Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2003: Challenges in Survey Taking for the Next Decade

2003

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2003
Challenges in Survey Taking for the Next Decade

REPLICATION VARIANCE ESTIMATION FOR THE TWO-PHASE NATIONAL RESOURCES INVENTORY

Wayne A. Fuller[1]

ABSTRACT

The National Resources Inventory (NRI) is a study of land use based on a large area sample of the United States. The sample for 2002 is a two-phase sample in which the observations for 2002 are a subsample of the 1997 sample. A replication variance estimation method for two phase samples is described and applied to the NRI. The procedure reproduces the estimated variance of the first phase sample for those attributes of the first phase sample used as controls for the second phase sample. Only the second phase sample is required by the analyst for variance estimation.

KEYWORDS: Double Sampling; Jackknife; Multi-Phase Sampling; Reweighted Expansion Estimator.

1. INTRODUCTION

In two-phase sampling, also known as double sampling, observations are made on a vector of auxiliary variables in a large sample, called the first phase sample. Then observations are made on the variables of interest using a smaller sample, called the second phase sample. The smaller sample is typically a subsample of the original sample. Often the vector of auxiliary variables is a vector of indicator variables defining subgroups (strata) of the original sample used in selecting the second phase sample.

Rao (1973) and Cochran (1977) give formulas for variance estimation when the first phase is a simple random sample and the second phase is a stratified random sample. Särndal, Swensson and Wretman (1992, ch. 9) discuss two phase sampling. Kott (1990) derived a formula for variance estimation when the first phase is a stratified random sample and the second phase is a stratified random sample of the first phase sample based on strata defined by first-phase information. Rao and Shao (1992) proposed a jackknife variance estimation method in the context of hot deck imputation where the second phase strata correspond to imputation cells. Binder (1996) illustrated a "cookbook approach" to variance estimation for the two phase ratio estimator. Binder et al. (1997) derived formulas for variance estimation for various estimators for two-phase re-stratified sampling. Rao and Sitter (1995) and Sitter (1997) consider variance estimation for two-phase samples. Fuller (1998) proposed a replicate variance estimation method for the two-phase regression estimator that is particularly convenience for multipurpose surveys.

Kott and Stukel (1997) pointed out that, for a stratified second phase sample with the reweighted expansion estimator, the replication method proposed by Rao and Shao (1992) produces consistent variance estimates. Kim, Navarro and Fuller (2004) give a replication method for estimating the variance of the direct expansion estimator for a second phase stratified sample. We give the corresponding replication variance estimator for the two-phase regression estimator. In the last section, an application of replication variance estimation in the Natural Resources Inventory is described.

[1] Department of Statistics, Iowa State University, Ames, IA, 50011, U.S.A.

2. ASYMPTOTIC PROPERTIES

Let the finite population be of size N and indexed from 1 to N. Let the parameter of interest be the population total $Y = \Sigma_{i=1}^{N} y_i$, where y_i is the study variable and N is assumed to be known. Suppose we have a first phase sample of size n_1. We observe $\mathbf{x}_i$ for all $i \in A_1$ where $\mathbf{x}_i$ is a vector of auxiliary variables and A_1 is the set of indices of the n_1 first phase elements. If the second phase sample is stratified, indicators for strata are included in $\mathbf{x}_i$ where x_{ig} takes the value one if unit i belongs to the g-th group and is zero otherwise. A second phase sample of size n_2 is selected from the n_1 first phase elements.

