BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES
SRD Research Report Number: Census/SRD/RR-87/17

# THE EFFECTS OF MEASUREMENT ERRORS ON VARIANCE ESTIMATION

by

Lawrence R. Ernst and Kirk M. Wolter
Statistical Research Division
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: August 8, 1987

Report issued: August 8, 1987

## 1. INTRODUCTION

In Wolter (1985) the problem of estimating variances when the data are contaminated by measurement (or response) errors was considered for linear estimators. Under a simple additive error model it was shown that design-unbiased variance estimators are in general biased as estimators of total variance and that in certain circumstances this bias can be important. It was also shown that with additional conditions a random group variance estimator can shift the bias entirely to the sampling error component, generally with an accompanying reduction in the total variance.

This paper builds on the work in Wolter, extending and generalizing it in several ways. It is first shown that by viewing the variance estimator as a general quadratic function of the responses, an estimator can always be obtained with bias independent of the response error. Unfortunately, the residual terms in the bias can be large. However, with additional conditions that are more general than those considered in conjunction with the random group estimator in Wolter, a variance estimator is obtained which removes the bias due to response error and also yields a total bias that is typically reasonably small.

The key results just described are presented in Section 2. In Section 3 it is shown that the results on the random group variance estimator in the presence of measurement errors presented in Wolter are a special case of the results in Section 2 and that there are important situations where only the more general results are applicable. In Section 4 the random group estimator results are extended to the jackknife and balanced half-sample methods of estimating variances. Finally, in Section 5, the extension of this work to nonlinear estimators is considered. It is demonstrated, by example, that the asymptotic results in terms of sample size that hold for sampling variance do not in general hold for total variance in the presence of measurement errors. In particular, it is shown that this difficulty occurs with the Taylor series method, even when a variance estimator exists for the Taylor series approximation with bias (as an estimator of the variance of the approximation) independent of the response error. Situations for which such an estimator of variance is asymptotically unbiased are also illustrated.

## 2. PRINCIPAL RESULTS

To establish a framework for the work to be presented in this paper, we first review the notation and terminology employed in Appendix D of Wolter (1985). It is assumed that the response, say $Y_i$, in a population of size N is adequately described by the additive error model,

$$Y_i = \mu_i + e_i, \qquad i=1,\ldots,N. \tag{2.1}$$

The errors $e_i$ are assumed to be $(0, \sigma_i^2)$ random variables and the means $\mu_i$ are taken to be the "true values."

We assume it is desired to estimate some parameter $\theta$ of the finite population with an estimator $\hat{\theta}$ of the form

$$\hat{\theta} = \sum_{i=1}^{N} w_i t_i Y_i, \tag{2.2}$$

where the $w_i$ are fixed weights attached to the units in the population, the $t_i$ are indicator random variables,

$$\begin{aligned} t_i &= 1 \quad \text{if } i \in s \\ &= 0 \quad \text{if } i \notin s, \end{aligned}$$

and s denotes the sample.

We let $E_d$ and $Var_d$ denote the expectation and variance operators with respect to the sampling design; $E$ and $\mathit{Var}$ are these operators with respect to the distribution, say $\xi$, of the measurement (or response) errors; and finally unsubscripted E and Var denote the total expectation and variance.

It is established in Wolter that

$$\mathrm{Var}(\hat{\theta}) = \sum_{i=1}^{N} w_i^2 \mu_i^2 \pi_i (1 - \pi_i) + \sum_{i \neq j}^{N}\sum w_i w_j \mu_i \mu_j (\pi_{ij} - \pi_i \pi_j)$$

$$+ \sum_{i=1}^{N} w_i^2 \pi_i \sigma_i^2 + \sum_{i \neq j}^{N}\sum w_i w_j \pi_{ij} \sigma_{ij}, \qquad (2.3)$$

where $\pi_i$ denotes the probability that the i-th unit is drawn into the sample, $\pi_{ij}$ the probability that both the i-th and j-th units are drawn into the sample, and $\sigma_{ij} = \mathcal{E}(\varepsilon_i \varepsilon_j)$. The sum of the first two terms of (2.3) is the sampling variance, $\mathrm{Var}_d\ \mathcal{E}(\hat{\theta})$, and the sum of the remaining two terms is the response variance, $E_d\ \mathcal{V}ar(\hat{\theta})$.

