SSC Annual Meeting, June 1996
Proceedings of the Survey Methods Section

JACKKNIFE VARIANCE ESTIMATION FOR THE POSTSTRATIFIED ESTIMATOR UNDER IMPUTATION

W. Yung and J.N.K. Rao[1]

ABSTRACT

Poststratification is a common technique used in sample surveys to improve the efficiency of estimates. In addition, imputation is usually used to compensate for item nonresponse. Unfortunately, treating the imputed values as true values and applying standard variance estimation techniques can lead to severe underestimates of the true variance. Rao and Shao (1992) proposed a design consistent jackknife variance estimation technique under a particular hot deck imputation scheme for the basic expansion estimator. In this paper, the Rao-Shao (1992) jackknife method is extended to the poststratified estimator under both mean imputation and hot deck imputation. The proposed jackknife variance estimator, as well as a linearization variance estimator obtained by linearizing the jackknife, are design consistent under both imputation methods. Theoretical results are presented as well as results from limited simulation studies.

KEY WORDS: Imputation; Jackknife variance estimator; Poststratification; Uniform response mechanism.

RÉSUMÉ

La stratification a posteriori est une technique standard utilisée dans les sondages pour améliorer l'efficacité des estimés. De plus, l'imputation est habituellement utilisée pour compenser la non-réponse partielle. Malheureusement, traiter les valeurs imputées comme de vraies valeurs et appliquer les techniques standard d'estimation de la variance, peut mener à sous-estimer de façon considérable la vraie variance. Rao et Shao (1992) ont proposé une technique d'estimation de la variance de type jackknife qui serait cohérente avec le plan d'enquête, sous un schéma particulier d'imputation hot-deck pour l'estimateur d'expansion élémentaire. Dans cet article, la méthode du jackknife de Rao-Shao (1992) est étendue à l'estimateur stratifié a posteriori sous les hypothèses d'imputation par la moyenne et d'imputation hot-deck. L'estimateur de la variance de type jackknife proposé, ainsi qu'un estimateur de la variance linéaire obtenue en linéarisant l'estimateur jackknife, sont convergent pour le plan avec les deux méthodes d'imputation. Des résultats théoriques sont présentés ainsi que des résultats d'études de simulation limitée.

MOTS CLÉS: Imputation; estimateur de la variance de type jackknife; stratification a posteriori; mécanisme de réponse uniforme.

1. INTRODUCTION

In complex large-scale surveys, poststratification is commonly used to increase the efficiency of estimators. Survey weights are adjusted so that the estimated number of units in each poststratum is equal to or will benchmark to known population counts. These adjusted weights are then used in estimating totals or means of variables collected in the survey. Along with poststratification, some form of imputation is commonly used to handle item nonresponse. Treating the imputed values as true values will lead to valid point estimates but applying standard variance estimation formulae can lead to serious underestimation when the proportion of missing values is appreciable. In order to correctly estimate the variance of an estimator under imputation, Rao and Shao (1992) proposed a jackknife variance estimation procedure based on adjusted imputed values. Under mild conditions, they showed that the proposed variance estimator is design consistent.

In this paper, we will extend the Rao-Shao jackknife to the poststratified estimator under mean imputation and a weighted hot deck imputation. In the following section, we describe the sample design used throughout this paper as well as introduce the poststratified estimator and its corresponding jackknife variance estimator under full response. The

[1] Wesley Yung, Household Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6; J.N.K. Rao, Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, K1S 5B6.

nonresponse situation is considered in section 3 with section 3.1 dealing with the poststratified estimator under mean imputation and section 3.2 dealing with the poststratified estimator under hot deck imputation. Simulation results are presented in section 4 for both the mean imputation and hot deck imputation cases.

2. POSTSTRATIFICATION UNDER FULL RESPONSE

Suppose we have L design strata with N_h clusters or primary sampling units (PSU's) in the h-th stratum. Within the h-th stratum, we select n_h (≥ 2) clusters and perform further subsampling within selected clusters according to some probability sampling design. We assume only that we have unbiased estimators of the cluster totals, Y_{hi}, $h = 1, \ldots, L$; $i = 1, \ldots, n_h$. From the survey design, we obtain the design weights, w_{hik}, associated with the k-th unit in the i-th sampled cluster within the h-th stratum . Also, associated with the (hik)-th unit (element) is the variable of interest, y_{hik}. An estimator of the total Y is given by

$$\hat{Y} = \sum_{(hik)\in s} w_{hik}\, y_{hik},$$

where s denotes the sampled elements.