We define the two-phase regression estimator of the mean of y by

$$\bar{y}_{reg} = \bar{y}_{\pi 2} + (\bar{\mathbf{x}}_{\pi 1} - \bar{\mathbf{x}}_{\pi 2})\hat{\boldsymbol{\beta}}_2 , \tag{2.1}$$

where

$$\hat{\boldsymbol{\beta}}_2 = \left(\sum_{i \in A_2} \pi_{2i}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{\pi 2})' (\mathbf{x}_i - \bar{\mathbf{x}}_{\pi 2}) \right)^{-1} \sum_{i \in A_2} \pi_{2i}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{\pi 2})' (y_i - \bar{y}_{\pi 2})$$

$$(\bar{\mathbf{x}}_{\pi 2}, \bar{y}_{\pi 2}) = \left(\sum_{i \in A_2} \pi_{2i}^{-1} \right) \sum_{i \in A_2} \pi_{2i}^{-1} (\mathbf{x}_i , y_i),$$

$$\bar{\mathbf{x}}_{\pi 1} = \left(\sum_{i \in A_1} \pi_{1i}^{-1} \right)^{-1} \sum_{i \in A_1} \pi_{1i}^{-1} \mathbf{x}_i ,$$

π_{1i} is the first phase selection probability for element i, $\pi_{2i} = \pi_{1i}\pi_{2i|1i}$ is the second phase selection probability for element i, and $\pi_{2i|1i}$ is the conditional probability of selecting element i for the second phase sample given that i is in the first phase sample. To derive the asymptotic properties of the two-phase regression estimator we assume a sequence of samples and finite populations such as that described in Fuller (1975).

Theorem 1. Let $\{F_N\}$ be a sequence of finite populations of size N. Let the G_N dimensional vectors $(1, \mathbf{x}_{N,i}, y_{Ni}) = \mathbf{z}_{Ni}$, $i = 1, 2, \ldots, N$, be a random sample from an infinite population with finite $2+\tau, \tau > 0$, moments. Assume $E\{\mathbf{z}'_{Ni}\mathbf{z}_{Ni}\}$ is positive definite for all N. Let first phase samples of size $n_{1,N}$, $n_{1,N} \geq n_{1,N-1}$, be selected by designs such that

$$V\{(\bar{\mathbf{z}}_{\pi,1,N} - \bar{\mathbf{z}}_N) | F_N\} = O_p\left(n_{1,N}^{-1/2}\right), \tag{2.2}$$

where

$$\bar{\mathbf{z}}_{\pi 1,N} = \left(\sum_{i \in A_{N1}} \pi_{1,Ni}^{-1} \right)^{-1} \sum_{i \in A_{N1}} \pi_{1,Ni}^{-1} \mathbf{z}_{Ni} .$$

A_{N1} is the set of indices in the first phase sample selected from population N, and $\pi_{1,Ni}$ is the probability of selecting element i for the first phase sample. Assume

$$E\left\{ \left| \bar{\mathbf{x}}_{\pi 2,N} - \bar{\mathbf{x}}_N \right|^2 \right\} = O\left(n_N^{-\delta}\right), \tag{2.3}$$

$$E\left\{ \left| \bar{\mathbf{x}}_{\pi 1,N} - \bar{\mathbf{x}}_N \right|^2 \right\} = O\left(n_N^{-\delta}\right), \tag{2.4}$$

$$E\left\{ \left| \hat{\boldsymbol{\beta}}_{2,N} - \boldsymbol{\beta}_N \right|^2 \right\} = O\left(n_N^{-\gamma}\right), \tag{2.5}$$

for some $\delta > 0, \gamma > 0$ and $\delta + \gamma = \lambda > 0.5$, where $\bar{\mathbf{x}}_{\pi 1}$ and $\hat{\boldsymbol{\beta}}_2$ are defined in (2.1), and

$$\boldsymbol{\beta}_N = \left[\sum_{i=1}^{N} \left(\mathbf{x}_i - \bar{\mathbf{x}}_N\right)' \left(\mathbf{x}_i - \bar{\mathbf{x}}_N\right) \right]^{-1} \sum_{i=1}^{N} \left(\mathbf{x}_i - \bar{\mathbf{x}}_N\right)' \left(y_i - \bar{y}_N\right).$$