It is also established in Wolter that if

$$\tilde{\theta} = \sum_{i=1}^{N} w_i t_i \mu_i$$

is an estimator with the same functional form as $\hat{\theta}$ with means $\mu_i$ replacing the responses $Y_i$, and $v(\tilde{\theta})$ is a design unbiased estimator of the design-variance of $\tilde{\theta}$, and the variance estimator $v_c(\hat{\theta})$ of $\mathrm{Var}(\hat{\theta})$ is obtained by replacing the $\mu_i$ in $v(\tilde{\theta})$ by the responses $Y_i$, then

$$\mathrm{Bias}\ [v_c(\hat{\theta})] = -\ \mathcal{V}ar\ E_d(\hat{\theta})$$

$$= -\sum_{i=1}^{N} w_i^2 \pi_i^2 \sigma_i^2 - \sum_{i \neq j}^{N}\sum w_i w_j \pi_i \pi_j \sigma_{ij}. \qquad (2.4)$$

With srs wor and $\pi$ps sampling as illustrations, it is shown that this bias can be important in some situations. For example, for $\pi$ps sampling with $\hat{\theta} = \hat{Y}$, the Horvitz-Thompson estimator of the population total, Bias $[v_c(\hat{\theta})]$ is independent of the sample size and generally of order $N^2$ as a function of the population size. Additional assumptions are then presented for which a random group variance estimator of $Var(\hat{\theta})$ is obtained with bias arising solely from the sampling distribution, not the $\xi$-distribution , and which for many common situations has a substantially smaller total bias than in (2.4). The basis of these assumptions is that the correlated component of response error arises strictly from the effect of interviewers. The specific assumptions follow:

(a) There are k random groups of equal size and identical distributions.

(b) Interviewer assignments are completely nested within random groups.

(c) Interviewers have a common effect on the $\xi$-distribution , i.e.,

$$E(e_i) = 0 ;$$

$$E(e_i^2) = \sigma_i^2 ;$$

$$E(e_i e_j) = \sigma_{ij} \quad \text{if units } i \text{ and } j \text{ are enumerated by the same interviewer;}$$

$$= 0 \quad \text{if units } i \text{ and } j \text{ are interviewed by different interviewers;}$$

and these moments do not depend on which interviewer enumerates the i-th and j-th units.

In this section more general additional assumptions than (a)-(c) are considered, together with a general class of variance estimators that includes the random group estimator. It is shown that for appropriate choice of a variance estimator from this class, the bias arises solely from the sampling distribution and is typically reasonably small. However, to motivate the need for the additional assumptions, we first, under the simple assumptions that lead to (2.3) and (2.4), consider the following class of estimators of $Var(\hat{\theta})$:

$$v(\hat{\theta}) = \sum_{i \varepsilon s} a_i w_i^2 Y_i^2 + \sum_{i \neq j \varepsilon s}\sum b_{ij} w_i w_j Y_i Y_j, \tag{2.5}$$

where the $a_i$ and $b_{ij}$ are fixed coefficients associated with the i-th sample unit and the (i,j)-th pair respectively. Then, since

$$E(Y_i^2) = \mu_i^2 + \sigma_i^2 \tag{2.6}$$

and

$$E(Y_i Y_j) = \mu_i \mu_j + \sigma_{ij},$$

it follows that

$$E[v(\hat{\theta})] = E(E_d[v(\hat{\theta})]) = E(\sum_{i=1}^{N} a_i w_i^2 Y_i^2 \pi_i + \sum_{i \neq j}^{N}\sum b_{ij} w_i w_j Y_i Y_j \pi_{ij})$$

$$= \sum_{i=1}^{N} a_i w_i^2 \mu_i^2 \pi_i + \sum_{i \neq j}^{N}\sum b_{ij} w_i w_j \mu_i \mu_j \pi_{ij}$$

$$+ \sum_{i=1}^{N} a_i w_i^2 \pi_i \sigma_i^2 + \sum_{i \neq j}^{N}\sum b_{ij} w_i w_j \pi_{ij} \sigma_{ij}, \tag{2.7}$$

which together with (2.3) yield

$$\text{Bias } [v(\hat{\theta})] = \sum_{i=1}^{N} w_i^2 \mu_i^2 \pi_i (a_i - 1 + \pi_i)$$

$$+ \sum_{i \neq j}^{N}\sum w_i w_j \mu_i \mu_j (b_{ij}\pi_{ij} - \pi_{ij} + \pi_i \pi_j)$$

$$+ \sum_{i=1}^{N} w_i^2 \pi_i (a_i - 1)\, \sigma_i^2$$

$$+ \sum_{i \neq j}^{N}\sum w_i w_j \pi_{ij}(b_{ij} - 1)\, \sigma_{ij}. \tag{2.8}$$

In general, it is not possible to make the entire bias expression (2.8) equal zero. However, either the first or the third term, and either the second or the fourth term can be removed from this expression by the appropriate choice of $a_i$ and $b_{ij}$. That is, the first and third term in (2.8) would drop out with $a_i = 1 - \pi_i$ and $a_i = 1$ respectively, while the second and the fourth would be removed with $b_{ij} = (\pi_{ij} - \pi_i\pi_j)/\pi_{ij}$ and $b_{ij} = 1$ respectively. In particular, with $a_i = 1 - \pi_i$ and $b_{ij} = (\pi_{ij} - \pi_i\pi_j)/\pi_{ij}$, (2.8) reduces to (2.4), while with $a_i = 1$, $b_{ij} = 1$,

