The Annals of Statistics
1981, Vol. 9, No. 5, 1010–1019

INFERENCE FROM STRATIFIED SAMPLES: PROPERTIES OF THE LINEARIZATION, JACKKNIFE AND BALANCED REPEATED REPLICATION METHODS

BY D. KREWSKI[1] AND J. N. K. RAO[2]

Health and Welfare Canada and Carleton University

The asymptotic normality of both linear and nonlinear statistics and the consistency of the variance estimators obtained using the linearization, jackknife and balanced repeated replication (BRR) methods in stratified samples are established. The results are obtained as $L \to \infty$ within the context of a sequence of finite populations $\{\Pi_L\}$ with L strata in Π_L and are valid for any stratified multistage design in which the primary sampling units (psu's) are selected *with* replacement and in which independent subsamples are taken within those psu's selected more than once. In addition, some exact analytical results on the bias and stability of these alternative variance estimators in the case of ratio estimation are obtained for small L under a general linear regression model.

1. Introduction. Many large scale surveys now involve large numbers of strata with relatively few primary sampling units (psu's) selected within each stratum. One segment of the Current Population Survey (CPS) conducted by the U.S. Bureau of the Census, for example, involved 110 strata with only three psu's selected within each stratum (Gurney and Jewett, 1975). In recent years, problems of statistical inference based on data from such stratified cluster samples have received considerable attention. In particular, three general methods of estimating the variance of nonlinear statistics such as regression and correlation coefficients have been advanced: Taylor expansion or linearization, the jackknife and balanced repeated replication (BRR).

Previous work on the properties of these methods in stratified samples has been largely empirical. Using data from the March 1967 CPS and sample designs involving the selection of two psu's from each of $L = 6,12$ or 30 strata, Kish and Frankel (1974) studied the degree to which the statistic $T = (\hat{\theta} - \theta)/v^{1/2}(\hat{\theta})$ followed a t distribution with L degrees of freedom. Here, $\hat{\theta}$ is an estimator of some parameter θ and $v(\hat{\theta})$ is an estimator of the variance of $\hat{\theta}$ based on one of the three methods mentioned earlier. They found the t approximation to be adequate for purposes of constructing two-sided confidence intervals for a variety of population parameters (excepting possibly multiple correlation coefficients) with as few as 6 or 12 strata. The BRR method performed consistently better than the jackknife which in turn performed better than the linearization method, although the differences were small for relatively simple statistics such as ratios. When the stability of these variance estimators was examined, however, their performance was found to be in the reverse order. Further empirical investigations of the properties of these three methods have been made by Bean (1975), Campbell and Meyer (1978), Lemeshow and Levy (1978) and Shah, Holt and Folsom (1977).

Received January, 1978; revised December, 1980.

[1] Research done at Carleton University during an educational leave of absence from Health and Welfare Canada. Additional support was provided by a scholarship from the Natural Sciences and Engineering Research Council of Canada.

[2] Research supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

AMS 1970 *subject classification.* Primary 62D05.

Key words and phrases. Stratified sampling, variance estimation, linearization method, jackknife method, balanced repeated replication, central limit theorem, ratio estimation, bias, stability.

In this paper, a theoretical basis for the construction of approximate confidence intervals for θ using the statistic T is provided for designs involving large numbers of strata. The three methods of variance estimation are outlined in Section 2 and the asymptotic normality of T is established for each method in Section 3. These asymptotic results, obtained as $L \to \infty$ within the context of a sequence of finite populations $\{\Pi_L\}$ with L strata in Π_L, are valid for any stratified multistage design in which the psu's are selected *with* replacement and in which independent subsamples are taken within those psu's selected more than once. Finally, some exact analytical results on the bias and stability of these alternative variance estimators are given in Section 4 for the special case of ratio estimation in stratified simple random sampling. These results are obtained under a fairly general linear regression model, facilitating an examination of the impact of changes in the values of the model parameters on the properties of the variance estimators.

2. Variance estimation. Many parameters such as ratios, correlation and regression coefficients may be expressed as a nonlinear function of the vector of population totals, say $\theta = g(\mathbf{Y})$, such that $g(\mathbf{Y}) \propto g(\bar{\mathbf{Y}})$. Here, $\bar{\mathbf{Y}} = \mathbf{Y}/M$ where M denotes the number of units in the population which may be unknown. A natural estimator of θ is given by $\hat{\theta} = g(\hat{\mathbf{Y}})$ where $\hat{\mathbf{Y}} = (\hat{Y}_1, \cdots, \hat{Y}_p)'$ is an unbiased linear estimator of $\mathbf{Y} = (Y_1, \cdots, Y_p)'$. Since $T = (\hat{\theta} - \theta)/v^{1/2}(\hat{\theta}) \equiv (g(\bar{\mathbf{y}}) - g(\bar{\mathbf{Y}})/v^{1/2}(g(\bar{\mathbf{y}}))$, where $\bar{\mathbf{Y}} = (\bar{Y}_1, \cdots, \bar{Y}_p)'$ and $\bar{\mathbf{y}} = (\bar{y}_1, \cdots, \bar{y}_p)'$ with $\bar{y}_k = \hat{Y}_k/M$ $(k = 1, \cdots, p)$, we may write $\theta = g(\bar{\mathbf{Y}})$ and $\hat{\theta} = g(\bar{\mathbf{y}})$ when investigating the asymptotic distribution of the statistic T even though M may be unknown.

EXAMPLE 2.1. The ratio $\theta = Y_1/Y_2 = \bar{Y}_1/\bar{Y}_2$ is often estimated by $\hat{\theta} = \hat{Y}_1/\hat{Y}_2 = \bar{y}_1/\bar{y}_2$.

2.1. The linearization method. This well known method is valid for any sample design provided that an unbiased estimator $v(\bar{y}_k)$ of $V(\bar{y}_k)$ is available. (Such an estimator can usually be obtained from standard sampling theory for linear statistics.) Noting that $\hat{\theta} - \theta \approx \sum(\bar{y}_k - \bar{Y}_k)g_k(\bar{\mathbf{Y}})$, where $g_k(\mathbf{t}) = \partial g(\mathbf{t})/\partial t_k$ with $\mathbf{t} = (t_1, \cdots, t_p)'$, the "linearization" variance estimator is given by

$$v_L(\hat{\theta}) = \sum g_k^2(\bar{\mathbf{y}})v(\bar{y}_k) + 2\sum\sum_{k<l} g_k(\bar{\mathbf{y}})g_l(\bar{\mathbf{y}})\mathrm{cov}(\bar{y}_k, \bar{y}_l). \tag{2.1}$$

Here, $2\,\mathrm{cov}(\bar{y}_k, \bar{y}_l)\ ) = v(\bar{y}_{kl}) - v(\bar{y}_k) - v(\bar{y}_l)$ where $\bar{y}_{kl}$ corresponds to the variable $y_{kl} = y_k + y_l$. The calculation of p variance and $p(p-1)/2$ covariance terms may be avoided by writing $v_L(\hat{\theta}) = v(\bar{z})$, where $\bar{z}$ corresponds to the derived variable $z = \sum g_k(\bar{\mathbf{y}})y_k$ (Woodruff, 1971).

