Journal of the Korean Statistical Society (2000), **29**: *4, pp 385–394*

Jackknife Variance Estimation under Imputation for Nonrandom Nonresponse with Follow-ups

Jinwoo Park [1]

Abstract

Jackknife variance estimation based on adjusted imputed values when nonresponse is nonrandom and follow-up data are available for a subsample of nonrespondents is provided. Both hot-deck and ratio imputation method are considered as imputation method. The performance of the proposed variance estimator under nonrandom response mechanism is investigated through numerical simulation.

Keywords: Double sampling estimate; Hot-deck imputation; Ratio imputation; Multiple imputation; Nonrandom nonresponse; Jackknife variance estimation.

1. INTRODUCTION

Almost all sample surveys suffer from nonresponse. If nonresponse is ignored, it may cause some bias. Unit nonresponse is customarily handled by weighting adjustment method. On the other hand, item nonresponse is usually handled by some form of imputation to fill in missing item values. Since imputation results in a complete data set, such method allows the use of complete data methods of analysis. However, after imputation, treating the imputed values as observed values may lead to serious underestimation of variances of point estimators (Hansen et al. 1953).

Many researchers attempted to get remedies for this problem. Rubin(1987) proposed multiple imputation(MI) to estimate the variance due to imputation by replicating the process a number of times and estimating the between replicate variation. However, Rao(1996) pointed out several drawbacks in MI, such as high costs for storage, inconsistency etc.. Rao and Shao (1992) proposed an adjusted jackknife variance estimation under hot-deck. Since hot-deck imputation is rather convenient and needs smaller data set than MI, many statistical agencies

[1]Department of Applied Statistics, The University of Suwon, Suwon, 445-743, Korea

prefers hot-deck imputation. Furthermore, Kovar and Chen(1994) showed that it performs well in many cases.

All the methods cited above have assumed only random nonresponse which is appropriate when the nonresponse happens at random. To get an appropriate estimate with nonrandom nonresponse, more information on the nonrespondents is necessary. Glynn et al.(1993) introduced many possible ways to obtain more information on the nonrespondents. Furthermore, they applied MI to the mean estimation when nonresponse is nonrandom and follow-up data are available for a subsample of nonrespondents. Since MI has some drawbacks as Rao(1996) pointed out, it seems to be desirable to find method based on single imputation for nonrandom nonresponse.

The purpose of this paper is to suggest new estimates based on single imputation method to mean estimation when nonresponse is nonrandom and follow-up data are available for a subsample of nonrespondents. Section 2 cites the double sampling estimate with follow-up data and provides a new mean estimator and its jackknife variance estimator after a single imputation. Hot-deck imputation and ratio imputation is considered as imputation methods. In section 3, some numerical study results are briefly outlined. Finally, some concluding remarks are offered in section 4.

2. ESTIMATION WITH FOLLOW-UP DATA

2.1. The Double sampling Estimate

Suppose that the initial random sample of size n results in n_1 respondents and n_0 nonrespondents. Since nonresponses are nonrandom, a follow-up random sample of size n_{01} is taken from the nonrespondents. Assume that all follow-up sample data are acquired. The following Figure 1. shows the population and the sample, where the shaded parts represent sample respondents.

Let $\bar{y}$ be the sample mean and let s^2 be the sample variance of the sample of size n, which may contain missing values due to nonresponse. Further, let $\bar{y}_1$ and s_1^2 are the sample mean and the sample variance for the initial respondents, and $\bar{y}_{01}$, s_{01}^2 those for the n_{01} followed-up respondents among nonrespondents.