$$\begin{aligned} \text{Bias } [v(\hat{\theta})] &= \sum_{i=1}^{N} w_i^2 \mu_i^2 \pi_i^2 + \sum_{i \neq j}^{N}\sum w_i w_j \mu_i \mu_j \pi_i \pi_j \\ &= \Big(\sum_{i=1}^{N} w_i \mu_i \pi_i\Big)^2, \end{aligned} \tag{2.9}$$

which is independent of the response variance. Unfortunately, despite the fact that the response variance component of the bias has been removed in (2.9), this expression is typically quite large. To illustrate, for the $\pi$ps sampling example considered earlier, $w_i = 1/\pi_i$ and consequently,

$$\text{Bias } [v(\hat{\theta})] = \Big(\sum_{i=1}^{N} \mu_i\Big)^2,$$

which typically is of order $N^2$, as it was in (2.4). Furthermore, (2.9) is not directly a function of the sampling error. In particular, a small sampling variance, as would occur if the sample size was fixed and the quantities $w_i\mu_i$ did not vary much, would not generally imply that (2.9) is small.

The difficulty, illustrated by (2.4) and (2.9), in attempting to obtain a variance estimator with bias that is both independent of the response error and reasonably small can be viewed as algebraically arising from the fact that under the conditions that lead to (2.3), no more than two of the four terms in (2.8) can be removed. The additional assumptions (a)-(c) allow more control of the bias of the variance estimator, accounting for the results on the bias of the random group estimator in Wolter. We now proceed to consider the following more general additional conditions, which will allow for similar reductions in the bias of the variance estimator. It is assumed that each ordered pair of sample units (i,j), $i \neq j$, falls into one of two sets, U and C. As is illustrated below, a pair need not be in the same set for all samples or even for a particular sample. In fact, the only assumptions in this regard are that (i,j) and (j,i) are in the same set and that

$$\alpha_{ij} > 0, \quad \text{where } \alpha_{ij} = P((i,j)\varepsilon U | i,j \text{ are in sample}). \tag{2.10}$$

The other assumptions are that for each (i,j),

$$E(e_i|(i,j)\varepsilon U) = E(e_i|(i,j)\varepsilon C) = 0, \tag{2.11}$$

$$E(e_i e_j|(i,j)\varepsilon U) = 0. \tag{2.12}$$

We also let

$$\sigma'_{ij} = E(e_i e_j|(i,j)\varepsilon C) \tag{2.13}$$

and note that (2.3) still holds, where now

$$\sigma_{ij} = \sigma'_{ij}(1 - \alpha_{ij}). \tag{2.14}$$

Before explaining how these conditions enable us to obtain variance estimators with generally smaller biases, we illustrate the rather abstract formulation of these conditions by considering the situation where the assumptions (a)-(c) hold. Then conditions (2.10)-(2.12) also hold if C and U are taken to be the sample pairs in the same random group and different random groups respectively. In Section 3 we will discuss further examples where the conditions (2.10)-(2.12) are met.

We now consider the following modification of the variance estimator (2.5), for which the coefficients corresponding to the pair $(i,j)$ depend on whether $(i,j)\varepsilon U$ or $C$. Let

$$v'(\hat{\theta}) = \sum_{i\varepsilon s} a_i w_i^2 Y_i^2 + \sum_{(i,j)\varepsilon U} b_{ij1} w_i w_j Y_i Y_j$$

$$+ \sum_{(i,j)\varepsilon C} b_{ij2} w_i w_j Y_i Y_j, \tag{2.15}$$

where the $a_i$, $b_{ij1}$ and $b_{ij2}$ are all constants. Then, using the relations (2.15), (2.6), (2.14),

$$E(Y_i Y_j | (i,j)\varepsilon U) = \mu_i \mu_j,$$

which follows from (2.11) and (2.12), and

$$E(Y_i Y_j | (i,j)\varepsilon C) = \mu_i \mu_j + \sigma'_{ij},$$

which follows from (2.11) and (2.13), we obtain

$$E[v'(\hat{\theta})] = E(E_d[v'(\hat{\theta})]) = \sum_{i=1}^{N} a_i w_i^2 \; E(Y_i^2)\pi_i$$

$$+ \sum_{i\neq j}^{N}\sum b_{ij1} w_i w_j \; E(Y_i Y_j | (i,j)\varepsilon U)\alpha_{ij}\pi_{ij}$$

$$+ \sum_{i\neq j}^{N}\sum b_{ij2} w_i w_j \; E(Y_i Y_j | (i,j)\varepsilon C)(1 - \alpha_{ij})\pi_{ij}$$

$$= \sum_{i=1}^{N} a_i w_i^2 \mu_i^2 \pi_i + \sum_{i\neq j}^{N}\sum w_i w_j \mu_i \mu_j [b_{ij1}\alpha_{ij} + b_{ij2}(1-\alpha_{ij})]\pi_{ij}$$

$$+ \sum_{i=1}^{N} a_i w_i^2 \pi_i \sigma_i^2 + \sum_{i\neq j}^{N}\sum b_{ij2} w_i w_j \pi_{ij} \sigma_{ij}. \tag{2.16}$$