One drawback of the linearization method is that the evaluation of the partial derivatives $g_k(\cdot)$ may be difficult for certain parameters (e.g., multiple and partial correlation coefficients). However, useful approximations to the required partial derivatives may be obtained using the numerical methods of Woodruff and Causey (1976). In the case of a multiple regression equation with $m + 1$ coefficients, Tepping (1968) has developed a systematic method of evaluating $g_k(\bar{\mathbf{y}})$ which involves solving $(m + 1)(m + 4)/2$ sets of linear equations with $(m + 1)$ equations in each set.

EXAMPLE 2.1 (continued). For $\hat{\theta} = \bar{y}_1/\bar{y}_2$,

$$v_L(\hat{\theta}) = \{v(\bar{y}_1) + \hat{\theta}^2 v(\bar{y}_2) - 2\hat{\theta}\,\mathrm{cov}(\bar{y}_1, \bar{y}_2)\}/\bar{y}_2^2.$$

2.2. The jackknife method. Estimates of variance obtained using the jackknife technique are based on the variability among a number of replicate estimates of θ computed from overlapping subsamples of the total sample. Unlike the linearization method, the partial derivatives of $g(\cdot)$ are thus not required.

Properties of the jackknife in the case of simple random sampling (srs) with replacement have been extensively investigated (see Miller (1974) for an excellent review). Some large sample results for srs without replacement are also available (Krewski, 1978a; Majumdar and Sen, 1978). Letting $\hat{\theta}^i$ denote the estimator of θ computed from the sample after omitting the ith observation, the jackknife estimator of the variance of $\hat{\theta}$ when sampling with replacement is given by

(2.2) $$v_J(\hat{\theta}) = n^{-1}(n-1)^{-1} \sum_{i=1}^{n} (\tilde{\theta}^i - \tilde{\theta}_J)^2 = n^{-1}(n-1) \sum_{i=1}^{n} (\hat{\theta}^i - \bar{\hat{\theta}})^2,$$

where $\tilde{\theta}^i = n\hat{\theta} - (n-1)\hat{\theta}^i$ are pseudovalues, $\tilde{\theta}_J = \sum \tilde{\theta}^i/n$ is the jackknife estimator of θ and $\bar{\hat{\theta}} = \sum \hat{\theta}^i/n$.

The extension to stratified samples is not immediate, however, and several different versions have been proposed. Both the jackknife and BRR methods developed to date, moreover, require that the psu's be selected *with* replacement, except in the special case of stratified srs (Jones, 1974; McCarthy, 1966). We will thus confine ourselves to stratified multistage designs in which the psu's are selected with replacement and in which independent subsamples are taken within those psu's selected more than once.

Suppose that $n_h \geq 2$ psu's are selected from the N_h psu's in the hth stratum with probabilities $p_{hi} > 0 (i = 1, \cdots, N_h; h = 1, \cdots, L)$, where $\sum_{i=1}^{N_h} p_{hi} = 1$. Then an unbiased estimator of the stratum total $\mathbf{Y}_h$ is given by $\hat{\mathbf{Y}}_h = \sum_{i=1}^{n_h} \hat{\mathbf{Y}}_{hi}/(n_h p_{hi})$, where $\hat{\mathbf{Y}}_{hi}$ is an unbiased estimator of the total $\mathbf{Y}_{hi}$ for a selected psu based on sampling at the second and subsequent stages. Hence, $\mathbf{Y}$ is unbiasedly estimated by $\hat{\mathbf{Y}} = \sum \hat{\mathbf{Y}}_h$ and $\bar{\mathbf{Y}}$ is unbiasedly estimated by $\bar{\mathbf{y}} = \hat{\mathbf{Y}}/M$, where $M = \sum M_h$ and M_h denotes the total number of units in stratum h. Letting $W_h = M_h/M$ denote the weight of the hth stratum, we can write $\bar{\mathbf{y}} = \sum W_h \bar{\mathbf{y}}_h$, where $\bar{\mathbf{y}}_h = \sum_{i=1}^{n_h} \mathbf{y}_{hi}/n_h$ and $\mathbf{y}_{hi} = \hat{\mathbf{Y}}_{hi}/(M_h p_{hi})$. Note that for each h, the $\mathbf{y}_{hi}$'s are independent identically distributed random variables whereas for $h \neq h'$, $\mathbf{y}_{hi}$ and $\mathbf{y}_{h'j}$ are independent but not necessarily identically distributed.

Let $\bar{\mathbf{y}}^{hi}$ denote the estimator of $\bar{\mathbf{Y}}$ computed from the sample after omitting $\mathbf{y}_{hi}$ $(i = 1, \cdots, n_h; h = 1, \cdots, L)$, i.e.,

$$\bar{\mathbf{y}}^{hi} = \sum_{h' \neq h} W_{h'}\bar{\mathbf{y}}_{h'} + W_h(n_h\bar{\mathbf{y}}_h - \mathbf{y}_{hi})/(n_h - 1).$$

Then Jones' (1974) jackknife estimator of $V(\hat{\theta})$ is given by

(2.3) $$v_J^{(1)}(\hat{\theta}) = \sum_{h=1}^{L} n_h^{-1}(n_h - 1) \sum_{i=1}^{n_h}(\hat{\theta}^{hi} - \hat{\theta}^h)^2,$$

where $\hat{\theta}^{hi} = g(\bar{\mathbf{y}}^{hi})$ and $\hat{\theta}^h = \sum_{i=1}^{n_h} \hat{\theta}^{hi}/n_h$ (see also Brillinger, 1977). Replacing $\hat{\theta}^h$ by $\hat{\theta}$ in (2.3) leads to the modified jackknife variance estimator

(2.4) $$v_J^{(2)}(\hat{\theta}) = \sum_{h=1}^{L} n_h^{-1}(n_h - 1) \sum_{i=1}^{n_h}(\hat{\theta}^{hi} - \hat{\theta})^2.$$

Kish and Frankel (1974) and Lee (1973) have considered $v_J^{(2)}(\hat{\theta})$ in the special case of all $n_h = 2$, although Kish and Frankel actually compute $\bar{\mathbf{y}}^{hi}$ by deleting $\mathbf{y}_{hi}$ from the sample and including $\mathbf{y}_{hj}$ $(j \neq i = 1, 2)$ twice. Two other variations of $v_J^{(1)}(\hat{\theta})$, denoted by $v_J^{(3)}(\hat{\theta})$ and $v_J^{(4)}(\hat{\theta})$, may be obtained by replacing $\hat{\theta}^h$ in (2.3) by $\sum\sum \hat{\theta}^{hi}/n$ and $\sum \hat{\theta}^h/L$ respectively, where $n = \sum n_h$.