The standard estimator for the population mean, $\bar{Y}$, of the finite population is the weighted mean of $\bar{y}_1$ and $\bar{y}_{01}$.

$$\hat{\bar{Y}} = \frac{(n_1\bar{y}_1 + n_0\bar{y}_{01})}{n} \tag{2.1}$$

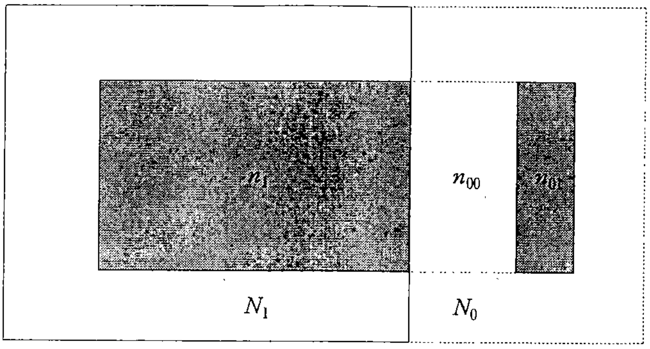

Figure 1. The population and the sample

Cochran(1977) derived the variance of $\hat{\bar{Y}}$, and suggested the associated unbiased variance estimator for large N and n, which are as follows.

$$V(\hat{\bar{Y}}) = \frac{(1-f)}{n}S^2 + W_2\frac{(1-k)}{n}S_0^2, \tag{2.2}$$

$$v(\hat{\bar{Y}}) = (\frac{1}{n} - \frac{1}{N})\frac{n_1}{n}s_1^2 + (\frac{n_0}{nn_{01}} - \frac{1}{N})\frac{n_0}{n}s_{01}^2 + \frac{N-n}{(N-1)n}\frac{n_0 n_1}{n^2}(\bar{y}_{01} - \bar{y}_1)^2, \tag{2.3}$$

where $W_2 = N_2/N$ and $k = n_{01}/n_1$.

2.2. Estimation after Imputation

Suppose that the initial random sample of size n results in n_1 respondents and n_0 nonrespondents. A follow-up random sample of size n_{01} is taken from the n_0 nonrespondents. Further, an imputation is applied to the final nonrespondents, whose size is $n_{00} = n_0 - n_{01}$. In this case n_{01} follow-up data serve as doners.

First of all, the estimator of initial respondents mean is the usual mean, $\bar{y}_1$, and its variance esimator is usual also. Let $\bar{y}_{01}$ be the average of n_{01} follow-up data , and y^*_{0j}, $j \in A_{mis}$ be the imputed value. A_{mis} and A_{res} represent the sample of nonrespondents and respondents respectively. The estimator of nonrespondents mean after imputation, $\bar{y}_{0I}$, is

$$\bar{y}_{0I} = (\sum_{i \in A_{res}} y_{0i} + \sum_{j \in A_{mis}} y^*_{0j})/n_0.$$

Thus, a mean estimator, $\hat{\bar{Y}}_I$, can be written as a weighted average of $\bar{y}_1$ and $\bar{y}_{0I}$. That is,

$$\hat{\bar{Y}}_I = w_1\bar{y}_1 + w_0\bar{y}_{0I}, \tag{2.4}$$

where $w_i = n_i/n$, $i = 0, 1$.

The following theorem shows that the estimator given by (2.4) is unbiased, and derives the variance of sample mean, $\hat{\bar{Y}}$. The proof of theorem 2.1 is given in the Appendix.

Theorem 2.1. *Suppose that the initial random sample of size n results in n_1 respondents and n_0 nonrespondents. A follow-up random sample of size n_{01} is taken from the n_0 nonrespondents. Assume that all follow-up sample data are acquired, and that imputation is applied to the final nonrespondents, whose size is $n_{00} = n_0 - n_{01}$. Then $\hat{\bar{Y}}_I$ at (2.4) is unbiased, and the variance of it, $V(\hat{\bar{Y}}_I)$, is as follow.*

$$\begin{aligned} V(\hat{\bar{Y}}_I) &= \frac{W_0[1+(n-1)W_0]}{n} \cdot V(\bar{y}_{0I}) + \frac{W_1[1+(n-1)W_1]}{n} \qquad (2.5)\\ &\cdot \Big(W_0(2-\frac{n_{01}}{n_0})\frac{S_0^2}{n} + \frac{W_1}{n}S_1^2 + \frac{W_0}{n}(\bar{Y}_0-\bar{Y})^2 + \frac{W_1}{n}(\bar{Y}_1-\bar{Y})^2\Big). \end{aligned}$$