Let $\bar{y}_{reg,N}$ be defined by (2.1). Then

$$\bar{y}_{reg,N} - \bar{y}_N = \bar{e}_{\pi 2,N} + \left(\bar{\mathbf{x}}_{\pi 1,N} - \bar{\mathbf{x}}_N\right)\boldsymbol{\beta}_N + O_p\left(n^{-0.5\lambda}\right), \tag{2.6}$$

where

$$\bar{e}_{\pi 2,N} = \left(\sum_{i \in A_2} \pi_{2,Ni}^{-1} \right)^{-1} \sum_{i \in A_2} \pi_{2,Ni}^{-1} e_{Ni},$$

and $e_{Ni} = y_{Ni} - \bar{y}_N - \left(\mathbf{x}_{Ni} - \bar{\mathbf{x}}_N\right)\boldsymbol{\beta}_N$.

Proof. To simplify the notation we omit the subscript *N*, except to identify finite population parameters. The estimator can be written.

$$\begin{aligned}
\bar{y}_{reg} - \bar{y}_N &= \bar{y}_{\pi,2} + \left(\bar{\mathbf{x}}_{\pi,1} - \bar{\mathbf{x}}_{\pi,2}\right)\hat{\boldsymbol{\beta}}_2 = \bar{y}_{\pi,2} + \left(\bar{\mathbf{x}}_{\pi,1} - \bar{\mathbf{x}}_{\pi,2}\right)\boldsymbol{\beta}_2 + \left(\bar{\mathbf{x}}_{\pi,1} - \bar{\mathbf{x}}_{\pi,2}\right)\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}\right) \\
&= \bar{y}_{\pi,2} - \left(\bar{\mathbf{x}}_{\pi,2} - \bar{\mathbf{x}}_N\right)\hat{\boldsymbol{\beta}}_2 + \left(\bar{\mathbf{x}}_{\pi,1} - \bar{\mathbf{x}}_N\right)\hat{\boldsymbol{\beta}}_2 + O_p\left(n^{-0.5\lambda}\right) \\
&= \bar{e}_{\pi,2} - \left(\bar{\mathbf{x}}_{\pi,2} - \bar{\mathbf{x}}_{\pi,1}\right)\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_N\right) + \left(\bar{\mathbf{x}}_{\pi,1} - \bar{\mathbf{x}}_N\right)\boldsymbol{\beta}_N + O_p\left(n^{-0.5\lambda}\right) \\
&= \bar{e}_{\pi,2} + \left(\bar{\mathbf{x}}_{\pi,1} - \bar{\mathbf{x}}_N\right)\boldsymbol{\beta}_N + O_p\left(n^{-0.5\lambda}\right).
\end{aligned} \tag{2.7}$$

establishing result (2.6).

The result permits the size of the vector **x** to increase with sample size. Thus a design that has a relatively large number of second phase strata is covered by the theorem. For a second phase stratified sample in which **x** contains stratum indicators, assumptions (2.3), (2.4), and (2.5) require the number of strata to increase at a slower rate than the sample size. The variance of the approximating random variable is

$$V\left\{d_{reg,N}\right\} = V\left\{\bar{e}_{\pi 2,N}\right\} + V\left\{\left(\bar{\mathbf{x}}_{\pi 1,N} - \bar{\mathbf{x}}_N\right)\boldsymbol{\beta}_N\right\} + 2C\left\{\bar{e}_{\pi 1,N}, \left(\bar{\mathbf{x}}_{\pi 1,N} - \bar{\mathbf{x}}_N\right)\boldsymbol{\beta}_N\right\}, \tag{2.8}$$

where $d_{reg,N} = \bar{e}_{\pi 2,N} + \left(\bar{\mathbf{x}}_{\pi 1,N} - \bar{\mathbf{x}}_N\right)\boldsymbol{\beta}_N$. Note that $V\left\{\bar{e}_{\pi 2}\right\}$ is the unconditional variance.