With pseudovalues defined by $\tilde{\theta}^{hi} = n_h\hat{\theta} - (n_h - 1)\hat{\theta}^{hi}$, both

(2.5) $$\tilde{\theta}_J^{(1)} = \sum_{h=1}^{L} \sum_{i=1}^{n_h} \tilde{\theta}^{hi}/n \qquad \text{and} \qquad \tilde{\theta}_J^{(2)} = \sum_{h=1}^{L} n_h^{-1} \sum_{i=1}^{n_h} \tilde{\theta}^{hi}/L$$

represent natural extensions of the jackknife estimator of θ proposed by McCarthy (1966) in the special case $n_h = 2$ for all h. The corresponding jackknife variance estimators $v_J^{(5)}(\hat{\theta})$ and $v_J^{(6)}(\hat{\theta})$ are given by

(2.6) $$v_J^{(5)}(\hat{\theta}) = \sum_{h=1}^{L} n_h^{-1}(n_h - 1)^{-1} \sum_{i=1}^{n_h} (\tilde{\theta}^{hi} - \tilde{\theta}_J^{(1)})^2$$

with $v_J^{(6)}(\hat{\theta})$ obtained by replacing $\tilde{\theta}_J^{(1)}$ by $\tilde{\theta}_J^{(2)}$ in (2.6). The jackknife statistic $\tilde{\theta}_J^{(2)}$ has also been considered by Folsom, Bayless and Shah (1971). Their jackknife variance estimator is obtained by replacing $\tilde{\theta}_J^{(1)}$ by $\sum_{i=1}^{n_h} \tilde{\theta}^{hi}/n_h$ in (2.6) and is identical to $v_J^{(1)}(\hat{\theta})$.

Jackknife variance estimators requiring less computational effort may be constructed using only a random sample of size $m_h (< n_h)$ of the n_h psu's in the sample in stratum h. For example, $v_J^{(2)}(\hat{\theta})$ would be computed as

(2.7) $$\tilde{v}_J^{(2)}(\hat{\theta}) = \sum_{h=1}^{L} m_h^{-1}(n_h - 1) \sum_{i=1}^{m_h}(\hat{\theta}^{hi} - \hat{\theta})^2.$$

When $n_h > 2$ for some h, however, such estimators will be less efficient in the linear case $g(\bar{\mathbf{Y}}) = \sum_{k=1}^{p} c_k \bar{Y}_k$ than the usual estimator

$$v(\mathbf{c}'\bar{\mathbf{y}}) = \mathbf{c}'\hat{\mathbf{D}}(\bar{\mathbf{y}})\mathbf{c} \tag{2.8}$$

of $V(\mathbf{c}'\bar{\mathbf{y}}) = \mathbf{c}'\mathbf{D}(\bar{\mathbf{y}})\mathbf{c}$, where $\mathbf{c} = (c_1, \cdots, c_p)'$ (Krewski, 1978b). Here,

$$\hat{\mathbf{D}}(\bar{\mathbf{y}}) = \sum W_h^2 \hat{\mathbf{\Gamma}}_h / n_h \tag{2.9}$$

is an unbiased estimator of the covariance matrix $\mathbf{D}(\bar{\mathbf{y}}) = \sum W_h^2 \mathbf{\Gamma}_h / n_h$, where

$$\hat{\mathbf{\Gamma}}_h = \sum_{i=1}^{n_h} (\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)'/(n_h - 1) \tag{2.10}$$

is an unbiased estimator of $\mathbf{\Gamma}_h = E(\mathbf{y}_{hi} - \bar{\mathbf{Y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{Y}}_h)'$, with $\bar{\mathbf{Y}}_h = \mathbf{Y}_h / M_h$. All of the jackknife variance estimators discussed previously reduce to (2.8) in the linear case, as does (2.7) when all $n_h = 2$.

As in the case of simple random sampling, the jackknife can also be used to reduce the bias of $\hat{\theta}$. The jackknife estimators $\tilde{\theta}_J^{(1)}$, $\tilde{\theta}_J^{(2)}$ and

$$\tilde{\theta}_J^{(3)} = (1 + n - L)\hat{\theta} - \sum_{h=1}^{L} (n_h - 1)\hat{\theta}^h \tag{2.11}$$

(Jones, 1974) may be considered in this regard, although empirical evidence (Kish, Namboodiri and Pillai, 1962; Frankel, 1971; Bean, 1975) suggests that the bias of $\hat{\theta}$ is largely negligible in large scale surveys. (One exception to this consensus is the multiple correlation coefficient for which Frankel found the relative bias to be as high as 20–25%.) We will thus focus our attention on the jackknife variance estimators and associated T statistics and refer the reader to Krewski and Rao (1978, 1981) for results on the properties of the jackknife estimators of θ.

2.3. Balanced repeated replication. When $n_h = 2$ for all h, McCarthy (1966, 1969) has proposed a method of variance estimation based on a number of half-samples formed by deleting one psu from the sample in each stratum. The set of S half-samples used may be defined by an $S \times L$ matrix $\mathbf{\Delta} = ((\delta_{jh}))$ where $\delta_{jh} = \pm 1$ depending on whether the first or second sample psu in the hth stratum is in the jth half-sample, and $\sum_{j=1}^{S} \delta_{jh} = \sum_{j=1}^{S} \delta_{jh}\delta_{jh'} = 0$ $(h \neq h')$. A minimal set of $L + 1 \leq S \leq L + 4$ balanced half-samples may be constructed using the methods of Plackett and Burman (1946).

Let $\bar{\mathbf{y}}^{(j)}$ denote the estimator of $\bar{\mathbf{Y}}$ based on the jth half-sample and let $\hat{\theta}^{(j)} = g(\bar{\mathbf{y}}^{(j)})$. A BRR variance estimator is then given by

$$v_B^{(1)}(\hat{\theta}) = \sum (\hat{\theta}^{(j)} - \hat{\theta})^2 / S. \tag{2.12}$$

Letting $\hat{\theta}_c^{(j)} = g(\bar{\mathbf{y}}_c^{(j)})$ denote the estimator of θ based on the complement of the jth half-sample, two additional BRR variance estimators are given by

$$v_B^{(2)}(\hat{\theta}) = \sum (\hat{\theta}^{(j)} - \hat{\theta}_c^{(j)})^2 / (4S) \tag{2.13}$$

and

$$v_B^{(3)}(\hat{\theta}) = \sum \{(\hat{\theta}^{(j)} - \hat{\theta})^2 + (\hat{\theta}_c^{(j)} - \hat{\theta})^2\} / (2S). \tag{2.14}$$

Due to the orthogonality constraints on $\mathbf{\Delta}$, all of the BRR variance estimators reduce to the usual variance estimator in (2.8) in the linear case.