Since the variance of $\bar{y}_{0I}$ (Kim, 1997) can be represented as

$$V(\bar{y}_{0I}) = (\frac{1}{n_0} - \frac{1}{N_0})S_0^2 + (\frac{1}{n_{01}} - \frac{1}{n_0})S_0^2 + \frac{n_{00}}{n_0^2}(1 - \frac{n_{00}}{n_{01}})S_0^2,$$

we can estimate S_0^2 by

$$\hat{S}_0^2 = \frac{\hat{V}(\bar{y}_{0I})}{1/n_{01} + (1 - n_{00}/n_{01})n_{00}/n_0^2}.$$

Thus, by plugging $\hat{S}_0^2$, w_0 and w_1, and $\hat{V}(\bar{y}_{0I})$ into (2.5), we can get the following variance estimator which is an asymptotically unbiased estimator of $V(\hat{\bar{Y}})$.

$$\begin{aligned} \hat{V}(\hat{\bar{Y}}) &= \frac{w_0[1+(n-1)w_0]}{n} \cdot \hat{V}(\bar{y}_{0I}) + \frac{w_1[1+(n-1)w_1]}{n} \qquad (2.6)\\ &\cdot \Big(w_0(2-\frac{n_{01}}{n_0})\frac{\hat{S}_0^2}{n} + \frac{w_1}{n}s_1^2 + \frac{w_0}{n}(\bar{y}_{0I}-\hat{\bar{Y}}_I)^2 + \frac{w_1}{n}(\bar{y}_1-\hat{\bar{Y}}_I)^2\Big). \end{aligned}$$

2.2.1. Hot Deck Imputation

When hot deck imputation is used, Rao and Shao (1992) proposed the following adjusted jackknife variance estimator of $\bar{y}_{0I}$, which is asymptotically unbiased.

$$v_{0I(HD)} = \frac{n_0 - 1}{n_0} \sum_{i=1}^{n_0} (\bar{y}_{0I}^a(-i) - \bar{y}_{0I})^2, \qquad (2.7)$$

where the i-th replicate is

$$\bar{y}^a_{0I}(-i) = \begin{cases} \frac{1}{n_0-1}[n_0\bar{y}_{0I} - y_{0i} + \sum_{j\in A_{mis}} z_{0ji}] & \text{if i} \in A_{res}, \\ \frac{1}{n_0-1}[n_0\bar{y}_{0I} - y^*_{0i}] & \text{if i} \in A_{mis}. \end{cases}$$

and $z_{0ji} = y^*_{0j} + \bar{y}_{01}(-i) - \bar{y}_{01}$. Therefore, we can get $\hat{V}(\hat{\bar{Y}}_I)$ by substituting (2.7) to $\hat{V}(\bar{y}_{0I})$ in (2.6).

2.2.2. Ratio Imputation

Suppose that an auxiliary variable, x, closely related to an item y is observed on all sample units, and that the ratio $\bar{y}_{01}/\bar{x}_{01}$ and $\bar{y}_0/\bar{x}_0$ is the same, that is, the follow-up sample is randomly selected from A_{mis} . Under ratio imputation, we impute the predicted value in place of the missing y_{0j} as follows:

$$y^*_{0j} = \frac{\bar{y}_{01}}{\bar{x}_{01}} \cdot x_{0j}.$$

Then the adjusted jackknife variance estimator of $\bar{y}_{0I}$ is

$$v_{0I(Ratio)} = \frac{n_0 - 1}{n_0} \sum_{i=1}^{n_0} (\bar{y}^a_{0I}(-i) - \bar{y}_{0I})^2, \tag{2.8}$$