Now, suppose we can partition the population into C mutually exclusive and exhaustive groups called poststrata. Poststrata are typically defined using variables which are available only after the sample is obtained (for example, age, sex or education). If the poststratum population counts are known, this auxiliary information can be used to increase the efficiency of the estimates. To define the poststratified estimator, we first define the poststratum indicator variable

$${}_c\delta_{hik} = \begin{cases} 1 & \text{if } (hik)\text{-th unit is in poststratum } c \\ 0 & \text{otherwise.} \end{cases}$$

Note that we use the prescript c to denote a poststratum. An estimator of the poststratum count ${}_cM$ is given by

$${}_c\hat{M} = \sum_{(hik)\in s} w_{hik}\, {}_c\delta_{hik}.$$

The poststratified weight is then defined as

$${}_cw_{hik} = w_{hik} \frac{{}_cM}{{}_c\hat{M}}$$

for (hik) in poststratum c and the poststratified estimator of the total is

$$\hat{Y}_{PS} = \sum_c \sum_{(hik)\in s} {}_cw_{hik} y_{hik}\, {}_c\delta_{hik}$$
$$= \sum_c \frac{{}_cM}{{}_c\hat{M}}\, {}_c\hat{Y}$$

where ${}_c\hat{Y} = \sum_{(hik)\in s} w_{hik}\, y_{hik}\, {}_c\delta_{hik}$ is the estimated poststratum total. We note that the poststratified weights, ${}_cw_{hik}$, benchmark to the known poststratum totals. By this, we mean that if the poststratified weights are used to estimate a poststratum count, one recovers the known population total.

For variance estimation we will use the jackknife variance estimator. To define the jackknife variance estimator for the poststratified estimator, we first define the jackknife weights when the (gj)-th cluster is deleted as

$$w_{hik(gj)} = \begin{cases} 0 & \text{if } (hi) = (gj) \\ n_g/(n_g - 1)\, w_{gik} & \text{if } h = g,\ i \neq j \\ w_{hik} & \text{otherwise.} \end{cases}$$

The estimator of the poststratum count, based on the sample with the (gj)-th cluster deleted, is

$${}_c\hat{M}_{(gj)} = \sum_{(hik)\in s} w_{hik(gj)}\, {}_c\delta_{hik}$$

and the corresponding poststratified jackknife weight is

$${}_cw_{hik(gj)} = w_{hik(gj)} \frac{{}_cM}{{}_c\hat{M}_{(gj)}}.$$

The estimator ${}_c\hat{M}_{(gj)}$ must be used instead of ${}_c\hat{M}$ in the definition of ${}_cw_{hik(gj)}$ or else the resulting variance estimator will overestimate the true variance (see Yung and Rao, 1996). Using the poststratified jackknife weights, the poststratified estimator when the (gj)-th cluster has been deleted is given as

$$\hat{Y}_{PS(gj)} = \sum_c \sum_{(hik)\in s} {}_cw_{hik(gj)}\, y_{hik}\, {}_c\delta_{hik}.$$

The jackknife variance estimator is then given by

$$v_J(\hat{Y}_{PS}) = \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\hat{Y}_{PS(gj)} - \hat{Y}_{PS}\right)^2.$$

Design consistency of the jackknife variance estimator is established in Yung (1996).

3. POSTSTRATIFICATION UNDER NONRESPONSE

In practice, even a carefully planned survey will suffer from nonresponse of some type. As outlined in Kalton and Kasprzyk (1986), it is common practice to distinguish between unit nonresponse, when none of the survey responses are available for a sampled element and item nonresponse, when some but not all of the responses are available. The distinction between unit and item nonresponse is useful since different methods are used to compensate for the missing data. In general, weighting adjustments are used for unit nonresponse while imputation methods are used for item nonresponse. In this paper, we consider the poststratified estimator under item nonresponse only. The poststratified estimator under weighting adjustments for unit nonresponse is considered in Yung (1996).

Some common imputation methods are: mean imputation, ratio imputation, regression imputation, random imputation, sequential hot deck imputation and weighted hot deck imputation. These imputation procedures are usually divided into two major groups: (1) Deterministic imputation which includes the first 3 methods listed above. The imputed values are fixed given the sample of respondents and any auxiliary information on the nonrespondents. (2) Hot deck imputation which includes the last three methods listed above. In hot deck imputation, the missing responses are replaced by values selected from the respondents.

