), \qquad (A.3)$$

$$H_\gamma(\Pi) = E_2H_\gamma(\Pi,\varphi), \qquad (A.4)$$

and

$$\overline{H}(\Pi) = \sum_{\gamma=1}^{L} H_\gamma(\Pi)/L. \qquad (A.5)$$

To determine $V_3(\hat{\sigma}_b^2(2)) = V_3(\overline{B}_{II} - \overline{B}_I)/(k-r)^2$, we find from (4.14) and (4.15),

$$V_3(\overline{B}_I) = \sum_{\gamma\varepsilon\varphi} V_3[B_\gamma(\underset{\sim}{y})]/\ell^2$$

$$= \sum_{\gamma\varepsilon\varphi} [2\ tr(\underset{\sim}{\Phi}V_3(\overline{\underset{\sim}{e}}_\gamma))^2 + 4\overline{\underset{\sim}{n}}_\gamma'\underset{\sim}{\Phi}V_3(\overline{\underset{\sim}{e}}_\gamma)\underset{\sim}{\Phi}\overline{\underset{\sim}{n}}_\gamma]/\ell^2$$

$$= \sum_{\gamma\varepsilon\varphi} V_\gamma(\Pi,\varphi)/\ell^2; \qquad (A.6)$$

$$V_3(\overline{B}_{II}) = 2(k-r)\sigma_b^4/(L-\ell) + \sum_{\gamma \in \mathcal{S}} [V_\gamma(\Pi,\mathcal{S})$$

$$+ 4\sigma_b^2 H_\gamma(\Pi,\mathcal{S})]/(L-\ell)^2, \tag{A.7}$$

and

$$\mathrm{Cov}_3(\overline{B}_I,\overline{B}_{II}) = 0. \tag{A.8}$$

Thus from (A.6), (A.7), and (A.8),

$$E_2V_3(\hat{\sigma}_b^2(2)) = \{[2(k-r)\sigma_b^4 + \overline{V}(\Pi) + 4\sigma_b^2\overline{H}(\Pi)]/(L-\ell)$$

$$+ \overline{V}(\Pi)/\ell\}/(k-r)^2 \tag{A.9}$$

where

$$\overline{V}(\Pi) = \sum_{\gamma=1}^{L} E_2V_\gamma(\Pi,\ )/L. \tag{A.10}$$

Combining (A.1), (A.2), and (A.9) yields (4.22).

APPENDIX B

To confirm (5.1) and (5.2), we return to the notation of model (4.8). We define the left-hand Kronecker product as $\underset{\sim}{A} \otimes \underset{\sim}{B} = [\underset{\sim}{A}b_{ij}]$ where $\underset{\sim}{B} = [b_{ij}]$. Then $\dot{\underset{\sim}{X}} = \underset{\sim}{1}_m \otimes \underset{\sim}{I}_\ell$; $\dot{\underset{\sim}{U}}_1 = \underset{\sim}{1}_f \otimes [\underset{\sim}{I}_k \cdots \underset{\sim}{I}_k]_{mk\times k} \otimes \underset{\sim}{I}_\ell$, with $f = m/k$; and $\dot{\underset{\sim}{U}}_i' = [\underset{\sim}{0}_m, \ldots, \underset{\sim}{I}_m, \ldots, \underset{\sim}{0}_m]$, $i=1,\ldots,\ell$, with the identity matrix in the $(i-1)^{th}$ position. From (4.3), $\underset{\sim}{A}_1 = \underset{\sim}{1}_{ff} \otimes (\underset{\sim}{I}_k - k^{-1}\underset{\sim}{1}_{kk}) \otimes \underset{\sim}{I}_\ell$ and $\underset{\sim}{A}_i = (\underset{\sim}{0}_m, \ldots, \underset{\sim}{I}_m - m^{-1}\underset{\sim}{1}_{mm}, \ldots, \underset{\sim}{0}_m)$, $i=2,\ldots,\ell$, with the non-zero matrix in the $(i-1)^{th}$ position.