Finally, (2.16) and (2.3) yield

$$\text{Bias }[v'(\hat{\theta})] = \sum_{i=1}^{N} w_i^2 \mu_i^2 \pi_i (a_i - 1 + \pi_i)$$

$$+ \sum_{i \neq j}\sum w_i w_j \mu_i \mu_j [b_{ij1}\alpha_{ij}\pi_{ij} + b_{ij2}(1-\alpha_{ij})\pi_{ij} - \pi_{ij} + \pi_i \pi_j]$$

$$+ \sum_{i=1}^{N} w_i^2 \pi_i (a_i - 1)\sigma_i^2 + \sum_{i \neq j}^{N}\sum w_i w_j \pi_{ij}(b_{ij2} - 1)\sigma_{ij}. \qquad (2.17)$$

The additional set of coefficients in (2.17) in comparison with (2.8) is what allows for greater control over the bias of the variance estimator. For example, the second and fourth terms of (2.17) can now both be made to drop out with

$$b_{ij1} = 1 - \frac{\pi_i \pi_j}{\alpha_{ij}\pi_{ij}}, \; b_{ij2} = 1 .$$

If additionally, $a_i = 1 - \pi_i$, then the first term will drop out and

$$\text{Bias }[v'(\hat{\theta})] = -\sum_{i=1}^{N} w_i^2 \pi_i^2 \sigma_i^2, \qquad (2.18)$$

while if $a_i = 1$, then

$$\text{Bias }[v'(\hat{\theta})] = \sum_{i=1}^{N} w_i^2 \mu_i^2 \pi_i^2. \qquad (2.19)$$

In the latter case, the portion of the bias arising from the response variance is eliminated. Furthermore, if the $\mu_i$ are nonnegative, (2.19) cannot exceed (2.9), and in the $\pi$ps example considered previously, the bias would now be of order N, instead of $N^2$. However, (2.19) still suffers from the fact that like (2.9), a small sampling variance does not necessary result in a small bias. In the remainder of this section it will be demonstrated how this drawback can be overcome by a different choice of coefficents while still eliminating the portion of the bias arising from the response variance. The results obtained will directly generalize the results of Wolter, as will be explained in Section 3.

To obtain the appropriate coefficients, first note that in order to eliminate the response variance portion of the bias it is required that for all i, j,

$$a_i = 1 \text{ and } b_{ij2} = 1, \tag{2.20}$$

but this does not restrict the $b_{ij1}$. To obtain the appropriate values for the $b_{ij1}$, we first let n denote the sample size and $n_1$, $n_2$ the number of pairs (i,j) in U and C respectively. For now, we consider the case where n, $n_1$ and $n_2$ are the same for all possible samples. For example, this would hold if the correlated component of response error arises strictly from the effect of interviewers; the number of interviews each interviewer conducts is fixed, although not necessarily the same for all interviewers; and C and U are the sample pairs interviewed by the same interviewer and different interviewers respectively. Now, if in addition to n being fixed the quantities $w_i \mu_i$ are the same for all i, then clearly $\mathrm{Var}_d\, E(\hat{\theta}) = 0$. For $v'(\hat{\theta})$ to be unbiased when these conditions and (2.20) hold, it suffices for (2.15) to be 0 with the $Y_i$ replaced by $\mu_i$ since Bias $[v'(\hat{\theta})]$ is independent of the response error. However, this is equivalent to

$$n + \sum_{(i,j)\varepsilon U} b_{ij1} + n_2 = 0;$$

this relation is satisfied if for all (i,j)

$$b_{ij1} = -\frac{n + n_2}{n_1} = 1 - \frac{n^2}{n_1}, \tag{2.21}$$

where the last equality follows since $n_2 = n(n - 1) - n_1$. The special case of $v'(\hat{\theta})$ for which (2.20) and (2.21) hold is denoted $v''(\hat{\theta})$, that is

$$v''(\hat{\theta}) = \sum_{i \varepsilon s} w_i^2 Y_i^2$$

$$+ \sum_{(i,j)\varepsilon U} (1 - \frac{n^2}{n_1}) w_i w_j Y_i Y_j + \sum_{(i,j)\varepsilon C} w_i w_j Y_i Y_j . \qquad (2.22)$$

Thus $v''(\hat{\theta})$ is an unbiased estimater of $\mathrm{Var}(\hat{\theta})$ if $n$, $n_1$, $n_2$ are fixed and the quantities $w_i \mu_i$ are the same for all i in the population. Consequently, if the $w_i \mu_i$ do vary, but the variability is sufficiently small, then the bias of $v''(\hat{\theta})$ is small, as desired. Furthermore, if $n$, $n_1$ or $n_2$ vary, but $(n + n_2)/n_1$ is fixed, then (2.21) is still fixed and hence $v''(\hat{\theta})$ is defined. An example of $v''(\theta)$ with variable n will be presented in Section 3.