3. REPLICATE VARIANCE ESTIMATION

Let the second phase sample be a stratified sample with *G* strata and n_{2g} elements selected in the *g* th stratum. Let $\bar{y}_{2g}$ be the mean of the elements in second phase stratum *g*. Assume the first phase sample is such that the replication variance estimator of the estimated total is unbiased,

$$E\left\{ \sum_{k=1}^{L} c_k \left(\hat{T}_{1x}^{(k)} - \hat{T}_{1x}\right)^2 \right\} = V\left\{\hat{T}_{1x}\right\}, \tag{3.1}$$

where $\hat{T}_{1x}^{(k)}$ is the k th replicate of the estimated total, $\hat{T}_{1x}$ is the estimated total based on the first phase, L is the total number of replicates, and c_k, $k = 1, 2, \ldots, L$, are constants determined by the replication method. The two-phase estimated mean for y is

$$\bar{y}_{reg} = \sum_{g=1}^{G} \bar{x}_{1g} \bar{y}_{2g} , \tag{3.2}$$

where $\bar{x}_{1g}$ is the estimated fraction in second phase stratum g estimated from the first phase sample and $\bar{y}_{2g}$ is the mean of the second phase elements in second phase stratum g,

$$\bar{y}_{2g} = n_{2g}^{-1} \sum_{i \in A_{2g}} y_i , \tag{3.3}$$

n_{2g} is the number of elements in second phase stratum g and A_{2g} is the set of indices of elements in the g th stratum of the second phase sample.

To consider replicate variance estimation, assume the k th first phase replicate removes $n_{2,kg}$ elements from the g th second phase stratum. ($n_{2,kg}$ can be zero). If the $n_{2,kg}$ are a random sample of the n_{2g},

$$E\left\{ \left(\bar{y}_{2g}^{(k)} - \bar{y}_{2g}\right)^2 \middle| A_1 \right\} = \left[\left(n_{2g} - n_{2,kg}\right)^{-1} + n_{2g}^{-1} - 2n_{2g}^{-1}\right] s_{yg}^2 = \frac{n_{2,kg}}{(n_{2g} - n_{2,kg})\, n_{2g}} S_{yg}^2 ,$$

where S_{yg}^2 is the second phase variance of the g th stratum,

$$\bar{y}_g^{(k)} = \left(n_{2g} - n_{2,kg}\right)^{-1} \left(\sum_{i \in A_{2g}} y_i - \sum_{i \in A_{2g(k)}} y_i \right)$$

and $A_{2g(k)}$ is the set of indices of elements removed to create the k th replicate. Assume the multiplier for the first phase is c_k. Let the k th replicate of $\bar{y}_g$ be replaced with

$$\bar{y}_{2g}^{(k)*} = \bar{y}_{2g} + c_k^{-1/2} \left[n_{2g}^{-1}\left(n_{2g} - n_{2,kg}\right)\right]^{1/2} \left(y_{2g}^{(h)} - \bar{y}_{2g}\right). \tag{3.4}$$

Under the second phase random sample assumption,

$$\sum_{k=1}^{L} E\left\{ \left(\bar{y}_{2g}^{(k)*} - \bar{y}_g\right)^2 \right\} = n_{2g}^{-1} S_{yg}^2 .$$

Let

$$\bar{y}_{reg}^{(k)} = \sum_{g=1}^{G} \bar{x}_{1g}^{(k)} \bar{y}_{2g}^{(k)*} . \tag{3.5}$$

Then

$$\bar{y}_{reg}^{(k)} - \bar{y}_{reg} = \sum_{g=1}^{G} \bar{x}_{1g} \left(\bar{y}_{2g}^{(k)*} - \bar{y}_{2g}\right) + \sum_{g=1}^{G} \left(\bar{x}_{1g}^{(k)} - \bar{x}_{1g}\right) \bar{y}_{2g} + \sum_{g=1}^{G} \left(\bar{x}_{1g}^{(k)} - \bar{x}_{1g}\right) \left(\bar{y}_{2g}^{(k)*} - \bar{y}_{2g}\right), \tag{3.6}$$

where the last term on the right of the equality is of smaller order than the first two terms. The first phase design will determine the nature of the expectation of $\left(\bar{y}_{reg}^{(k)} - \bar{y}_{reg}\right)^2$. In a first phase simple random sample, $\bar{y}_g$ is conditionally unbiased for all positive $\bar{x}_{1g}$. In other designs it is possible for $\bar{x}_{1g}$ and $\bar{y}_g$ to be correlated.