Then from (4.2), we have

$$Q_1(\underset{\sim}{y}) = \sum_{\gamma=1}^{\ell} \sum_{j=1}^{k} (y_{\gamma.j.} - \bar{y}_{\gamma...})^2 \tag{B.1}$$

and

$$Q_q(\underset{\sim}{y}) = \sum_{j=1}^{k} \sum_{s=1}^{f} (y_{\gamma tjs} - \bar{y}_{\gamma t..})^2 \text{ for } q = k(\gamma-1)+a+1, \tag{B.2}$$

with $\bar{y}_{\gamma...} = y_{\gamma...}/k$ and $\bar{y}_{\gamma t..} = y_{\gamma t..}/m$. From (4.5) and (4.6), one can show that

$$\underset{\sim}{\Lambda} = \left[\begin{array}{c|c} \ell m^2(k-1) & (k-1)f\underset{\sim}{1}_{\ell}' \\ \hline (k-1)f\underset{\sim}{1}_{\ell} & (m-1)\underset{\sim}{I}_{\ell} \end{array}\right]$$

Thus

$$\underset{\sim}{\Lambda}^{-1} = \left[\begin{array}{c|c} \dfrac{k(m-1)}{(k-1)\ell m^2[k(m-2)+1]} & \delta\underset{\sim}{1}' \\ \hline \delta\underset{\sim}{1}_{\ell} & \beta\underset{\sim}{1}_{\ell\ell} + (\alpha-\beta)\underset{\sim}{I}_{\ell} \end{array}\right] \tag{B.3}$$

where $\delta = -1/m\ell[k(m-2)+1]$, $\alpha = \{1 + \frac{(k-1)}{k\ell[k(m-2)+1]}\}/(m-1)$,

and $\beta = (k-1)/(m-1)k\ell[k(m-2)+1]$.

From (4.7) and (B.1) through (B.3), we have $\hat{\sigma}_b^2(1)$ as given in (5.1). From (4.9), V(1) is shown to be given by (5.2).

APPENDIX C

We shall establish 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:

$$\mathrm{Cov}(\hat{\sigma}_b^2(1), \hat{\sigma}_b^2(2)) = \mathrm{Cov}_1 E_2(\hat{\sigma}_b^2(1)), E_2(\hat{\sigma}_b^2(2))$$

$$+ E_1 \mathrm{Cov}_2(\hat{\sigma}_b^2(1), \hat{\sigma}_b^2(2))$$

where $\mathrm{Cov}_2$ denotes covariance given the samples $\Pi$ and $\mathcal{F}$, $\mathrm{Cov}_1$ denotes covariance over all possible samples $\Pi$ and $\mathcal{F}$ and $E_1$, $E_2$ are analogously defined. Since $E_2(\hat{\sigma}_b^2(1)) = \sigma_b^2$, the first term on the right is zero. We shall show that

$$\mathrm{Cov}_2(\hat{\sigma}_b^2(1), \hat{\sigma}_b^2(2)) = 0 \qquad \text{(C.1)}$$

Given $\Pi$ and $\mathcal{F}$, $\boldsymbol{\lambda}_o$ in (4.7) is fixed since it depends only on the design matrices $\mathbf{X}$ and $\mathbf{U}_q$, $q=1,\ldots,I+1$. Furthermore, $\hat{\sigma}_b^2(2)$ defined in (4.19), is a linear combination of the $B_\gamma(\mathbf{y})$, $\gamma=1,\ldots,L$, whose coefficients are fixed given the sets $\Pi$ and $\mathcal{F}$. Therefore, (C.1) is true if the $Q_q(\mathbf{y})$ given by (4.2) are uncorrelated with the $B_\gamma(\mathbf{y})$ given by (4.7).

Assume the vector $\mathbf{y}$ is ordered by block and within block by EA's. Let $m_{\gamma t}$ denote the number of units sampled in the t-th EA of block $\gamma$. Define the block diagonal matrix

$$\mathbf{T} = \mathrm{diag}\{m_{\gamma t}^{-1} \mathbf{1}_{m_{\gamma t}}\}$$

ordered as for $\mathbf{y}$. Then, the vector of EA means for block $\gamma$, can be rewritten as

$$\bar{\mathbf{y}}_\gamma = \mathbf{G}_\gamma \mathbf{T}' \mathbf{y}$$

where $\mathbf{G}_\gamma = [g_{\gamma ti}]$ with $g_{\gamma ti} = 1$ if the t-th element of $\bar{\mathbf{y}}_\gamma$ corresponds to the i-th EA of the vector $\mathbf{y}$ and 0 otherwise. Thus, $B_\gamma(\mathbf{y})$ can be rewritten as

$$B_\gamma(\underset{\sim}{y}) = \underset{\sim}{y}' \, \underset{\sim}{T} \, \underset{\sim}{G}_\gamma' \, \underset{\sim}{\Phi} \, \underset{\sim}{G}_\gamma \, \underset{\sim}{T}' \, \underset{\sim}{y}.$$

Assume that the conditional distribution of $\underset{\sim}{y}$ given $\Pi$ and $\varphi$ is normal with mean $\underset{\sim}{\mu} = \underset{\sim}{X} \, \underset{\sim}{\bar{\eta}}$ and variance $\underset{\sim}{\Psi} = \sum_{q}^{I+1} \underset{\sim}{U}_q \, \underset{\sim}{U}_q' \, \sigma_q^2$.