Often the sample is divided into groups, called imputation classes, based on characteristics available for all individuals, and imputation is performed independently within each imputation class. For simplicity, we consider only the single imputation class case but note that the extension to multiple imputation classes is relatively straightforward. We also consider the poststratified estimator under mean and weighted hot deck imputation only. Throughout this paper we assume a uniform response mechanism. That is, we assume that the probability of response is the same for all sampled elements, and that the probability of response is independent between elements. In the case of multiple imputation classes, one would assume a uniform response mechanism within each imputation class.

3.1 Mean Imputation

One of the simplest and most intuitive imputation methods is mean imputation which simply imputes the weighted mean of the respondents for the missing values. Although simple, mean imputation has the disadvantage of leading to a spike at the mean of the respondents, thus distorting the distribution of values. In the traditional mean imputation procedure, for nonrespondents we impute the weighted mean of the respondents,

$$\tilde{y} = \frac{\sum_{(hik)\in s} w_{hik} y_{hik} a_{hik}}{\sum_{(hik)\in s} w_{hik}\, a_{hik}}$$

where a_{hik} is the response indicator variable defined as

$$a_{hik} = \begin{cases} 1 & \text{if } (hik)\text{-th unit responds} \\ 0 & \text{otherwise.} \end{cases}$$

After imputation, poststratification can be performed assuming that we know to which poststratum each sampled individual belongs regardless of response status. Alternatively, we can impute the poststratified estimator of the mean of the respondents,

$$\begin{aligned} \bar{y} &= \frac{\sum_c \sum_{(hik)\in s} {}_c w_{hik} y_{hik} a_{hik}\, {}_c\delta_{hik}}{\sum_c \sum_{(hik)\in s} {}_c w_{hik} a_{hik}\, {}_c\delta_{hik}} \\ &= \frac{\hat{S}}{\hat{T}} \end{aligned}$$

where

$$\hat{S} = \sum_c \sum_{(hik)\in s} {}_c w_{hik} y_{hik} a_{hik}\, {}_c\delta_{hik} \qquad (1)$$

and

$$\hat{T} = \sum_c \sum_{(hik)\in s} {}_c w_{hik} a_{hik}\, {}_c\delta_{hik}. \qquad (2)$$

This estimator is a more efficient estimator of the mean than $\tilde{y}$ and should thus lead to more "efficient" imputation.

Using the poststratified mean for imputation, the poststratified estimator of the total under imputation is

$$\hat{Y}^{I}_{PS} = \sum_c \sum_{(hik)\in s} {}_c w_{hik} y_{hik} a_{hik}\, {}_c\delta_{hik} + \sum_c \sum_{(hik)\in s} {}_c w_{hik} y^{*}_{hik} (1 - a_{hik})\, {}_c\delta_{hik}$$

where $y^{*}_{hik} = \bar{y} = \hat{S}/\hat{T}$ are the imputed values. It is easy to show that this estimator still benchmarks to the known population totals.

Turning to variance estimation, it is common practice to treat the imputed values as if they are true values and then compute the variance estimates using standard formulae. But this procedure can lead to serious underestimation of the true variance of the estimates when the proportion of missing values is appreciable. In order to correctly estimate the variance of an estimator under imputation, Rao and Shao (1992) proposed a new jackknife variance estimation method, for the weighted hot deck imputation procedure, based on adjusted imputed values. Assuming a uniform response mechanism and no poststratification, this new method leads to design consistent variance estimators. We now extend the Rao-Shao jackknife method to the poststratified estimator.

Under mean imputation, we adjust the imputed value, y^*_{hik}, by the amount $y^*_{hik(gj)} - y^*_{hik}$ where $y^*_{hik(gj)}$ is the value one would imputed for the (hik)-th nonrespondent if the (gj)-th cluster is deleted. That is, we replace the imputed values with

$$z^*_{hik(gj)} = y^*_{hik} + y^*_{hik(gj)} - y^*_{hik} = y^*_{hik(gj)}$$

where $y^*_{hik(gj)} = \hat{S}_{(gj)} / \hat{T}_{(gj)}$ with

$$\hat{S}_{(gj)} = \sum_c \sum_{(hik) \in s} {}_c w_{hik(gj)} y_{hik} a_{hik} \, {}_c\delta_{hik}, \qquad (3)$$