where the i-th replicate is

$$\bar{y}^a_{0I}(-i) = \begin{cases} \frac{1}{n_0-1}[n_0\bar{y}_{0I} - y_{0i} + \sum_{j\in A_{mis}} (y^*_{0j} + z_{oji})] & \text{if i} \in A_{res}, \\ \frac{1}{n_0-1}[n_0\bar{y}_{0I} - y^*_{0i}] & \text{if i} \in A_{mis}. \end{cases}$$

and $z_{0ji} = \frac{\bar{y}_{01}(i)}{\bar{x}_{01}(i)} \cdot x_{0j} - \frac{\bar{y}_{01}}{\bar{x}_{01}} \cdot x_{0j}$, $\bar{y}_{01}(i) = (n_{01}\bar{y}_{01} - y_{0i})/(n_{01} - 1)$ and a similar expression $\bar{x}_{01}(i)$. Therefore, we can get $\hat{V}(\hat{\bar{Y}})$ by substituting (2.8) to $\hat{V}(\bar{y}_{0I})$ in (2.6). Furthermore, Rao (1996) suggested a linearized version of as following $v_L(\bar{y}_{0I})$.

$$v_L(\bar{y}_{0I}) \doteq (\frac{\bar{x}_0}{\bar{x}_{01}})^2 \cdot \frac{A}{n_{01}} + 2(\frac{\bar{x}_0}{\bar{x}_{01}}) \cdot \frac{B}{n_0} + \frac{C}{n_0}, \tag{2.9}$$

where

$$\begin{aligned} A &= \sum_1^{n_{01}} (y_{0i} - \frac{\bar{y}_{01}}{\bar{x}_{01}} \cdot x_{0i})^2/(n_{01} - 1), \\ B &= \frac{\bar{y}_{01}}{\bar{x}_{01}} \cdot \sum_1^{n_{01}} (y_{0i} - \frac{\bar{y}_{01}}{\bar{x}_{01}} \cdot x_{0i}) \cdot x_{0i}/(n_{01} - 1), \\ C &= (\frac{\bar{y}_{01}}{\bar{x}_{01}})^2 \cdot \sum_1^{n_{01}} (x_{0i} - \bar{x}_0)^2/(n_0 - 1). \end{aligned}$$

So we can also get $\hat{V}(\hat{\bar{Y}})$ by substituting (2.9) to $\hat{V}(\bar{y}_{0I})$ in (2.6).

3. SIMULATION RESULTS

To evaluate the performance of the proposed estimators in section 2, some numerical simulations are considered. Rao and Sitter (1995) suggested the following model.

$$y_i = \beta x_i + \sqrt{x}_i \cdot \epsilon_i,$$

where

$$\begin{aligned} x_i &\sim \Gamma(g, h), \\ \epsilon_i &\sim N(0, \sigma^2). \end{aligned}$$

Thus,

$$\begin{aligned} \mu_x &= gh, \sigma_x^2 = gh^2, C_x = \sigma_x/\mu_x = 1/\sqrt{g}, \\ \mu_y &= \beta\mu_x, \sigma_y^2 = \beta^2\sigma_x^2 + \mu_x\sigma_x, corr(x_i, y_i) = \rho = \beta\sigma_x/\sigma_y. \end{aligned}$$

We generated 500 samples of size $n = 400$ units with $\beta = 1$, $\mu_x = 100$ and choosing σ_x, σ_y to match specified values of ρ and C_x. The probability of nonresponse is assumed to be related to the $x-$ variable in two distinct ways:

$$\begin{aligned} P_L &= 1 - exp(-c_L x), \\ P_S &= exp(-c_S x), \end{aligned}$$

where the constants c_L and c_S are chosen such that an expected 25(Kovar and Chen, 1994). In the model, P_L implies that large units are more likely not to respond, and P_S implies that small units are more likely not to respond. The results are summarized in Table 1 and Table 2 .