From (2.17), (2.20) and (2.21) it follows that in general

$$\mathrm{Bias}\ [v''(\hat{\theta})] = \sum_{i=1}^{N} w_i^2 \mu_i^2 \pi_i^2$$

$$- \sum_{i \neq j}^{N}\sum w_i w_j \mu_i \mu_j (\frac{n^2}{n_1} \alpha_{ij} \pi_{ij} - \pi_i \pi_j) . \qquad (2.23)$$

An important special case of (2.22) occurs when $n$, $n_1$ and $n_2$ are fixed and the $\alpha_{ij}$ are the same for all pairs $(i,j)$; as for example if C is the set of all pairs interviewed by the same interviewer and each interviewer interviews the same number of sample cases, which are randomly distributed among the interviewers. In this case, for all $(i,j)$,

$$\alpha_{ij} = \frac{n_1}{n(n-1)} ,$$

and (2.23) reduces to

$$\mathrm{Bias}\ [v''(\hat{\theta})] = \sum_{i=1}^{N} w_i^2 \mu_i^2 \pi_i^2$$

$$- \sum_{i \neq j}^{N}\sum w_i w_j \mu_i \mu_j (\frac{n}{n-1} \pi_{ij} - \pi_i \pi_j) . \qquad (2.24)$$

An application of (2.24) will be presented in the next section.

## 3. COMPARISON WITH RESULTS IN WOLTER (1985)

It will be demonstrated here that the key results in Wolter (1985), in particular Theorem D.4, on the use of a random group estimator in the presence of measurement errors, can be considered a particular case of the results at the end of the previous section. It will also be illustrated that the results in this paper can be applied in some important situations where the results in Wolter are not applicable.

In Wolter, under assumptions (a)-(c), the random group variance estimator

$$v_{RG}(\hat{\theta}) = k^{-1}(k - 1)^{-1} \sum_{\alpha=1}^{k} (\hat{\theta}_\alpha - \hat{\theta})^2$$

$$= 2^{-1}k^{-2}(k - 1)^{-1} \sum_{\alpha \neq \beta}^{k}\sum (\hat{\theta}_\alpha - \hat{\theta}_\beta)^2 \qquad (3.1)$$

is considered, where

$$\hat{\theta}_\alpha = \sum_{i=1}^{N} kw_i t_{i(\alpha)} Y_i \qquad (3.2)$$

and

$$t_{i(\alpha)} = 1 \quad \text{if the i-th unit is included in } \alpha\text{-th random group}$$

$$= 0 \quad \text{otherwise.} \qquad (3.3)$$

It is shown there, as Theorem D.4, that with these assumptions,

$$\text{Bias } [v_{RG}(\hat{\theta})] = \sum_{i=1}^{N} w_i^2 \mu_i^2 \pi_i^2 - \sum_{i \neq j}^{N}\sum w_i w_j \mu_i \mu_j (k\upsilon_{j|i}\pi_{ij} - \pi_i \pi_j), \qquad (3.4)$$

where $\upsilon_{j|i}$ is the conditional probability that unit j is included in random group $\beta$, given that unit i is included in random group $\alpha$ ($\alpha \neq \beta$) and that both i and j are in the parent sample.

We demonstrate that $v_{RG}(\hat{\theta})$ is a special case of $v''(\hat{\theta})$, and (3.4) is the corresponding special case of (2.23). Let C and U be the set of ordered pairs of sample units (i,j) for which i and j are in the same random group and different groups respectively. Then since the random groups are each of size n/k (where n is not assumed to be constant), it follows that

$$n_2 = n(\frac{n}{k} - 1)$$

and hence

$$n_1 = n(n - 1) - n_2 = \frac{n^2(k-1)}{k}.$$

We substitute this last relation in (2.22) and (2.23), obtaining

$$v''(\hat{\theta}) = \sum_{i \varepsilon S} w_i^2 Y_i^2 - \frac{1}{k-1} \sum_{(i,j) \varepsilon U} w_i w_j Y_i Y_j + \sum_{(i,j) \varepsilon C} w_i w_j Y_i Y_j \qquad (3.5)$$

and

$$\text{Bias}[v''(\hat{\theta})] = \sum_{i=1}^{N} w_i^2 \mu_i^2 \pi_i^2 - \sum_{i \neq j}^{N}\sum w_i w_j \mu_i \mu_j (\frac{k}{k-1} \alpha_{ij} \pi_{ij} - \pi_i \pi_j). \quad (3.6)$$

Now, if the terms in (3.1) are expanded and collected, (3.1) reduces to (3.5). Furthermore, because the random groups are identically distributed, $\alpha_{ij} = (k-1)\,\nu_{j|i}$, and thus (3.4) and (3.6) are equivalent.