$$\hat{T}_{(gj)} = \sum_c \sum_{(hik) \in s} {}_c w_{hik(gj)} a_{hik} \, {}_c\delta_{hik} \qquad (4)$$

and ${}_c w_{hik(gj)}$ is the poststratified jackknife weight. We see that the adjustment simply corrects the imputed values to reflect the deletion of the (gj)-th cluster. Using these adjusted imputed values, we define the adjusted imputed poststratified estimator, when the (gj)-th cluster is deleted as

$$\hat{Y}^{Ia}_{PS(gj)} = \sum_c \sum_{(hik) \in s} {}_c w_{hik(gj)} y_{hik} a_{hik} \, {}_c\delta_{hik} + \sum_c \sum_{(hik) \in s} {}_c w_{hik(gj)} z^*_{hik(gj)} (1 - a_{hik}) \, {}_c\delta_{hik} .$$

The jackknife variance estimator is given by

$$v_J(\hat{Y}^I_{PS}) = \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\hat{Y}^{Ia}_{PS(gj)} - \hat{Y}^I_{PS} \right)^2 .$$

Asymptotic consistency of $v_J(\hat{Y}^I_{PS})$ has been established in Yung (1996).

3.2 Hot Deck Imputation

Hot deck imputation is commonly employed for item nonresponse for the following reasons: (a) It preserves the distribution of the item values unlike mean imputation which leads to a "spike" at the mean of the respondents for the item y. (b) Results obtained from different analyses are consistent with one another, unlike results from analyses from an incomplete data set. (c) It permits the use of the same survey weight for all items, unlike the weighting adjustment method.

In the simplest form of hot deck imputation, a simple random sample is selected with replacement from the sample respondents to an item y and the associated item values are used as donors. While this simple hot deck provides unbiased estimates for totals and means for simple random sampling under uniform response, the resulting estimator under a stratified multi-stage design will be biased. Unbiased estimation can be achieved by using a weighted hot deck procedure. In this procedure, donors are selected with replacement with probabilities proportional to their poststratified weights. That is, a donor is selected with probability

$${}_c w_{gjl} / \sum_c \sum_{(hik) \in s} {}_c w_{hik} a_{hik} \, {}_c\delta_{hik} .$$

The use of the poststratified weights will lead to "efficient" imputation. The poststratified estimator under imputation is

$$\hat{Y}^I_{PS} = \sum_c \sum_{(hik) \in s} {}_c w_{hik} y_{hik} a_{hik} \, {}_c\delta_{hik} + \sum_c \sum_{(hik) \in s} {}_c w_{hik} y^*_{hik} (1 - a_{hik}) \, {}_c\delta_{hik}$$

where y^*_{hik} are the imputed values obtained by using the poststratified weights in the weighted hot deck imputation procedure. This estimator benchmarks to known population totals.

To construct a correct jackknife variance estimator we define the adjusted imputed values when the (gj)-th cluster has been deleted as

$$z^*_{hik(gj)} = y^*_{hik} + \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}}$$

where $\hat{S}$, $\hat{T}$, $\hat{S}_{(gj)}$ and $\hat{T}_{(gj)}$ are as previously defined (see equations (1), (2), (3) and (4)). Letting E_* denote expectation with respect to the hot deck imputation given the sample of respondents, we have

$$E_*(z^*_{hik(gj)}) = E_*(y^*_{hik}) + \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}} - \frac{\hat{S}}{\hat{T}} = \frac{\hat{S}_{(gj)}}{\hat{T}_{(gj)}}$$

since $E_*(y^*_{hik}) = \hat{S}/\hat{T}$. Thus, on average we impute the properly adjusted mean of the respondents $\hat{S}_{(gj)}/\hat{T}_{(gj)}$. The adjusted imputed poststratified estimator when the (gj)-th cluster is deleted is given by

$$\hat{Y}^{Ia}_{PS(gj)} = \sum_c \sum_{(hik)\in s} {}_c w_{hik(gj)} y_{hik} a_{hik} {}_c\delta_{hik} + \sum_c \sum_{(hik)\in s} {}_c w_{hik(gj)} z^{*}_{hik(gj)} (1 - a_{hik}) {}_c\delta_{hik}.$$

The corresponding jackknife variance estimator is defined as

$$v_J(\hat{Y}^I_{PS}) = \sum_g \frac{n_g - 1}{n_g} \sum_j \left(\hat{Y}^{Ia}_{PS(gj)} - \hat{Y}^I_{PS}\right)^2.$$

Asymptotic consistency of $v_J(\hat{Y}^I_{PS})$ has been established in Yung (1996). Note that the adjusted imputed values are used only for variance estimation.