More generally, when $n_h = q$ (a prime) for all h, Gurney and Jewett (1975) discuss the use of orthogonal arrays in constructing balanced subsamples comprised of one psu in each stratum. Since no such method is at present available for arbitrary n_h, however, BRR is less widely applicable than the jackknife.

3. Large sample results. A framework for the development of asymptotic theory is provided by the concept of a sequence of finite populations $\{\Pi_L\}_{L=1}^{\infty}$ with L strata in Π_L. For simplicity of notation, the population index L will be suppressed in what follows and all limiting processes will be understood to be as $L \to \infty$. Writing $\mathbf{y}_{hi} = (y_{hi1}, \cdots, y_{hip})'$ and

$\bar{\mathbf{Y}}_h = (\bar{Y}_{h1}, \cdots, \bar{Y}_{hp})'$, we shall first study the asymptotic properties of $\bar{\mathbf{y}}$ and $\hat{\mathbf{D}}(\bar{\mathbf{y}})$ under the following regularity conditions.

C1. $\sum_{h=1}^{L} W_h E\,|y_{hik} - \bar{Y}_{hk}|^{2+\delta} = O(1)$ for some $\delta > 0$ $(k = 1, \cdots, p)$.
C2. $\max_{1 \le h \le L} n_h = O(1)$.
C3. $\max_{1 \le h \le L} W_h = O(L^{-1})$.
C4. $n \sum W_h^2 \mathbf{\Gamma}_h / n_h \to \mathbf{\Gamma}$ (positive definite).

Condition C1 is a standard Liapounov-type condition on the $2 + \delta$ absolute moments. Condition C2 (*bounded allocation*) reflects our intention to focus on surveys with large numbers of strata with relatively few psu's selected within each stratum. Thus, we will require that no strata be of disproportionate size (C3). Except in the case of $v_{\mathscr{J}}^{(5)}(\hat{\theta})$ and $v_{\mathscr{J}}^{(6)}(\hat{\theta})$ in Theorem 3.4 below, we note that it is possible to replace C2 and C3 in what follows by the single and somewhat weaker condition $\max_{1 \le h \le L} W_h / w_h = O(1)$, where $w_h = n_h / n$. Finally, it is assumed that the limit of the dispersion matrix of $\bar{\mathbf{y}}$ exists when multiplied by the normalizing factor n (C4).

Since $\{\mathbf{y}_{hi}\}$ is a double sequence of rowwise independent random variables (recall that the row subscript L is omitted), we can make use of an established central limit theorem (Hoadley, 1971) and law of large numbers (Sen, 1970) for independent nonidentically distributed random variables $\{\mathbf{X}_t\}_{t=1}^{T}$ with $E(\mathbf{X}_t) = \boldsymbol{\mu}_t$ and $D(\mathbf{X}_t) = \mathbf{\Gamma}_t^*$, where $\mathbf{X}_t = (X_{t1}, \cdots, X_{tp})'$ and $\boldsymbol{\mu}_t = (\mu_{t1}, \cdots, \mu_{tp})'$.

LEMMA 3.1 (Central limit theorem). *If* $\sum_{t=1}^{T} \mathbf{\Gamma}_t^* / T \to \mathbf{\Gamma}^*$ (pos. def.) *and* $T^{-1} \sum_{t=1}^{T} E\,|X_{tk} - \mu_{tk}|^{2+\delta} = O(1)$ *as* $T \to \infty$ *for some* $\delta > 0$ *and* $k = 1, \cdots, p$, *then* $T^{1/2}(\bar{\mathbf{X}} - \bar{\boldsymbol{\mu}}) \to_d N(\mathbf{0}, \mathbf{\Gamma}^*)$, *where* $\bar{\mathbf{X}} = \sum \mathbf{X}_t / T$ *and* $\bar{\boldsymbol{\mu}} = \sum \boldsymbol{\mu}_t / T$.

LEMMA 3.2 (Law of large numbers). *If* $T^{-1} \sum_{t=1}^{T} E\,|X_{tk}|^{1+\delta} = O(1)$ *as* $T \to \infty$ *for some* $\delta > 0$ *and* $k = 1, \cdots, p$, *then for any* $\epsilon > 0$, $P\{|\bar{X}_k - \bar{\mu}_k| \ge \epsilon\} = O(T^{-r})$, *where* $r = \delta$ *when* $\delta \le 1$ *and* $r = (1 + \delta)/2$ *when* $\delta > 1$. *Here*, $\bar{\mathbf{X}} = (\bar{X}_1, \cdots, \bar{X}_p)'$ *and* $\bar{\boldsymbol{\mu}} = (\bar{\mu}_1, \cdots, \bar{\mu}_p)'$.

Applying Lemma 3.1 to the random variables $\mathbf{X}_{hi} = W_h(\mathbf{y}_{hi} - \bar{\mathbf{Y}}_h)$ we get, noting that $T = n$,

THEOREM 3.1. *Under conditions* C1–C4, $n^{1/2}(\bar{\mathbf{y}} - \bar{\mathbf{Y}}) \to_d N(\mathbf{0}, \mathbf{\Gamma})$.

Lemma 3.2 can be used to establish

THEOREM 3.2. *Under conditions* C1–C3, $n\{\hat{\mathbf{D}}(\bar{\mathbf{y}}) - \mathbf{D}(\bar{\mathbf{y}})\} \to \mathbf{0}$ *in probability*.

PROOF. We can write the kth variance term in $n\hat{\mathbf{D}}(\bar{\mathbf{y}})$ as $nv(\bar{y}_k) = n\{\sum_{h=1}^{L} W_h^2 n_h^{-1} \cdot \sum_{i=1}^{n_h} (y_{hik} - \bar{Y}_{hk})^2/(n_h - 1) - \sum_{h=1}^{L} W_h^2 (\bar{y}_{hk} - \bar{Y}_{hk})^2/(n_h - 1)\} = \sum_h \sum_i X_{hik} - \sum_h X_{hk}$, say. Applying Lemma 3.2 to $\{X_{hik}\}$ and $\{X_{hk}\}$ separately we get $n\{v(\bar{y}_k) - V(\bar{y}_k)\} \to 0$ in probability. The covariance terms may be handled in a similar fashion. □

Thus, $\bar{\mathbf{y}}$ is asymptotically normally distributed and $n\hat{\mathbf{D}}(\bar{\mathbf{y}})$ is a consistent estimator of $n\mathbf{D}(\bar{\mathbf{y}})$ as $L \to \infty$. It is also easy to show that $\bar{\mathbf{y}} - \bar{\mathbf{Y}} \to \mathbf{0}$ in probability under C1–C3. (This last result also holds with only bounded $1 + \delta$ rather than $2 + \delta$ moments in C1.)