If in addition to conditions (a)-(c), n is fixed, then the conditions for (2.24) hold, and thus

$$\text{Bias}\ [v_{RG}(\hat{\theta})] = \sum_{i=1}^{N} w_i^2 \mu_i^2 \pi_i^2 - \sum_{i \neq j}^{N}\sum \mu_i \mu_j w_i w_j \ (\frac{n}{n-1} \pi_{ij} - \pi_i \pi_j).$$

To illustate, example D.6 of Wolter, for $\pi$ps sampling, is a special case of this last relation for which $\pi_i = 1/w_i$, and hence

$$\text{Bias }[v_{RG}(\hat{\theta})] = \sum_{i=1}^{N} \mu_i^2 - \sum_{i \neq j}^{N}\sum \mu_i \mu_j \left(\frac{n}{n-1} \frac{\pi_{ij}}{\pi_i \pi_j} - 1\right).$$

Having established that the results in Section 2 generalize the results of Wolter, we now demonstrate that there are important situations where the former results but not the latter are applicable. Among the assumptions in Wolter is the rather stringent condition (c), which requires that the first and second moments on the $\xi$-distribution do not depend upon which interviewer enumerates which units. In many circumstances this is an unrealistic assumption. In its place, in this paper are the less restrictive conditions (2.11) and (2.12). For example, if the interviewer assignments are of equal size, the sample units are randomly distributed among the interviewers, and C and U are the set of distinct sample pairs interviewed by the same and different interviewers respectively, then (2.11) holds even if $E(e_i) \neq 0$ for each interviewer, as long as the expected value of this error is 0 averaged over all interviewers; while (2.12) is just a formal statement for the assumption that the correlated component arises strictly from the effect of interviewers.

Consider also the case when the interviewer assignments are not of equal size. Then with C and U as above, (2.11) and (2.12) would still hold, but in general only if we added back the condition that $E(e_i) = 0$ for each interviewer. (This is because $P((i,j) \in U)$ is no longer independent of which interviewer interviews the i-th unit.) However, there are still advantages to using $v''(\hat{\theta})$ as opposed to $v_{RG}(\hat{\theta})$ in this case. If the random group estimator were to be used, it would be necessary to combine interviewer assignments in order to meet the requirement of equal sized random groups, and even then this condition might be only approximately met. Furthermore, because interviewer assignments would be combined in the random group estimator, the precision of $v_{RG}(\hat{\theta})$ would generally be lower than $v''(\hat{\theta})$.

## 4. RESULTS FOR OTHER VARIANCE ESTIMATION METHODS

It is shown here that results analogous to those established in Wolter for the random group estimator also hold for the jackknife and balance half-sample estimators of variance.

The jackknife estimators will be considered first. It will be proven that with the same assumptions considered in Wolter for the random group estimator, including conditions (a)-(c), the bias expression (3.4) still holds if $v_{RG}(\hat{\theta})$ is replaced by either of the two jackknife estimators presented in Section 4.2.1 of Wolter. This will be done by simply demonstrating that with these assumptions the jackknife estimators are identical to the random group estimator.

First, generally following the notation of Wolter, we proceed to define these jackknife estimators. Let $\hat{\theta}$ be as in (2.2). Corresponding to each group $\alpha$, let $\hat{\theta}_{(\alpha)}$ be the estimator of the same functional form as $\hat{\theta}$ but with the $\alpha$-th group omitted and let

$$\hat{\theta}'_\alpha = k\hat{\theta} - (k-1)\hat{\theta}_{(\alpha)},$$

$$\hat{\hat{\theta}} = \sum_{\alpha=1}^{k} \hat{\theta}'_\alpha / k \ .$$

(Note that the notation $\hat{\theta}'_\alpha$ is used here in place of $\hat{\theta}_\alpha$ to distinguish this estimator from (3.2).) The two jackknife estimators, $v_1(\hat{\hat{\theta}})$, $v_2(\hat{\hat{\theta}})$ are then

$$v_1(\hat{\hat{\theta}}) = \frac{1}{k(k-1)} \sum_{\alpha=1}^{k} (\hat{\theta}'_\alpha - \hat{\hat{\theta}})^2,$$

$$v_2(\hat{\hat{\theta}}) = \frac{1}{k(k-1)} \sum_{\alpha=1}^{k} (\hat{\theta}'_\alpha - \hat{\theta})^2.$$

Now,

$$\hat{\theta}_{(\alpha)} = \frac{k}{k-1} \sum_{i=1}^{N} w_i t'_{i(\alpha)} Y_i,$$

where

$$t'_{i(\alpha)} \begin{array}{l} = 1 \text{ if i-th unit is in sample and not in} \\ \quad \text{the } \alpha\text{-th group} \\ = 0 \text{ otherwise} \end{array}$$