Table 1 provides estimates under the response probability P_L . All three estimates - the double sampling estimate, the estimate under hot deck imputation, the estimate under ratio imputation - seem to give valid estimates. The standard error (s.e.) of the estimate under imputation is slightly larger than that of the double sampling estimate, but the difference is negligible. The efficiency of the estimate under ratio imputation increases slightly as ρ increases.

Table 2 gives estimates under the response probability P_S which implies that small units are more likely not to respond. Here also all three estimates - the double sampling estimate, the estimate under hot deck imputation, the estimate under ratio imputation - seem to provide valid estimates. The standard error (s.e.) of each estimate is similar, irrespective of ρ . The above results implies that the proposed estimation methods after imputation is appropriate for a nonrandom nonresponse with follow-ups.

Table 1. Estimates of the mean when the nonresponse is nonrandom with response probability P_L.

$C_x = 1.0$		size	mean	s.e.	$C_x = 1.4$		size	mean	s.e.
sample		400	102.2	6.4388	sample		400	108.0	9.0410
respondents		292	73.9	5.8792	respondents		304	59.5	6.2653
follow-ups		28	170.7	20.7849	follow-ups		25	222.7	49.7560
double sampling		320	100.0	7.3844	double sampling		329	98.7	13.3199
$\rho = 0.7$	hot-deck imp		101.6	9.5855	$\rho = 0.7$	hot-deck imp		97.1	17.0305
	ratio imp		107.7	9.1095		ratio imp		114.2	16.2186
$\rho = 0.8$	hot-deck imp		101.8	9.5733	$\rho = 0.8$	hot-deck imp		96.6	16.4362
	ratio imp		108.1	8.5406		ratio imp		112.7	14.9481
$\rho = 0.9$	hot-deck imp		101.0	8.6875	$\rho = 0.9$	hot-deck imp		97.3	16.0987
	ratio imp		106.5	7.0786		ratio imp		112.7	13.5987

Table 2. Estimates of the mean when the nonresponse is nonrandom with response probability P_S.

$C_x = 1.0$		size	mean	s.e.	$C_x = 1.4$		size	mean	s.e.
sample		400	102.2	6.4388	sample		400	108.0	9.0410
respondents		305	126.0	7.8527	respondents		289	147.2	11.683
follow-ups		24	28.38	8.1404	follow-ups		28	5.04	3.7673
double sampling		329	102.9	6.6266	double sampling		317	107.8	9.0825
$\rho = 0.7$	hot-deck imp		102.2	6.8257	$\rho = 0.7$	hot-deck imp		108.3	10.0452
	ratio imp		101.8	6.5933		ratio imp		108.1	10.1127
$\rho = 0.8$	hot-deck imp		102.1	6.8040	$\rho = 0.8$	hot-deck imp		108.1	9.1024
	ratio imp		102.0	6.5950		ratio imp		108.1	9.1379
$\rho = 0.9$	hot-deck imp		102.5	7.0110	$\rho = 0.9$	hot-deck imp		107.9	9.0998
	ratio imp		102.5	6.4854		ratio imp		107.9	9.1283

4. CONCLUDING REMARKS

In this article, new estimation method after imputation for nonrandom nonresponse have been suggested. Considering on the fact that there were no available estimation method based on single imputation for nonrandom nonresponse in spite of the prevalance of single imputation, the work done by this study may contribute.

The suggested method is appropriate when follow-up data are available on a random subsample of the nonrespondent. For the ranges of cases considered, the suggested estimation method works quiet well. New estimate gives similar performance with the classical double sampling estimate. According to Glynn et al.(1993), performances of estimators based on MI is similar to the double sampling case. Hence, it seems that the performances of estimates based on single(hot-deck or ratio) imputation and on MI are similar. Since there are many drawbacks in MI, it is recommendable to use single imputation method.