Then the $Q_q(\underset{\sim}{y})$ and the $B_\gamma(\underset{\sim}{y})$ are independent if and only if

$$\underset{\sim}{T} \, \underset{\sim}{G}_\gamma' \, \underset{\sim}{\Phi} \, \underset{\sim}{G}_\gamma \, \underset{\sim}{T}' \, \underset{\sim}{\Psi} \, \underset{\sim}{A}_q = 0 \tag{C.2}$$

(see, for example, Searle 1971, p.59) which is satisfied if

$$\underset{\sim}{T} \, \underset{\sim}{G}_\gamma' \, \underset{\sim}{\Phi} \, \underset{\sim}{G}_\gamma \, \underset{\sim}{T}' \, \underset{\sim}{U}_q \, \underset{\sim}{U}_q' \, \underset{\sim}{A}_q = 0 \tag{C.3}$$

for all q and $\gamma$.

Because of the ordering of $\underset{\sim}{y}$, $\underset{\sim}{U}_q$, for q=2,...,I+1, is given by $\underset{\sim}{U}_q' = (\underset{\sim}{0}_1, \ldots \underset{\sim}{I}_{q-1}, \ldots, \underset{\sim}{0}_I)$ where the component matrices are square and are of the same dimension as the non-zero portion of the corresponding column in $\underset{\sim}{T}$. From (4.3), it follows that $\underset{\sim}{A}_q \, \underset{\sim}{X} = 0$ which implies by the similar form of $\underset{\sim}{T}$ to $\underset{\sim}{X}$ that

$$\underset{\sim}{T}' \, \underset{\sim}{U}_s \, \underset{\sim}{U}_s' \, \underset{\sim}{A}_q = 0$$

for s=2,...,I+1, q=1,...,I+1.

Now consider $\underset{\sim}{U}_1$. Clearly, the non-null columns of $\underset{\sim}{G}_\gamma \, \underset{\sim}{T}' \, \underset{\sim}{U}_1$ make up the columns of $\underset{\sim}{F}$. Hence, $\underset{\sim}{\Phi} \, \underset{\sim}{G}_\gamma \, \underset{\sim}{T}' \, \underset{\sim}{U}_1 = 0$ and, consequently (C.3) follows from the form of $\underset{\sim}{\Phi}$ in (4.8).

It can be shown that if rank $\underset{\sim}{F} = r = k$(full rank), the only matrix $\Phi$ satisfying (C.2) is the null matrix. Hence, $r < k$ is a necessary condition (C.2) to hold.

1. Parameters of Selected Characteristics and Performance of $\sigma_b^2$ for $k = 2$, $\Delta_o = .05$, $m = 125$, $\rho = .005$ and $.05$

| | $S_H/\sigma_e^2$ | $\overline{V}/\sigma_e^2$ | $RE(\hat{\sigma}_b^2, \hat{\sigma}_b^2(1))$ $\rho = .005$ | $\rho = .05$ |
|---|---|---|---|---|
| Sex | | | | |
| Male or Female | .015 | .003 | .70 | .44 |
| Race | | | | |
| White | .124 | .020 | .99 | .82 |
| Black | .079 | .001 | .97 | .68 |
| Marital Status | | | | |
| Single | .066 | .011 | .96 | .66 |
| Married | .041 | .013 | .91 | .58 |
| Divorced | .012 | .005 | .68 | .45 |
| Plumbing Facilities | | | | |
| All or Lack Complete | .110 | .021 | .98 | .79 |
| Units/Address | | | | |
| 1 | .321 | .127 | .99 | .89 |
| 2 - 9 | .226 | .080 | .99 | .87 |
| 10+ | .379 | .086 | .99 | .90 |