3.1. The linearization method. We require two additional regularity conditions in order to establish the asymptotic normality of $\hat{\theta} = g(\bar{\mathbf{y}})$ and the consistency of $nv_L(\hat{\theta})$.

C5. $\bar{Y}_k \to \mu_k$ (finite) for $k = 1, \cdots, p$.
C6. The first derivatives $g_k(\cdot)$ of $g(\cdot)$ are continuous in a neighborhood of $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)'$.

Under C5, it is assumed that the limit of the sequence of population means exists. Condition C6 is a standard requirement for the asymptotic normality of a nonlinear function of several random variables. Building on Theorems 3.1 and 3.2, Theorem 3.3 below can be readily established using standard arguments (C. R. Rao, 1973, page 387).

THEOREM 3.3. *Under conditions* C1–C6, (i) $n^{1/2}(\hat{\theta} - \theta) \to_d N(0, \sigma^2)$, (ii) $nv_L(\hat{\theta}) \to \sigma^2$ *in probability and* (iii) $T_L = (\hat{\theta} - \theta)/v_L^{1/2}(\hat{\theta}) \to_d N(0, 1)$, *where* $\sigma^2 = \sum\sum g_k g_{k'}, \gamma_{kk'}, g_k = g_k(\boldsymbol{\mu})$ *and* $\boldsymbol{\Gamma} = ((\gamma_{kk'}))$.

3.2. The jackknife method. Using the method of proof introduced by Miller (1964) for srs, we shall establish

THEOREM 3.4. *Under conditions* C1–C6, (i) $nv_J^{(i)}(\hat{\theta}) \to \sigma^2$ *in probability and* (ii) $T_J^{(i)} = (\hat{\theta} - \theta)/\sqrt{v_J^{(i)}(\hat{\theta})} \to_d N(0, 1)$ *for* $i = 1, \cdots, 6$.

PROOF. Only the case $p = 1$ is considered in detail; the extension to $p > 1$ is relatively straightforward. Let $I = (\mu - \Delta, \mu + \Delta)$, $\Delta > 0$, be any neighborhood of μ in which $g'(\cdot)$ is continuous. We need the following result.

LEMMA 3.3. $P\{$*all* $\bar{y}^{hi}, \bar{y} \in I$ *simultaneously*$\} \to 1$.

PROOF OF LEMMA 3.3. Since $\bar{y} - \bar{Y} \to 0$ in probability and $\bar{Y} \to \mu$, we have $P\{|\bar{y} - \mu| < \Delta/2\} \to 1$. Thus, we need only show that

$$P\{\max_{1\le i\le n_h; 1\le h\le L} |\bar{y}^{hi} - \bar{y}| < \Delta/2\} \to 1.$$

To this end, we note that $|\bar{y}^{hi} - \bar{y}| \le (n_h - 1)^{-1} W_h\{|y_{hi} - \bar{Y}_h| + |\bar{y}_h - \bar{Y}|\}$ where, using the Chebychev inequality and noting that $(n_h - 1)^{-1} \le 2/n_h$ since $n_h \ge 2$,

$$\begin{aligned} &P\{\max_{1\le i\le n_h; 1\le h\le L} (n_h - 1)^{-1} W_h |y_{hi} - \bar{Y}_h| \ge \Delta/4\} \\ &\quad \le \textstyle\sum_{h=1}^{L} \sum_{i=1}^{n_h} P\{(n_h - 1)^{-1} W_h |y_{hi} - \bar{Y}_h| \ge \Delta/4\} \\ &\quad \le (2^{5+2\delta}/\Delta^{2+\delta})(\max_{1\le h\le L} W_h)^{1+\delta} \textstyle\sum_{h=1}^{L} W_h E|y_{hi} - \bar{Y}_h|^{2+\delta} \\ &\quad = O(L^{-(1+\delta)}) \end{aligned}$$

under C1 and C3. Similarly, $P\{\max_{1\le h\le L} (n_h - 1)^{-1} W_h |\bar{y}_h - \bar{Y}_h| \ge \Delta/4\} \to 0$, establishing the lemma.

Since $\lim P(A_t) = \lim P(A_t B_t)$ for any two sequences of events with $\lim P(B_t) = 1$, we may tacitly assume that the events described in Lemma 3.3 hold in the remainder of the proof of Theorem 3.4. While details are provided only for $v_J^{(2)}(\hat{\theta})$, similar arguments hold in the case of the remaining jackknife variance estimators including $\tilde{v}_J^{(2)}(\hat{\theta})$.

When all $\bar{y}^{hi}, \bar{y} \in I$ we may write

$$\begin{aligned} \hat{\theta}^{hi} &= g(\bar{y}^{hi}) = g(\bar{y}) + (\bar{y}^{hi} - \bar{y}) g'(\xi^{hi}) \\ &= \hat{\theta} - (n_h - 1)^{-1} W_h (y_{hi} - \bar{y}_h) g'(\xi^{hi}) \end{aligned} \tag{3.1}$$

where ξ^{hi} lies between $\bar{y}^{hi}$ and $\bar{y}$. Since $m(t) = g'(t) - g'(\mu)$ is continuous at $t = \mu$, there exists for any $\epsilon > 0$ some $\Delta_\epsilon > 0$ such that $|m(t)| < \epsilon$ *for* $t \in I_\epsilon = (\mu - \Delta_\epsilon, \mu + \Delta_\epsilon)$. Thus

$$P\{\max_{1\le i\le n_h; 1\le h\le L} |m(\xi^{hi})| < \epsilon\} \ge P\{\text{all } \bar{y}^{hi}, \bar{y} \in I_\epsilon \text{ simultaneously}\} \to 1 \tag{3.2}$$

by Lemma 3.3, i.e., $\max_{1\le i\le n_h; 1\le h\le L} |m(\xi^{hi})| \to 0$ in probability (with ξ^{hi} defined arbitrarily when the events described above do not hold).

Substituting (3.1) into (2.4) we get

$$nv_J^{(2)}(\hat{\theta}) = n \sum_{h=1}^{L} W_h^2 n_h^{-1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 \{g'(\mu) + m(\xi^{hi})\}^2/(n_h - 1)$$

$$= nv(\bar{y}) | g'(\mu)|^2 + n \sum_{h=1}^{L} W_h^2 n_h^{-1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 m^2(\xi^{hi})/(n_h - 1) \tag{3.3}$$

$$+ \text{ cross-product term.}$$

Under C4, the first term on the right-hand side of (3.3) $\to \sigma^2 = |g'(\mu)|^2\gamma$ in probability due to Theorem 3.2 (where $\Gamma = \gamma$ for $p = 1$), while both the second term and the cross-product term $\to 0$ in probability due to (3.2). Hence, result (i) of Theorem 3.4. Result (ii) follows immediately from result (i) and Theorem 3.3. □

3.3. Balanced repeated replication. In the special case $n_h = 2$ for all h, we have

THEOREM 3.5. *Under conditions* C1–C6, (i) $nv_B^{(i)}(\hat{\theta}) \to \sigma^2$ *in probability and* (ii) $T_B^{(i)} = (\hat{\theta} - \theta)/\sqrt{v_B^{(i)}(\hat{\theta})} \to_d N(0, 1)$ *for* $i = 1, 2, 3$.