Appendix A: Proof of Theorem 2.1

First, to show the unbiasedness of $\hat{\bar{Y}}_I$, find the conditional expectation of it for given n_1.

$$\begin{aligned} E(\hat{\bar{Y}}_I|n_1) &= E_1E_2(\hat{\bar{Y}}_I) = E_1E_2(w_1\bar{y}_1 + w_0\bar{y}_{0I}) \\ &= E_1(w_1\bar{y}_1 + w_0\bar{y}_0) \\ &= w_1\bar{Y}_1 + w_0\bar{Y}_0, \end{aligned}$$

where E_1 denotes averaging over all possible selections from each stratum and E_2 denotes averaging over all possible imputations from follow-up data. Hence, averaging over repeated selections of n_1,

$$\begin{aligned} E(\hat{\bar{Y}}_I) &= EE(\hat{\bar{Y}}_I|n_1) \\ &= E(w_1\bar{Y}_1 + w_0\bar{Y}_0) \\ &= \bar{Y}_1E(w_1) + \bar{Y}_0E(w_0) \\ &= W_1\bar{Y}_1 + W_0\bar{Y}_0. \end{aligned}$$

Next, $V(\hat{\bar{Y}}_I)$ can be written as $V(\hat{\bar{Y}}_I) = EV(\hat{\bar{Y}}_I|n_1)$. $V(\hat{\bar{Y}}_I|n_1)$ can be decomposed as $V(\hat{\bar{Y}}_I|n_1) = V_1E_2(\hat{\bar{Y}}_I) + E_1V_2(\hat{\bar{Y}}_I)$, where E_1 denotes averaging over all possible selections from each stratum and E_2 denotes averaging over all possible imputations from follow-up data. The first term on the right is derived as

$$\begin{aligned} V_2(\hat{\bar{Y}}_I) &= V_2(w_0\bar{y}_{0I}) = w_0^2 V_2(\bar{y}_{0I}) \\ E_1V_2(\hat{\bar{Y}}_I) &= E_1(w_0^2 V_2(\bar{y}_{0I})) = w_0^2 V(\bar{y}_{0I}). \end{aligned}$$

On the other hand, the second term is

$$\begin{aligned} V_1E_2(\hat{\bar{Y}}_I) &= V_1(w_1\bar{y}_1 + w_0\bar{y}_{01}) \\ &= (1-n/N)\cdot S^2/n + W_0(1-n_{01}/n_0)\cdot S_0^2/n. \end{aligned}$$

. Then, by adding the above equations, it can be shown that

$$\begin{aligned} V(\hat{\bar{Y}}_I|n_1) &= E_1V_2\hat{\bar{Y}}_I + V_1E_2\hat{\bar{Y}}_I \\ &= w_0^2\cdot V(\bar{y}_{0I}) + w_1^2\cdot[(1-n/N)\cdot S^2/n + W_0(1-n_{01}/n_0)\cdot S_0^2/n] \\ &\approx w_0^2\cdot V(\bar{y}_{0I}) + w_1^2\cdot[S^2/n + W_0(1-n_{01}/n_0)\cdot S_0^2/n]. \end{aligned}$$

Following Cochran(1977), S^2 can be decomposed as follow.

$$\frac{S^2}{n} \approx \frac{W_1}{n}\cdot S_1^2 + \frac{W_0}{n}\cdot S_0^2 + \frac{W_0}{n}\cdot(\bar{Y}_0-\bar{Y})^2 + \frac{W_1}{n}\cdot(\bar{Y}_1-\bar{Y})^2.$$