Kim, Navarro and Fuller (2004) give conditions under which the jackknife variance estimator is consistent for the two-phase variance. It is required that:

(i) The approximation of (2.8) is adequate for the variance of the estimator
(ii) The second phase replicates furnish an adequate approximation for $V\left\{\bar{e}_{\pi,2}\right\}$.
(iii) The second order terms such as $\left(\bar{\mathbf{x}}_{\pi,1}^{(k)} - \bar{\mathbf{x}}_{\pi,2}^{(k)}\right)\left(\hat{\boldsymbol{\beta}}_2^{(k)} - \hat{\boldsymbol{\beta}}_2\right)$ are small relative to other terms.
(iv) The covariance term between $\bar{e}_{\pi,2}$ and $\bar{\mathbf{x}}_{\pi,1}$ is estimated by the replicates. If the k th second phase sample is composed of the second phase elements in the k-th first phase sample and the replicates meet other conditions, it seems reasonable to expect the covariance to be adequately estimated.

4. JACKKNIFE VARIANCE ESTIMATION FOR THE NRI

The U.S. National Resources Inventory (NRI) is conducted by the U.S. Natural Resources Conservation Service in cooperation with Center for Survey Statistics and Methodology of Iowa State University. The survey was conducted as a panel survey in 1982, 1987, 1992, and 1997. Data were collected on soil characteristics, land use, land cover, wind erosion, water erosion, and conservation practices. The sample is a stratified area sample of all states and Puerto Rico. The primary sampling units are areas of land called *segments.* Data are collected for the entire segment on items such as urban land, roads and water. Detailed data on soil properties and land use are collected at a random sample of points within the segment, where the typical number of points is three. The sample for 1997 contained about 300,000 segments and about 800,000 points. See Nusser and Goebel (1997) for a more complete description of the survey and Fuller (1999) for estimation procedures.

The 2002 sample is a subsample of about 102,000 segments selected from the 1997 sample. The subsample is a stratified sample of the 48 coterminous states, where the segments were placed in twelve strata in each state. The replication method used in NRI variance estimation is a form of “delete-a-group jackknife”. See Kott (2001) for a description. The 1997 NRI is the phase one sample, and the 2002 NRI is the phase two sample. For both the 1997 and the 2002 NRI, we suppose that we have completed an estimation process in which the estimation weights are calibrated for the auxiliary information at the population level and, in the case of the 2002 NRI, for a set of estimates from the 1997 NRI. The information at the population level is the total acres of federal land, acres in census water (water bodies greater than 40 acres and streams greater than 1/8 mile wide) and the acres in a federal program called the Conservation Reserve Program (CRP).

Let $w_{(1)i}$ and $w_{(2)i}$ represent the original phase one design weight (1997 NRI) and phase two (2002 NRI) design weight , respectively, for point *i*. Let $w_{(1)i}^*$ and $w_{(2)i}^*$ represent the fully calibrated regression weights. For any variable, y_i, the estimate for a 2002 total is computed as

$$\hat{t}_{y(2)} = \sum_{i\in A_2} w_{(2)i}^* y_i .$$

For a first phase variable, *x*, used to calibrate the second phase

$$\sum_{i\in A_1} w_{(1)i}^* x_i = \sum_{i\in A_2} w_{(2)i}^* x_i .$$

The goal of the variance estimation procedure for the 2002 NRI is to construct L new sets of weights $w_{(2)i}^{*(k)}$, $k = 1, 2, \ldots, L$, from which a user of the NRI data is able to compute replicate estimates $\hat{t}_{y(2)}^{(k)} = \sum_{A_2} w_{(2)i}^{*(k)} y_i$ for any variable y_i. The user can then calculate a variance estimate for $\hat{t}_{y(2)}$ as