PROOF. As in Theorem 3.4, only the case $p = 1$ is considered in detail. With the interval I defined as in Theorem 3.4, we need a result analogous to Lemma 3.3.

LEMMA 3.4. $P\{all\ \bar{y}^{(j)}, \bar{y} \in I\ simultaneously\} \to 1.$

PROOF OF LEMMA 3.4. We need only show that

$$P\{\max_{1\le j\le S} | \bar{y}^{(j)} - \bar{Y} | \ge \Delta/2\} \to 0. \tag{3.4}$$

Let $X_{h(j)} = LW_h(y_{h(j)} - \bar{Y}_h)$ where $y_{h(j)} = y_{h1}$ if $\delta_{jh} = +1$ and $y_{h(j)} = y_{h2}$ if $\delta_{jh} = -1$. Noting that $\max_{1\le h\le L} W_h^{1+\delta} = O(L^{-(1+\delta)})$ under C3, it follows that $L^{-1} \sum_{h=1}^{L} E|X_{h(j)}|^{2+\delta} = O(1)$. Hence

$$P\{| \bar{y}^{(j)} - \bar{Y}| \ge \Delta/2\} = O(L^{-(1+\delta/2)}) \tag{3.5}$$

by Lemma 3.2 (with $T = L$). Since $S = O(L)$, it now follows from (3.5) that

$$P\{\max_{1\le j\le S} | \bar{y}^{(j)} - \bar{Y} | \ge \Delta/2\} \le SP\{| \bar{y}^{(j)} - \bar{Y} | \ge \Delta/2\} = O(L^{-\delta/2}).$$

Hence the lemma.

The remainder of the proof of Theorem 3.5 now follows along the lines of Theorem 3.4 with the use of Lemma 3.4 and the orthogonality constraints on $\boldsymbol{\Delta}$. For example,

$$\begin{aligned} nv_B^{(1)}(\hat{\theta}) &= nS^{-1} \sum_{j=1}^{S} (\bar{y}^{(j)} - \bar{y})^2 | g'(\mu)|^2 \\ &\quad + nS^{-1} \sum_{j=1}^{S} (\bar{y}^{(j)} - \bar{y})^2 m^2(\xi^{(j)}) + \text{cross-product terms} \\ &= nv(\bar{y})| g'(\mu)|^2 + nS^{-1} \sum_{j=1}^{S} (\bar{y}^{(j)} - \bar{y})^2 m^2(\xi^{(j)}) \\ &\quad + \text{cross-product terms} \\ &\to_p | g'(\mu)|^2 \gamma = \sigma^2 \end{aligned}$$

since $\max_{1\le j\le S} | m(\xi^{(j)})| \to 0$ in probability. □

4. Exact results for ratio estimation. We now give some exact analytical results on the bias and stability of several alternative variance estimators based on the linearization, jackknife and BRR methods in the case of ratio estimation (Example 2.1) under stratified srs with proportional allocation ($n_h = W_h n$). These results are obtained under a general linear regression model given by (Rao and Ramachandran, 1974)

$$y_{hi1} = \alpha_h + \beta_h y_{hi2} + e_{hi} \qquad (i = 1, \cdots, n_h;\ h = 1, \cdots, L) \tag{4.1}$$

where, conditional on the y_{hi2}, the errors e_{hi} are uncorrelated with mean zero and variance $\delta_h y_{hi2}^{t_h}$ ($t_h \ge 0$). In practice, t_h has often been found to lie between 0 and 2. In addition, it is assumed that the variable y_{hi2} follows a gamma distribution with mean a_h and coefficient of variation (CV) $a_h^{-1/2}$. To simplify the derivations, we will assume that the number of units M_h in each stratum h is infinite. Hence,

$$\theta = \sum_{h=1}^{L} W_h E(y_{hi1}) / \sum_{h=1}^{L} W_h E(y_{hi2}) = (n/m)\bar{\alpha} + \bar{\beta} \quad \text{where} \quad \bar{\alpha} = \sum n_h \alpha_h / n,$$

$$\bar{\beta} = \sum m_h \beta_h / m \quad \text{and} \quad m = \sum m_h \quad \text{with} \quad m_h = n_h a_h.$$

4.1. Bias. After considerable algebra, the bias of $v_L(\hat{\theta})$ as an estimator of $\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$ is given by

$$\begin{aligned} B(v_L) &= -n^2(3m+2)[m(m+1)_4]^{-1}\bar{\alpha}^2 - 3[(m+1)(m+3)]^{-1}\Delta_\beta^2 \\ &\quad - \sum n_h [t_h^2 + t_h(2m+1) - m][(m + t_h + 1)_4]^{-1} f_h(t_h)\delta_h, \end{aligned} \tag{4.2}$$

provided $m > 2$, where $(m)_r = m(m-1) \cdots (m-r+1)$, $\Delta_\beta^2 = \sum m_h(\beta_h - \bar{\beta})^2/m$ and $f_h(t) = \Gamma(a_h + t)/\Gamma(a_h)$. Similarly, after some tedious algebra, the biases of the BRR variance estimators $v_B^{(1)}(\hat{\theta})$, $v_B^{(2)}(\hat{\theta})$ and $v_B^{(3)}(\hat{\theta})$ when all $n_h = 2$ may be expressed as

$$\begin{aligned} B(v_B^{(1)}) = B(v_B^{(3)}) &= 2n^2(m+2)(m-3)(3m-4)[m(m-2)(m)_5]^{-1}\bar{\alpha}^2 \\ &\quad + 4\sum [(m + 2t_h - 4)^{-1} - (m + t_h - 2)^{-1}](m + 2t_h - 2)^{-1} f_h(t_h)\delta_h \end{aligned} \tag{4.3}$$

and

$$\begin{aligned} B(v_B^{(2)}) &= n^2(m-3)(3m^2 + 4m - 16)[m(m-2)(m)_5]^{-1}\bar{\alpha}^2 - [(m+2)_2]^{-1}\Delta_\beta^2 \\ &\quad + 2\sum [(m + 2t_h - 2)^{-1}(m + 2t_h - 4)^{-1} - (m + t_h - 1)_2^{-1}] f_h(t_h)\delta_h, \end{aligned} \tag{4.4}$$

provided $m > 4$.