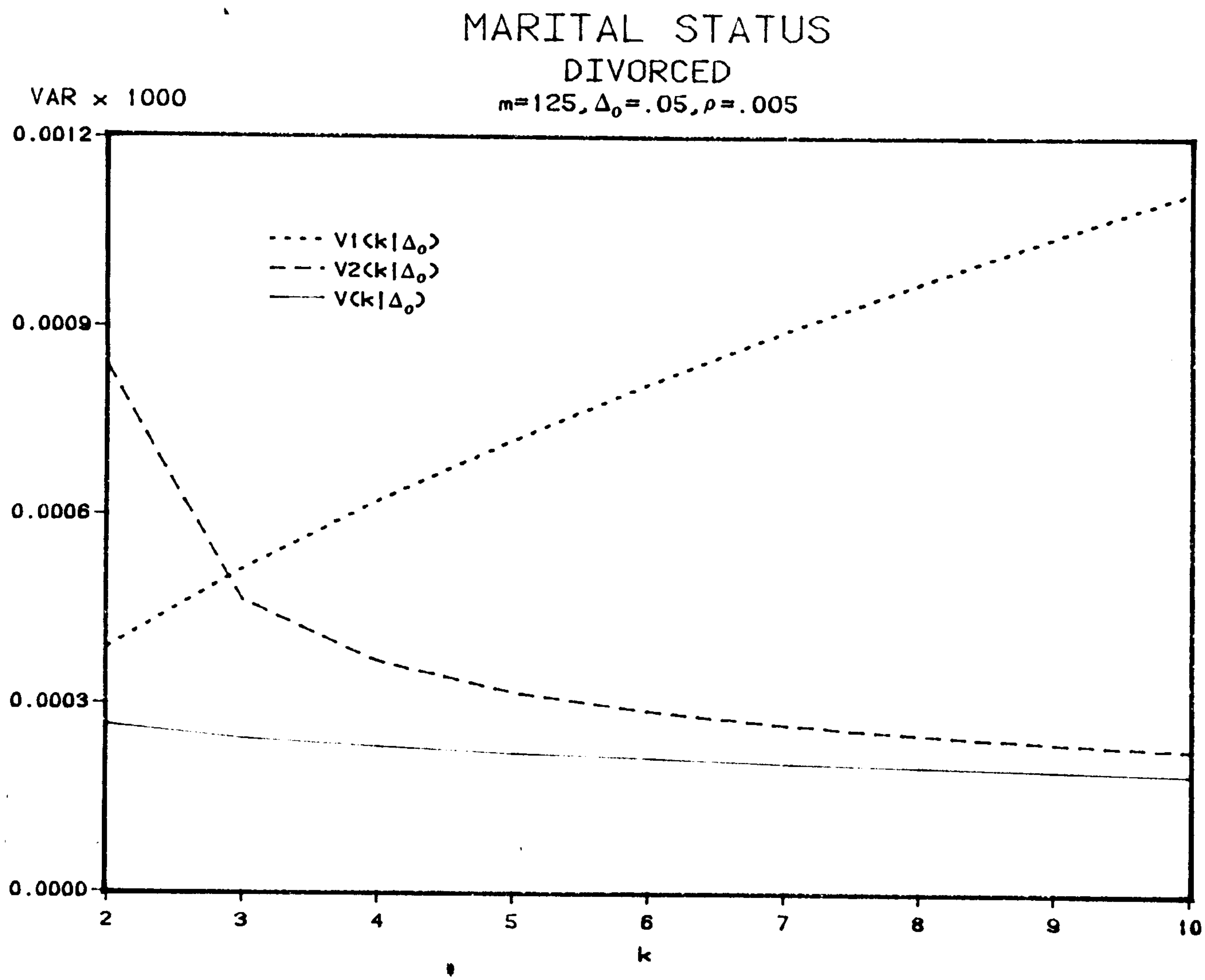
MARITAL STATUS
DIVORCED
m=125, Δo=.05, ρ=.005
VAR × 1000
0.0012
0.0009
0.0006
0.0003
0.0000
V1(k|Δo)
V2(k|Δo)
V(k|Δo)
2
3
4
5
6
7
8
9
10
k

Figure A Variances of $\sigma_b^2(1)$, $\sigma_b^2(2)$, and $\sigma_b^2$ for characteristic "Marital Status - Divorced."