$$\hat{V}\left\{\hat{t}_{y(2)}\right\} = \sum_{k=1}^{L} c_k \left(\hat{t}_{y(2)}^{(k)} - \bar{t}_{y(2)}\right)^2,$$

where c_k is a constant determined by the replication procedure and

$$\bar{t}_{y(2)} = L^{-1} \sum_{k=1}^{L} \hat{t}_{y(2)}^{(k)}.$$

As a first step, L replicate sets of phase one weights $w_{(1)i}^{*(k)}$ are constructed for the 1997 NRI. The sample is divided into L groups $G_{1,k}$, $k = 1, \ldots, L$, by ordering the segments geographically and then using systematic sampling to create 30 groups of approximately equal size. The program is flexible with L = 30 in the current version. The geographic ordering is used as an approximation to the stratification of the original sample.

Several options are available for forming weights. In one option with $c_k = 1$, the initial weights $w_{(1)i,0}^{*(k)}$ for jackknife replicate k are for $i \in G_{1,k}$,

$$w_{(1)i,0}^{*(k)} = \begin{cases} w_{(1)i}^{*} - a_L w_{(1)i} & \text{if } w_{(1)i}^{*} \geq w_{(1)i} \\ (1 - a_L) w_{(1)i}^{*} & \text{otherwise} \end{cases}$$

with $a_L = \sqrt{(L-1)/L}$, and for $i \notin G_{1,k}$,

$$w_{(1)i,0}^{*(k)} = \begin{cases} w_{(1)i}^{*} + (L-1)^{-1} a_L w_{(1)i} & \text{if } w_{(1)i}^{*} \geq w_{(1)i} \\ \left(1 + (L-1)^{-1} a_L\right) w_{(1)i}^{*} & \text{otherwise.} \end{cases}$$

Data are not collected on federal land and census water, but the points falling on federally owned land and falling on census water are kept in the data set so that the surface area of the country is fully represented. The acres in these classifications are known from external sources. Therefore we set $w_{(1)i,0}^{*(k)} = w_{(1)i}^{*}$ for every replicate if the point i is federal land or census water.

For the points that do not fall on federal or census water, the initial replicate weights are calibrated (through a raking procedure) to control acreages for CRP available for the previous years, resulting in the final replicate weights $w_{(1)i}^{*(k)}$.

Next, initial replicate weights $w_{(2)i,0}^{*(k)}$ are constructed for the phase two sample using a raking procedure and the first phase controls. Since the 2002 NRI sample was selected as a stratified sample from the 1997 NRI, the kth 2002 NRI replicate is the portion of the 2002 sample that is in the portion of the 2002 sample that is in the kth first phase replicate. The initial replicate weights are

$$w_{(2)i,0}^{*(k)} = \begin{cases} w_{(2)i}^{*} - a_L w_{(2)i} & \text{if } w_{(2)i}^{*} \geq w_{(2)i} \\ (1 - a_L) w_{(2)i}^{*} & \text{otherwise} \end{cases}$$

for $i \in G_{2,k}$, and

$$w_{(2)i,0}^{*(k)} = \begin{cases} w_{(2)i}^{*} + (L-1)^{-1} a_L w_{(2)i} & \text{if } w_{(2)i}^{*} \geq w_{(2)i} \\ \left(1 + (L-1)^{-1} a_L\right) w_{(2)i}^{*} & \text{otherwise} \end{cases}$$

for $i \notin G_{2,k}$, except if point i is in federal or census water. For federal and census water the final replicate weight $w_{(2)i}^{*(k)} = w_{(2)i}^{*}$, as for the first phase sample.