Then,

$$\hat{\theta}'_\alpha = k\hat{\theta} - (k-1)\hat{\theta}_{(\alpha)} = \sum_{i=1}^{N} kw_i t_{i(\alpha)} Y_i = \hat{\theta}_\alpha, \qquad (4.1)$$

where $\hat{\theta}_\alpha$ and $t_{i(\alpha)}$ are as in (3.2) and (3.3) respectively. Furthermore,

$$\hat{\hat{\theta}} = \sum_{i=1}^{N} w_i (\sum_{\alpha=1}^{k} t_{i(\alpha)}) Y_i = \sum_{i=1}^{N} w_i t_i Y_i = \hat{\theta}. \qquad (4.2)$$

It then immediately follows from (4.1) and (4.2) that $v_{RG}(\hat{\theta})$, $v_1(\hat{\hat{\theta}})$ and $v_2(\hat{\hat{\theta}})$ are identical.

Turning now to balanced half-sample estimators, we first fix a situation for which a balanced half-sample variance estimator would be appropriate in the context of the work in this paper. Assume that a population is divided into L strata and from each stratum two clusters are selected $\pi$ps. Furthermore, the sampling within the selected clusters is independent from cluster to cluster. It is also assumed that a characteristic $\theta$ can be estimated by an estimator $\hat{\theta}$ of the form

$$\hat{\theta} = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{j=1}^{M_i} w_{hij} t_{hij} Y_{hij},$$

where $Y_{hij}$ denotes the response for the j-th elementary unit in the i-th cluster of the h-th stratum and $w_{hij}$, $t_{hij}$ have similar interpretations. Finally, it is assumed that (2.1) holds with i replaced by hij, and that the analogous forms of (2.10) - (2.12) are satisfied with U the set of ordered pairs of elementary units in different clusters; as would hold, for example, if interviewer assignments are nested within clusters.

We let now $hi_\ell$, $\ell = 1,2$ denote the selected clusters in the h-th stratum and

$$\hat{\theta}_{h\ell} = 2\sum_{j=1}^{M_i} w_{hi_\ell j} t_{hi_\ell j} Y_{hi_\ell j};$$

take $\hat{\theta}_\alpha$, $\alpha=1,\ldots,k$ to be a balanced set of half-sample estimators of the form

$$\hat{\theta}_\alpha = \sum_{h=1}^{L} \sum_{\ell=1}^{2} \delta_{h\ell\alpha}\, \hat{\theta}_{h\ell} ,$$

where

$$\begin{aligned} \delta_{h\ell\alpha} &= 1 \text{ if cluster } hi_\ell \text{ is in the } \alpha\text{-th half-sample} \\ &= 0 \text{ otherwise,} \end{aligned}$$

and finally consider the balanced half-sample variance estimator

$$v_k(\hat{\theta}) = \sum_{\alpha=1}^{k} (\hat{\theta}_\alpha - \hat{\theta})^2/k . \tag{4.3}$$

Then, as explained in Section 3 of Wolter, since $\hat{\theta}$ is linear and the $\hat{\theta}_\alpha$ are a balanced set, upon expansion of (4.3) the cross product terms cancel, and hence

$$v_k(\hat{\theta}) = \sum_{h=1}^{L} \frac{1}{4} (\hat{\theta}_{h1} - \hat{\theta}_{h2})^2.$$

Now $4^{-1}(\hat{\theta}_{h1} - \hat{\theta}_{h2})^2$ can be viewed as a random group variance estimator, with 2 clusters and 2 random groups, of the variance of

$$\sum_{i=1}^{N_h} \sum_{j=1}^{M_i} w_{hij} t_{hij} Y_{hij} .$$

It follows from example D.7 of Wolter that

$$\text{Bias } [\tfrac{1}{4} (\hat{\theta}_{h1} - \hat{\theta}_{h2})^2] = 2[\text{Var}(\tilde{\theta}_{hwr}) - \text{Var}(\tilde{\theta}_{h\pi ps})],$$

where

$$\tilde{\theta}_h = \sum_{i=1}^{N_h} t_{hi} \sum_{j=1}^{M_j} w_{hij} \mu_{hij} \pi_{hij|hi},$$

$t_{hi}$ = 1 if the i-th cluster in the h-th stratum is in sample

= 0 otherwise,

$\pi_{hij|hi}$ is the conditional probability that unit hij is drawn into sample given that cluster hi is in sample, and $\text{Var}(\tilde{\theta}_{hwr})$ and $\text{Var}(\tilde{\theta}_{h\pi ps})$ denote the variances of $\tilde{\theta}_h$ assuming with and without replacement sampling, respectively. Consequently,

$$\text{Bias } [v_k(\hat{\theta})] = 2 \sum_{h=1}^{L} [\text{Var}(\tilde{\theta}_{hwr}) - \text{Var}(\tilde{\theta}_{h\pi ps})] .$$

Thus the bias of $v_k(\hat{\theta})$, with the conditions considered, is independent of the response error and is in the between cluster component of the sampling variance. Since $\hat{\theta}$ is linear, the bias would also be the same for the alternative balanced half-sample variance estimators defined in Section 3.4 of Wolter.