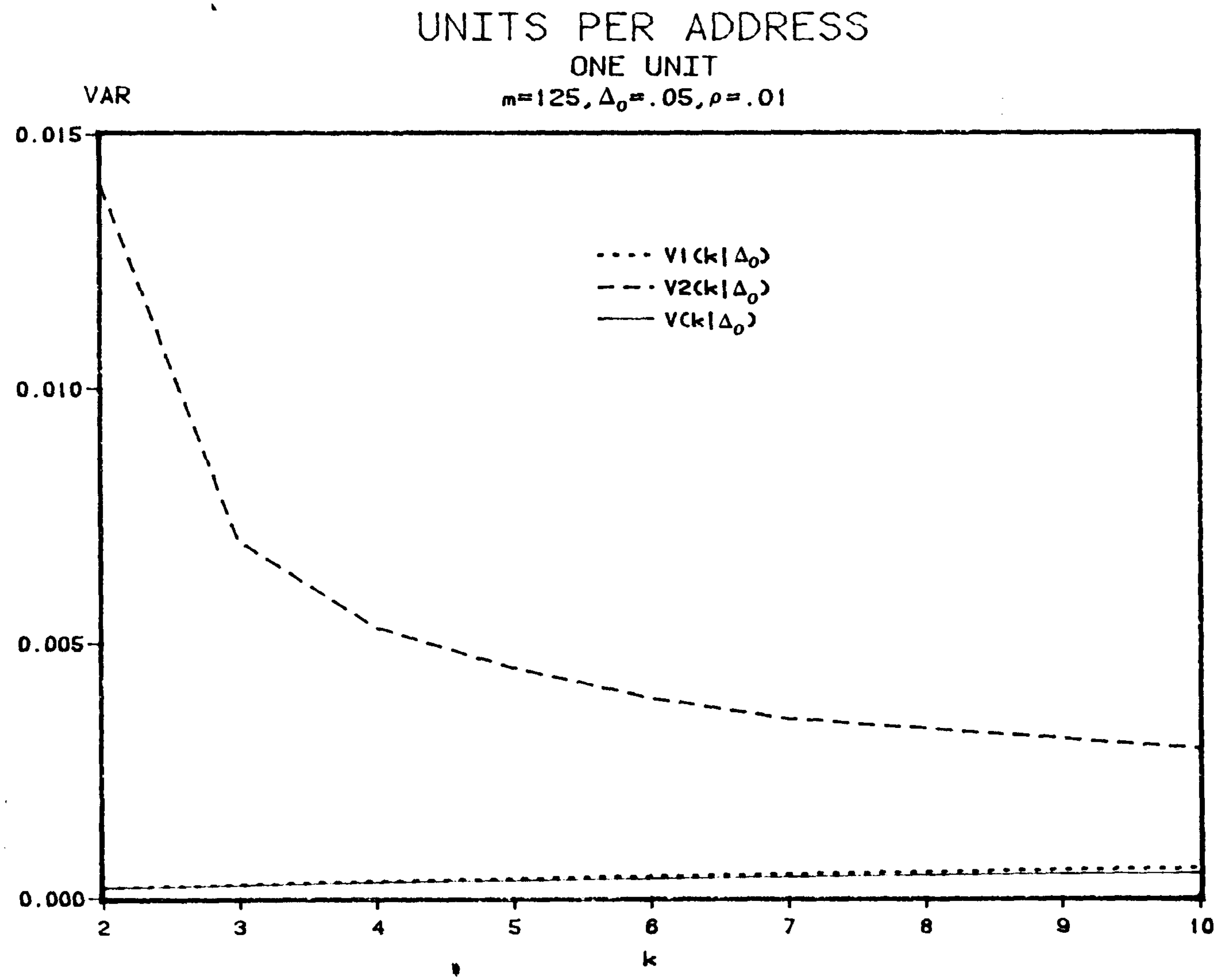
UNITS PER ADDRESS
ONE UNIT
m=125, Δo=.05, ρ=.01
VAR
0.015
0.010
0.005
0.000
V1(k|Δo)
V2(k|Δo)
V(k|Δo)
2
3
4
5
6
7
8
9
10
k

Figure B Variances of $\sigma_b^2(1)$, $\sigma_b^2(2)$, and $\sigma_b^2$ for characteristic "Number of Units - One Unit."

# REFERENCES

Bailey, L., Moore, T.F., and Bailar, B.A. (1978). "An Interviewer Variance Study for the Eight Impact Cities of the National Crime Survey Cities Sample," Journal of the American Statistical Association, 73, 16-23.

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. (1977). Sampling Techniques, John Wiley & Sons, New York.

Dalenius, T. (1977). Bibliography on non-sampling errors in surveys - I(A to G), II (H to Q), III (R to Z), International Statistical Review, 45, No. 1 71-89, No. 2 18197, No. 3 303-317.

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

Hansen, Morris H., Hurwitz, William N., and Bershad, Max A.. "Measurement Errors in Censuses and Surveys," Bulletin of the International Statistical Institute, 38 (August 1959), 359-74.

Hanson, Robert H. and Marks, E.S. (1958). "Influence of the Interviewer on the Accuracy of Survey Results," Journal of the American Statistical Association, 53, 635-655.

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

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

Kish, Leslie (1962). "Studies of Interviewer Variance for Attitudinal Variables," Journal of the American Statistical Association, 57,92-115.

Sukhatme, P.V. and Sukhatme, B.V. (1970). Sampling Theory of Surveys with Applications, The Iowa State University Press, Ames, Iowa.