4. SIMULATION STUDY

A limited simulation study was performed to investigate the finite sample properties of the proposed jackknife variance estimators. For this purpose, we used a fixed finite population, considered by Valliant (1993), consisting of 10,841 persons included in the September 1988 Current Population Survey (CPS) of the United States. The variable of interest, y, is the weekly wages for each person. The single poststratifier was defined on the basis of age, race and sex (see Table 1 for details). For simplicity, only the single imputation class case is considered in the simulation study.

Table 1. Assignment of Age/Race/Sex Categories to Poststrata

	Nonblack		Black	
Age	Male	Female	Male	Female
19 and under	1	1	1	1
20-24	2	3	3	3
25-34	5	6	4	4
35-64	7	8	4	4
65 and over	2	3	3	1

NOTE: Cell numbers (1-8) are poststratum identification numbers.

The study population contained 2,826 geographical segments, each composed of about four neighbouring households. One hundred design strata ($L=100$) were created with each stratum having about the same total number of households. We used a stratified two-stage sampling design with segments as clusters and persons as the second-stage units. In each stratum $n_h=2$ segments were selected with probability proportional to the number of persons in each segment, and a simple random sample of $m_{hi}=4$ persons was selected without replacement if the sample segment contained more than four persons. In sample segments with four or fewer persons, all persons in the segment were selected. To simulate nonresponse, for each sampled individual a Bernoulli trial was performed with probability p for the outcome "response" where $p=0.9$ or 0.7. Using this design, we selected two sets of 10,000 independent samples, one set where mean imputation was used to impute for the missing values and the other set where the weighted hot deck was used for imputation. In both imputation methods, the poststratified weights were used.

From each sample, we computed the poststratified estimator and three variance estimators: the jackknife v_J, the naive jackknife v_{JN} which treats the imputed values as true values and a linearization variance estimator v_{JL} obtained by linearizing the jackknife variance estimator. The jackknife linearization variance estimator, v_{JL}, has been shown to be asymptotically equivalent to the jackknife in Yung (1996) and thus is consistent for the true variance of the poststratified estimator under both mean and hot deck imputation. Also, we computed the corresponding 95% normal theory confidence intervals for each variance estimator. To compare the performances of the variance estimators we compute the empirical relative bias (RB) for each variance estimator: the RB of a variance estimator is

$$RB = \frac{1}{MSE}\left[\frac{1}{10{,}000}\sum_i v_i\right] - 1$$

where v_i is the value of v for the i-th simulated sample ($i = 1, \ldots, 10{,}000$) and MSE is the empirical MSE of the estimator, say $\tilde{Y}$:

$$MSE = \frac{1}{10{,}000}\sum_i (\tilde{Y}_i - Y)^2$$

where $\tilde{Y}_i$ is the value of $\tilde{Y}$ in the i-th simulated sample.

Error rates (ER) for the normal theory 95% confidence intervals on the total Y were also calculated for each variance estimator:

$$\text{ER} = 1 - \frac{1}{10{,}000}(\text{\# of samples with } L_i \le Y \le U_i),$$

where $L_i \leq Y \leq U_i$ is a confidence interval on Y for the i-th simulated sample. Lower and upper error rates were calculated as

$$\text{lower ER} = \frac{1}{10{,}000}(\text{\# of samples with } Y<L_i),$$

and

$$\text{upper ER} = \frac{1}{10{,}000}(\text{\# of samples with } Y>U_i).$$

We also calculated the average lengths of the confidence intervals as

$$\text{average length} = \frac{1}{10{,}000}\sum_i (U_i - L_i).$$

Table 2 presents results, averaged over the 10,000 samples, for the mean imputation case. The correct jackknife variance estimators, v_J and v_{JL}, appear to be estimating the true variance of $\hat{Y}^I_{PS}$ with relative biases less than 2% and error rates just slightly over the nominal 5%. The similar behaviour of v_J and v_{JL} is as expected due to their asymptotic equivalence. On the other hand, the naive jackknife, v_{JN}, is underestimating the true variance with the underestimation becoming more severe as the response rate decreases. With a response rate of 70%, v_{JN} suffers a 63% under-estimation. The error rates corresponding to v_{JN} are also at unacceptable levels varying from 10% to 24% for a nominal 5% level, as the response rate decreases from 0.9 to 0.7. Turning to the average length of the confidence intervals, we see that the average length corresponding to v_J and v_{JL} are increasing as the response rate decreases while for v_{JN} the average length actually decreases. This indicates that v_J and v_{JL} are correctly tracking the increasing MSE as the response rate decreases while v_{JN} is remaining relatively stable as p decreases.