Examination of (4.2)–(4.4) leads to the following conclusions. (i) $B(v_L) < 0$ when $t_h \geq \frac{1}{2}$ for all h, i.e., $v_L(\hat{\theta})$ underestimates $\text{MSE}(\hat{\theta})$; (ii) $B(v_B^{(1)}) \geq 0$ when $t_h \leq 2$ for all h; i.e., $v_B^{(1)}(\hat{\theta})$ overestimates $\text{MSE}(\hat{\theta})$; (iii) $B(v_B^{(2)}) > 0$ when $t_h \leq \frac{3}{2}$ and $\beta_h = \beta$ for all h, (iv) For $\beta_h = \beta$ and $n_h = 2$ for all h, $B(v_B^{(1)}) > B(v_B^{(2)}) > 0$ when all $t_h \leq \frac{3}{2}$ and $B(v_B^{(1)}) > B(v_B^{(2)}) > |B(v_L)|$ when all $t_h = 1$; (v) when $\bar{\alpha} = 0$ and $n_h = 2$ for all h, $|B(v_L)| > |B(v_B^{(2)})| > B(v_B^{(1)}) = 0$ when all $t_h = 2$.

In order to evaluate the biases of the jackknife variance estimators $v_J^{(1)}(\hat{\theta})$ and $v_J^{(2)}(\hat{\theta})$, we need the inverse moments $E(X_1 + \lambda X_2)^{-1}$, $E(X_1 + \lambda X_2)^{-2}$, $E(X_1 + \lambda X_2)^{-1}(X_1 + X_2 + X_3)^{-1}$ and $E(2X_1 + X_3)^{-1}(2X_2 + X_3)^{-1}$, where $\lambda > 0$ ($\neq 1$) and X_1, X_2 and X_3 are independent gamma variates. (We were unable to evaluate the exact biases of the remaining jackknife variance estimators.) An explicit expression for $E(X_1 + \lambda X_2)^{-1}$ is given in Lemma 4.1 below; expressions for the remaining moments may be derived using similar arguments (Krewski and Rao, 1981).

LEMMA 4.1. *Let X_1 and X_2 be independent gamma variates with means a and b respectively. Then for any positive constant λ ($\neq 1$) and integral values of a and b,*

$$E(X_1 + \lambda X_2)^{-1} = \{\Gamma(b)\Gamma(a)\}^{-1}[I^{(1)} + \lambda^{a-1}\Gamma(b + a - 1)\{I^{(2)} + I^{(3)}\}]$$

where

$$I^{(1)} = \sum_{k=1}^{a-1} (-1)^{k+1}\lambda^{k-1}\Gamma(b + k - 1)\Gamma(a - k),$$

$$I^{(2)} = (-1)^{a+1} \sum_{k=1}^{b+a-2} (-1)^{k+1}(\lambda - 1)^{-k}(b + a - k - 1)^{-1}$$

and

$$I^{(3)} = (-1)^{b-1}(\lambda - 1)^{1-b-a}\ln \lambda.$$

PROOF. We utilize a method introduced by P. S. R. S. Rao (1974). For $t \geq 0$, let $\phi(t) = E[\exp\{-t(X_1 + \lambda X_2)\}] = (1 + t)^{-a}(1 + \lambda t)^{-b}$. Then

$$E(X_1 + \lambda X_2)^{-1} = \int_0^\infty \phi(t)\,dt = \int_0^\infty (1+t)^{-a}(1+\lambda t)^{-b}\,dt = I(a, b; \lambda).$$

Applying integration by parts for $a \geq 2$ yields

$$\begin{aligned} I(a, b; \lambda) &= (a-1)^{-1}\{1 - bI(a-1, b+1; \lambda)\} \\ &= \{\Gamma(b)\Gamma(a)\}^{-1}\{I^{(1)} + (-1)^{a+1}\lambda^{a-1}\Gamma(b+a-1)I(1, b+a-1; \lambda)\}. \end{aligned} \tag{4.5}$$

By partial fractions, for $b + a \geq 3$,

$$\begin{aligned} I(1, b+a-1; \lambda) &= (\lambda - 1)^{-1}\{(b+a-2)^{-1} - I(1, b+a-2; \lambda)\} \\ &= \textstyle\sum_{k=1}^{b+a-2} (-1)^{k+1}(\lambda-1)^{-k}(b+a-k-1)^{-1} \\ &\quad + (-1)^{b+a}(\lambda-1)^{2-b-a}I(1, 1; \lambda). \end{aligned} \tag{4.6}$$

The desired result now follows from (4.5) and (4.6) since $I(1, 1; \lambda) = (\lambda - 1)^{-1}\ln \lambda$. □

In the special case $n_h = 2$, $a_h = a$, $\beta_h = \beta$ and $t_h = t$ for all h, the biases of the different variance estimators v may be expressed in the form $B(v) = C\,\bar{\alpha}^2 + D\bar{\delta}$, where $\bar{\delta} = \sum \delta_h/L$ (Krewski and Rao, 1981). Since the expressions for $B(v_J^{(1)})$ and $B(v_J^{(2)})$ involve expressions such as that given in Lemma 4.1, we computed these C and D coefficients for $a = 1, 2, 3$, $L \leq 12$ and $t = 0, 1, 2$.

This analysis led to the following conclusions: (i) $B(v_L) < 0$ when $t = 1$ or 2 as noted previously; (ii) both jackknife variance estimators $v_J^{(1)}(\hat{\theta})$ and $v_J^{(2)}(\hat{\theta})$ also underestimate $\text{MSE}(\hat{\theta})$ when $t = 1$ or 2 and $L > 4$ with $|B(v_L)| > |B(v_J^{(1)})| > |B(v_J^{(2)})|$ in this case; (iii) $B(v_B^{(1)}) > B(v_B^{(2)}) > |B(v_L)| > |B(v_J^{(1)})| > |B(v_J^{(2)})|$ when $t = 1$ and $L > 4$; (iv) When $\bar{\alpha} = 0$, all five variance estimators overestimate MSE $(\hat{\theta})$ for $t = 0$, with $B(v_B^{(1)}) > B(v_B^{(2)}) > B(v_J^{(2)}) > B(v_J^{(2)}) > B(v_L)$ in this case. When $t = 2$, all five variance estimators are underestimates when $\bar{\alpha} = 0$ with the absolute biases following the reverse order to that for $t = 0$.

4.2. Stability. The mean square errors (MSE's) of $v_L(\hat{\theta})$ and $v_B^{(1)}(\hat{\theta})$ may be expressed in the form $\text{MSE}(v) = F\bar{\alpha}^2 + G\bar{\alpha}^2\delta + H\delta^2$ under the model (4.1) with normally distributed errors e_{hi} and $n_h = 2$, $a_h = a$, $\beta_h = \beta$, $t_h = t$ and $\delta_h = \delta$ for all h, provided that the set of balanced half-samples for $v_B^{(1)}(\hat{\theta})$ is selected so that the number of units common to each pair is constant. (Evaluation of the MSE's of the remaining variance estimators appears difficult even in this special case.) The expressions for $\text{MSE}(v_L)$ and $\text{MSE}(v_B^{(1)})$ are lengthy, however, and are not reproduced here. Moreover, the latter quantity involves inverse moments of the form $E(X_1 + X_3)^{-a}(X_2 + X_3)^{-b}$ where X_1, X_2 and X_3 are independent gamma variates and a, $b = 1$ or 2. (P. S. R. S. Rao, 1974; Krewski and Chakrabarty, 1981).