## 5. NONLINEAR ESTIMATORS

In this section, variance estimation for nonlinear estimators of one or more random variables in the presence of response errors is considered, where each random variable satisfies a model of the form (2.1). We will concentrate on the problem encountered when using a Taylor series approximation in the variance estimation, although similar difficulties would also occur with other variance estimation methods.

We first examine the situation without the additional assumptions (2.11)-(2.13). Since even for linear estimators the variance estimators employed without these addition assumptions yield bias expressions such as (2.4) that can be quite large, it should be obvious that the same bias problem would occur with a nonlinear estimator $\hat{\theta}$ if, for example, the following approach is used: A linear approximation, denoted $f(\hat{\theta})$, to $\hat{\theta}$ is obtained, and $Var(\hat{\theta})$ is then estimated by $v_c[f(\hat{\theta})]$, where $v_c$ is as in Section 2. A specific example to illustrate this fact will now be presented anyway, since with a slight modification this example will later also serve to illustrate the difficulties that can arise even with the additional assumptions (2.11)-(2.13).

Assume srs wor and

$$\hat{\theta} = \bar{y}/\bar{x},$$

where for the i-th unit

$$y_i = \mu_i + \varepsilon_i, \tag{5.1}$$

$$x_i = \mu_i' + \varepsilon_i', \tag{5.2}$$

with the quantities in (5.1) and (5.2) satisfying conditions analogous to (2.1). Then to estimate $Var(\hat{\theta})$ by a linearization technique, the textbook variance estimator for srs wor would typically be used, where corresponding to the i-th sample unit the value

$$-\frac{E(\bar{y})}{E(\bar{x})^2}\, x_i + \frac{1}{E(\bar{x})}\, y_i \tag{5.3}$$

would ideally be used, although in practice this value is estimated by

$$- \frac{\bar{y}}{\bar{x}^2} x_i + \frac{1}{\bar{x}} y_i . \tag{5.4}$$

Now to simplify matters in this example it will be further assumed that $\mu_i = \mu_i' = 3$ for all units and that the survey from which the sample values are obtained is conducted by one interviewer selected at random from a pool of interviewers for which $e_i \equiv 1$, $e_i' \equiv -1$ for 1/2 the interviewers, while for the other half, $e_i \equiv -1$, $e_i' \equiv 1$. Then $\mathrm{Var}(\hat{\theta}) = 9/16$. However, irrespective of whether (5.3) or (5.4) is used in the variance estimation, the value would be the same for each sample unit, and hence the variance estimate would always be 0. Thus the bias of the variance estimator would be $- 9/16$, independently of sample size, and therefore of order 1.

Although this example is rather artificial, it can be modified in a number of ways to make it more realistic, while still retaining the order 1 bias. For example, if it is merely assumed that the $\mu_i$ and the $\mu_i'$ both average 3, instead of being 3 for each i, then the expected value of the variance estimator would still tend to 0 and $\mathrm{Var}(\hat{\theta})$ tend to 9/16 with increasing n.

Again, the above example should not be surprising, given that the same bias difficulties arise for linear estimators. What is much more interesting is that an order 1 bias in the variance estimator as a function of n may still remain if this example is modified so that assumptions (2.11)-(2.13) are satisfied and the appropriate random group variance estimator is used after linearization. To illustrate, maintain all the assumptions of the previous example, with the exception that the sample is now divided into two random groups and there are two interviewers chosen at random wr, each of whom is assigned one of the random groups to interview. (2.11)-(2.13) would then hold with U the set of sample pairs in different random groups. If either (5.3) or (5.4) is used to obtain a linear approximation $f(\hat{\theta})$, then $E(v_{RG}[f(\hat{\theta})]) = 2/9$ independently of n, while $Var(\hat{\theta}) = 19/64$, and thus the bias of the variance estimator is again of order 1.

To see what additional type of conditions must be satisfied in order for the bias of the variance estimator to be less important, consider the previous example but with k random groups and k interviewers in place of 2 random groups and 2 interviewers. Then $Var(\hat{\theta})$ can be shown to be of order 1/k while the bias of $v_{RG}[f(\hat{\theta})]$ as an estimator of $Var(\hat{\theta})$ is of order $1/k^2$, and thus if k is sufficiently large this bias will not be important. This illustrates that in order to develop rigorous conditions for which $E(v_{RG}[f(\hat{\theta})])$ will converge to $Var(\hat{\theta})$ for situations similiar to this example but with sampling variance allowed for, both n and k must be allowed to approach $\infty$. Thus, variance estimators for nonlinear estimators with biases that are unimportant do appear to exist but only under carefully drawn conditions.

# REFERENCE

Wolter, Kirk M. (1985), Introduction to Variance Estimation, New York: Springer-Verlag.