For the points that do not fall on federal or census water, the initial 2001 NRI replicate weights $w_{(2)i,0}^{*(k)}$ for each k are raked to a set of state-level estimates constructed with the corresponding final 1997 NRI replicate weights $w_{(1)i}^{*(k)}$. The categories for which the phase two estimates are calibrated to the phase one are the same as those used in the construction of the original weights $w_{(2)i}^{*}$, and include several broad landuse categories as well as a wetland classification. At the end of this raking step, we obtain the final replicate weights $w_{(2)i}^{*(k)}$, which are appended to the NRI 2002 dataset for the purpose of variance calculation.

At this writing the results of the 2002 NRI have not been released. Table 1 contains some illustrative results for one state. The entries should be multiplied by 10,000 to give reasonable acre values for a state, but the entries give realistic relative values. The 1997 values are the first phase sample estimates. The values for 2002 are based on the second phase sample that is about one third the size of the 1997 sample. Because of the correlation between 1997 and 2002 values, the standard errors for 2002 are less than the 1997 values multiplied by the square root of three. The computed standard error for large water is zero because large water is part of census water and is an auxiliary variable assumed to be known without error.

Table 1. Illustrative Results for a State

Coveruse	Year	
	1997	2002
Cultivated Cropland	2400	2300
	(11)	(14)
Noncultivated Cropland	110	120
	(6)	(7)
Urban	80	90
	(4)	(5)
Large Water	20	20
	(0)	(0)

ACKNOWLEDGEMENT

This research was funded in part by Cooperative Agreement 68-3A75-14 between the USDA Natural Resources Conservation Service and Iowa State University.

REFERENCES

Binder, D. A. (1996). "Linearization Methods for Single Phase and Two-Phase Samples: A Cookbook Approach", *Survey Methodology* 22, pp. 17-22.

Binder, D. A., C. Babyak, M. Brodeur, M. Hidiroglou and W. Jocelyn (1997), "Variance Estimation For Two-Phase Stratified Sampling", *ASA Proceedings of the Section on Survey Research Methods*, pp. 267-272.

Cochran, W. G. (1977), *Sampling techniques (3rd edition)*, New York: Wiley.

Fuller, W. A. (1975), "Regression analysis for sample survey", *Sankhyā, Series C, Indian Journal of Statistics* 37, pp. 117-132.

Fuller, W. A. (1998), "Replication variance estimation for two-phase samples", *Statistica Sinica* 8, pp. 1153-1164.

Fuller, W. A. (1999), "Estimation procedures for the United States National Resources Inventory", In *Proceedings of the Survey Methods Section of the Statistical Society of Canada*, pp. 39-44.

Kott, P. S. (1990), "Variance estimation when a first-phase area sample is restratified", *Survey Methodology* 16, pp. 99-103.

Kott, P. S. and D. M. Stukel (1997), "Can the Jackknife Be Used With a Two-Phase Sample?", *Survey Methodology* 23, pp. 81-89.

Kott, P. S. (2001), "The delete-a-group jackknife", *Journal of Official Statistics* 17, pp. 521-526.

Nusser, S. M. and J. J. Goebel (1997), "The national resources inventory: a long-term multi-resource monitoring programme", *Environmental and Ecological Statistics* 4, pp. 181-204.

Rao, J. N. K. (1973), "On double sampling for stratification and analytical surveys", *Biometrika* 60, pp. 125-133.

Rao, J. N. K., and J. SHAO (1992), "Jackknife variance estimation with survey data under hot deck imputation", *Biometrika* 79, pp. 811-822.

Rao, J. N. K. and R. R. SITTER (1995), "Variance estimation under two-phase sampling with application to imputation for missing data", *Biometrika* 82, pp. 453-460.

Särndal, C.-E., B. Swensson and J. Wretman (1992), *Model Assisted Survey Sampling*, Springer-Verlag.

Sitter, R. R. (1997), "Variance estimation for the regression estimator in two-phase sampling", *Journal of the American Statistical Association* 92, pp. 780-787.