We have evaluated the F, G and H coefficients in $\text{MSE}(v_L)$ and $\text{MSE}(v_B^{(1)})$ for the same values of a, L and t used earlier. We now summarize the results of this investigation. (i) Each of the coefficients in the expressions for the MSE's of $v_L(\hat{\theta})$ and $v_B^{(1)}(\hat{\theta})$ was found to decrease as L increases or as the CV $a^{-1/2}$, of the auxiliary variable decreases; (ii) for all values of a, L and t, $\text{MSE}(v_B^{(1)}) > \text{MSE}(v_L)$. The ratios $F_B^{(1)}/F_L$, $G_B^{(1)}/G_L$ and $H_B^{(1)}/H_L$ are all substantially greater than one for small values of L and a, but decrease as L or a increase. $H_B^{(1)}/H_L$ is particularly close to one for moderate values of La, indicating that the stability of $v_B^{(1)}(\hat{\theta})$ is comparable to that of $v_L(\hat{\theta})$ when $\bar{\alpha} \doteq 0$. (iii) The ratio $H_B^{(1)}/H_L$ decreases as t increases so that $\text{MSE}(v_B^{(1)})/\text{MSE}(v_L)$ decreases as t increases when $\bar{\alpha} \doteq 0$.

REFERENCES

BEAN, J. A. (1975). Distribution and properties of variance estimators for complex multistage probability samples. *Vital and Health Statist.* Ser. 2 No. 65. U.S. Government Printing Office, Washington, D.C.

BRILLINGER, D. R. (1977). Approximate estimation of standard errors of complex statistics based on sample surveys. *New Zealand Statistician* **11** 35–41.

CAMPBELL, C. and MEYER, M. (1978). Some properties of T confidence intervals for survey data. *Proc. Sec. Survey Res. Methods.* 437–442. (Amer. Statist. Assoc.) Washington, D.C.

FOLSOM, R. E., BAYLESS, D. L. and SHAH, B. V. (1971). Jackknifing for variance components in complex sample survey designs. *Proc. Soc. Statist. Sec.* 36–39.(Amer. Statist. Assoc.) Washington, D.C.

FRANKEL, M. R. (1971). *Inference from Survey Samples.* Inst. Social Res., Univ. Michigan, Ann Arbor.

GURNEY, M. and JEWETT, R. S. (1975). Constructing orthogonal replications for variance estimation. *J. Amer. Statist. Assoc.* **71** 819–821.

HOADLEY, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *Ann. Math. Statist.* **42** 1977–1991.

JONES, H. L. (1974). Jackknife estimation of functions of stratum means. *Biometrika* **61** 343–348.

KISH, L. and FRANKEL, M. R. (1974). Inference from complex samples (with discussion). *J. R. Statist. Soc. Ser. B* **36** 1–37.

KISH, L., NAMBOODIRI, N. K. and PILLAI, R. K. (1962). The ratio bias in surveys. *J. Amer. Statist. Assoc.* **57** 863–876.

KREWSKI, D. (1978a). Jackknifing *U*-statistics in finite populations. *Comm. Statist. Theor. Meth. A* **7** 1–12.

KREWSKI, D. (1978b). On the stability of some replication variance estimators in the linear case. *J. Statist. Plan. Inf.* **2** 45–51.

KREWSKI, D. and CHAKRABARTY, R. P. (1981). On the stability of the jackknife variance estimator. *J. Statist. Plan Inf.* **5** 71–79.

KREWSKI, D. and RAO, J. N. K. (1978). Inference from stratified samples I: large sample properties of the linearization, jackknife and balanced repeated replication methods. Tech. Report, Dept. Math., Carleton Univ.

KREWSKI, D. and RAO, J. N. K. (1981). Inference from stratified samples II: small sample properties of the linearization, jackknife and balanced repeated replication methods. Tech. Report, Dept. of Math., Carleton Univ.

LEE, K. (1973). Variance estimation in stratified sampling. *J. Amer. Statist. Assoc.* **68** 336–342.

LEMESHOW, S. and LEVY, P. (1978). Estimating the variance of ratio estimates in complex sample surveys with two primary units per stratum—a comparison of balanced replication and jackknife techniques. *J. Statist. Comp. Simul.* **8** 191–205.

MAJUMDAR, H. and SEN, P. K. (1978). Invariance principles for jackknifing *U*-statistics for finite population sampling and some applications. *Comm. Statist. Theor. Meth. A* **7** 1007–1025.

MCCARTHY, P. J. (1966). Replication: an approach to the analysis of data from complex surveys. *Vital and Health Statistics.* Ser. 2, No. 14, U. S. Government Printing Office, Washington, D.C.

MCCARTHY, P. J. (1969). Pseudoreplication: half-samples. *Rev. Internat. Statist. Inst.* **37** 239–264.

MILLER, R. G. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **35** 1594–1605.

MILLER, R. G. (1974). The jackknife—a review. *Biometrika* **61** 1–15.

PLACKETT, R. L. and BURMAN, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika* **33** 305–325.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). Wiley, New York.

RAO, J. N. K. and RAMACHANDRAN, V. (1974). Comparison of the separate and combined ratio estimators. *Sankhyā Ser. C* **36** 151–156.

RAO, P. S. R. S. (1974). Jackknifing the ratio estimator. *Sankhyā Ser. C.* **36** 84–97.

SEN, P. K. (1970). On some convergence properties of one-sample rank order statistics. *Ann. Math. Statist.* **41** 2140–2143.

SHAH, B. V., HOLT, M. M. and FOLSOM, R. E. (1977). Inference about regression models from sample survey data. *Bull. Internat. Statist. Inst.* **47** (Book 3) 43–57.

TEPPING, B. J. (1968). The estimation of variance in complex surveys. *Proc. Soc. Statist. Sec.* 11–18. (Amer. Statist. Assoc.) Washington, D.C.

WOODRUFF, R. S. (1971). Simple method for approximating the variance of a complex estimate. *J. Amer. Statist. Assoc.* **66** 411–414.

WOODRUFF, R. S. and CAUSEY, B. D. (1976). Computerized method for approximating the variance of a complex estimate. *J. Amer. Statist. Assoc.* **71** 315–321.

HEALTH AND WELFARE CANADA
355 RIVER ROAD
OTTAWA, CANADA K1A 1B8

DEPARTMENT OF MATHEMATICS AND STATISTICS
ARTS TOWER
CARLETON UNIVERSITY
OTTAWA, CANADA KIS 5B6