Results, averaged over the 10,000 samples, for the hot deck imputation case are presented in Table 3. We see that v_J and v_{JL} perform well with relative biases less than 3% and error rates slightly higher than the nominal 5%. We note that v_J and v_{JL} are expected to behave similarly due to their asymptotic equivalence. Turning to v_{JN}, we see that v_{JN} severely under-estimates the true variance of $\hat{Y}^I_{PS}$ with the difference increasing as the nonresponse rate increases. The error rates corresponding to v_{JN} are indicative of the underestimation as they are much larger than the nominal 5%. Turning to the average length of the confidence intervals, we see that the average length of the v_{JN} confidence intervals are constant while the lengths for the v_J and v_{JL} confidence intervals are increasing as the response rate decreases as was seen in the mean imputation case.

Table 2. Variance Estimation for the Poststratified Estimator under Mean Imputation

Performance Measure	v_J	v_{JL}	v_{JN}
		***p*=0.9**	
Relative Bias(%)	-1.77	-1.63	-28.19
Error Rate(%)	5.76	5.75	10.05
Lower Error Rate(%)	2.42	2.42	4.30
Upper Error Rate(%)	3.34	3.33	5.75
Average Length	4.06	4.05	3.47
		***p*=0.7**	
Relative Bias(%)	-1.48	-1.33	-63.52
Error Rate(%)	5.60	5.60	24.14
Lower Error Rate(%)	2.37	2.37	11.49
Upper Error Rate(%)	3.23	3.23	12.65
Average Length	4.66	4.66	2.83

Table 3. Variance Estimation for the Poststratified Estimator under Hot Deck Imputation

Performance Measure	v_J	v_{JL}	v_{JN}
		***p*=0.9**	
Relative Bias(%)	-0.77	-0.90	-18.24
Error Rate(%)	5.57	5.58	8.05
Lower Error Rate(%)	2.36	2.36	3.50
Upper Error Rate(%)	3.21	3.22	4.55
Average Length	4.27	4.27	3.87
		***p*=0.7**	
Relative Bias(%)	-2.61	-2.71	-44.15
Error Rate(%)	5.65	5.66	14.64
Lower Error Rate(%)	2.39	2.39	6.54
Upper Error Rate(%)	3.26	3.27	8.10
Average Length	5.21	5.20	3.94

5. CONCLUSIONS

We have proposed a new jackknife variance estimator for the poststratified estimator of a total under mean imputation and weighted hot deck imputation. The variance estimator is design consistent under uniform response within imputation classes. Our simulation results indicate that the proposed variance estimator and its linearized version perform well in terms of relative bias, whereas the naive jackknife variance estimator that treats the imputed values as true values leads to serious underestimation as the response rate decreases. Normal theory confidence intervals associated with the proposed variance estimators also perform well in terms of coverage rates.

Extensions to the case of two or more poststratifiers is under investigation. We also plan to study more complex parameters such as correlation and regression coefficients and cell proportions in a two-way table.

REFERENCES

Kalton, G., and Kasprzyk, D. (1986). "The Treatment of Missing Data", *Survey Methodology*, 12, 1-16.

Rao, J.N.K., and Shao, J. (1992). "Jackknife Variance Estimation with Survey Data under Hot Deck Imputation", *Biometrika*, 79, 811-822.

Valliant, R. (1993). "Poststratification and Conditional Variance Estimation", *Journal of the American Statistical Association*, 88, 89-96.

Yung, W. (1996). "Contributions to Poststratification in Stratified Multi-Stage Samples", unpublished Ph.D. thesis, Carleton University, Ottawa, Canada.

Yung, W., and Rao, J.N.K. (1996). "Jackknife Linearization Variance Estimators Under Stratified Multi-Stage Sampling", *Survey Methodology*, 22, 23-31.