So

$$\begin{aligned} V(\hat{\bar{Y}}_I|n_1) &\approx w_0^2\cdot V(\bar{y}_{0I}) + w_1^2\cdot[S^2/n + W_0(1-n_{01}/n_0)\cdot S_0^2/n] \\ &\approx w_0^2\cdot V(\bar{y}_{0I}) + w_1^2\cdot[W_0(1-n_{01}/n_0)\cdot S_0^2/n] \\ &+ w_1^2\cdot[\frac{W_1}{n}\cdot S_1^2 + \frac{W_0}{n}\cdot S_0^2 + \frac{W_0}{n}\cdot(\bar{Y}_0-\bar{Y})^2 + \frac{W_1}{n}\cdot(\bar{Y}_1-\bar{Y})^2]. \end{aligned}$$

Therfore, by averaging over repeated selections of n_1, we can show that

$$\begin{aligned} EV(\hat{\bar{Y}}_I|n_1) &\approx E(w_0^2\cdot V(\bar{y}_{0I}) + w_1^2\cdot[W_0(1-n_{01}/n_0)\cdot S_0^2/n] \qquad (A.1) \\ &+ w_1^2\cdot[\frac{W_1}{n}\cdot S_1^2 + \frac{W_0}{n}\cdot S_0^2 + \frac{W_0}{n}\cdot(\bar{Y}_0-\bar{Y})^2 + \frac{W_1}{n}\cdot(\bar{Y}_1-\bar{Y})^2]). \end{aligned}$$

Since n_1 is a random variable from hypergeometric distribution, we can say that

$$\begin{aligned} E(n_1) &= n\cdot N_1/N = n\cdot W_1 \\ V(n_1) &\approx n\cdot N_1/N\cdot(1-N_1/N) \\ E(n_1^2) &\approx V(n_1) + [E(n_1)]^2 = nW_1[1+(n-1)W_1]. \end{aligned}$$

By substituting the above results for (A.1), we can get the following formula.

$$\begin{aligned} V(\hat{\bar{Y}}_I) &= \frac{W_0[1+(n-1)W_0]}{n}\cdot V(\bar{y}_{0I}) + \frac{W_1[1+(n-1)W_1]}{n} \\ &\cdot \Big(W_0(2-\frac{n_{01}}{n_0})\frac{S_0^2}{n} + \frac{W_1}{n}S_1^2 + \frac{W_0}{n}(\bar{Y}_0-\bar{Y})^2 + \frac{W_1}{n}(\bar{Y}_1-\bar{Y})^2\Big). \end{aligned}$$

REFERENCES

Cochran, W. G.(1977), Sampling Techniques(3rd ed.), New York: John Wiley.

Ford, B. L.(1983), " An Overview of Hot-Deck Procedures", *In Incomplete Data in Sample Surveys*, Vol. 2, Academic Press, pp 185-207.

Glynn, R. J., Laird, N. M., and Rubin, D. B. (1993), "Multiple Imputation in Mixture Models for Nonignorable Nonresponse with Follow-ups", *Journal of American Statistical Association*, Vol. 88, 423, pp 984-993.

Hansen, M., Hurwitz, W. , and Madow, W.(1953), *Sample Survey Methods and Theory.(Vol. 2)*, New York: John Wiley.

Kim, J. G. (1997), "Jackknife Variance Estimation Under Hot-Deck Imputation", Unpublished handscript.

Kovar, J. G. and Chen, E. J.(1994), "Jackknife Variance Estimation of Imputed Survey Data", *Survey Methodology*, Vol. 20, 1, pp 45-52.

Rao, J. N. K. and Shao, J. (1992), "Jackknife Variance Estimation under Hot Deck Imputation", *Biometrika*, Vol. 79, pp811-822.

Rao, J. N. K. and Sitter, R. R. (1995), "Variance Estimation under Two-phase Sampling with Application to Imputation for Missing Data", *Biometrika*, Vol. 82, pp 453-460.

Rao, J. N. K.(1996), "On Variance Estimation with Imputed Survey Data", *Journal of American Statistical Association*, Vol. 91, pp 499-512.

Rubin, D. B.(1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley.