Journal of the Korean Statistical Society (2002), **31**: *4, pp 469–482*

A Combined Method Compensating for Wave Nonresponse†

Jinwoo Park[1]

Abstract

This paper suggests a new method of compensating for wave nonresponse in panel survey, which combines weighting adjustment and imputation. By deleting less frequent nonresponse patterns, we can get simplicity. A new mean estimator under the new combining method is provided and a limited simulation study employing a real data is conducted.

Keywords. Wave nonresponse, panel survey, attrition, weight adjustment, imputation, hot deck, adjusted jackknife variance estimator.
AMS 2000 subject classifications. Primary 62D05; Secondary 62G09.

1. Introduction

Nonresponse occurs frequently in surveys. Nonresponse is generally considered to be of two types: unit nonresponse and item nonresponse. Compensating for unit nonresponse is customarily made by weighting adjustment method. On the other hand, item nonresponse is usually handled by some form of imputation.

In addition to unit and item nonresponse, another type of nonresponse, named wave nonresponse, occurs in panel survey. Wave nonresponse occurs when one or more waves of panel data are missing for a unit that has provided data for at least one wave. Both weighting and imputation may be used to compensate for missing data due to wave nonresponse.

Rizzo *et al.* (1994), and Folsom and Witt (1994) used weighting adjustments methods for panel nonresponse in the Survey of Income and Program Participation (SIPP). Kalton and Miller (1986) conducted a simulation study among three-wave respondents from the first three waves of the SIPP 1984 panel. They

Received November 2001; accepted August 2002.
†This research was supported by a grant from Korea Science and Engineering Foundation (KOSEF 2000-1-10400-008-1).
[1]Department of Applied Statistics, Suwon University, Suwon, Gyeonggi-do 445-743, Korea

addressed the use of imputation and weighting for wave nonresponse and offered a number of insights into the quality of the two strategies for the problem of wave nonresponse. Lepkowski (1989) examined missing data compensation strategies for wave nonresponse. Weighting adjustment method is easier to implement and preserves the relationship in the observed data. However, it may give poorer quality nonresponse compensation than imputation. Furthermore, it requires many different sets of weights to execute longitudinal analysis. Imputation is easier to use and is simple to meet all analytic objectives. However, it reduces the ability to detect important relationships among survey variables through attenuation of the strength of observed covariances.

Because of the complementary strengths and weaknesses of the weighting and imputation strategies, it is natural to think of combinations of the two approaches. Lepkowski (1989) presented two kinds of combined approaches: "imputation for completing attrition nonresponse" approach and "imputation for completing wave nonrespondents" approach. The objective of this study is to suggest a new method of compensating for wave nonresponse in panel survey, which combines weighting adjustment and imputation. In Section 2, the idea of combining method is provided. Estimation of mean under the new combining method is described in Section 3. A numerical example is considered in Section 4 and finally some concluding remarks are given in Section 5.

2. New Combining Method

2.1. Wave nonresponse patterns

In preparation for the description of a new wave nonresponse compensation method, it is useful to introduce wave nonresponse patterns in panel survey. Figure 2.1 shows several patterns of wave nonresponse that may occur in a three wave panel survey. x and o denote a wave response and a wave nonresponse, respectively. The shaded part represents wave nonresponse. There are wave response/nonresponse patterns for the three-wave panel. Among wave nonresponse patterns, patterns in which the respondent appears in an early wave and then fails to respond at later waves are called attrition patterns, and all the other wave nonresponse patterns are called non-attrition patterns. While xxo and xoo belong to attrition patterns, xox, oxx, oxo, oox belong to non-attrition patterns.

The frequency distribution of the wave response/nonresponse patterns will vary across surveys depending on the survey topic, survey organization, and other factors. Table 2.1 presents the frequency distribution of respondents to

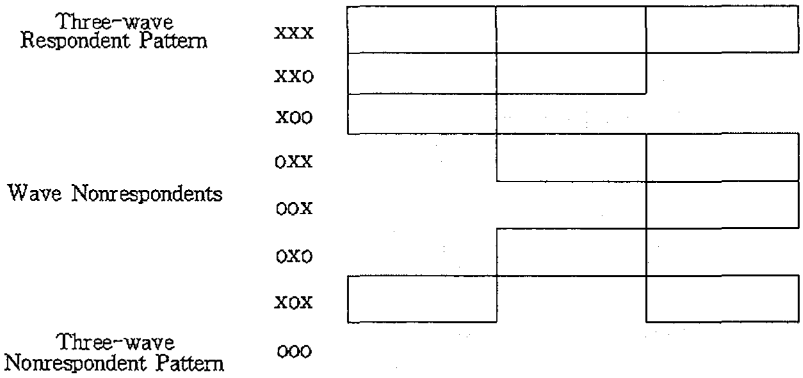

FIGURE 2.1 *Wave nonresponse patterns for a three-wave panel survey*

two panel surveys: The Income Survey Development Program 1979 (ISDP) and The 1984 Survey of Income and Program Participation (SIPP) (Lepkowski, 1989). The data in Table 2.1 are for the first three waves of each panel. In both ISDP and SIPP the largest percentage of persons are three-wave respondents (80.2%, 90.0%). Among wave nonreponse patterns, the attrition patterns are the next most frequent patterns (7.2%, 6.7%; 4.9%, 4.2%). The nonattrition patterns are the least frequent (2.3%, 2.2%, 0.6%, 0.9%; 0.9%). While the percentage distribution of ISDP contains wave 1 nonrespondents, the SIPP 1984 Panel does not since wave 1 nonrespondents were not followed at later waves.

TABLE 2.1 *Response patterns for the first three waves*

Response Pattern		*ISDP*	*SIPP*
Respondents	xxx	80.2%	90.0%
Attrition Wave	xxo	7.2%	4.9%
Nonrespondents	xoo	6.7%	4.2%
	xox	2.3%	0.9%
Nonattrition Wave	oxx	2.2%	-
Nonrespondents	oxo	0.6%	-
	oox	0.9%	-

2.2. New method

Considering the complementary strengths and weaknesses of the weighting and imputation methods, it is natural to try to combine the two approaches. Lepkowski (1989) introduced two combining methods such as the imputation for completing wave nonresponse approach and imputation for completing attribution nonresponse approach. Imputation for completing wave nonresponse approach uses imputation to complete wave nonresponses for those patterns in which only one wave is missing (*e.g.*, patterns xxo, xox and oxx), and the remaining wave nonresponse patterns deleted and compensated by weighting. On the other hand, imputation for completing attrition nonresponse approach first uses imputation to convert nonattrition patterns into attrition patterns and then employs attrition weighting methods to compensate for the remaining wave nonresponse. For example, the nonattrition patterns oxx and xox could be completed by imputation, and the nonattrition patterns oxo and oox are deleted.

As can be seen in Table 2.1, the percentage of the non-attrition pattern is very low. However, many complicated problems arise to compensate for wave nonresponse of that pattern. Hence, by deleting less frequent patterns, it is possible to get simplicity and practicability. From now on, only three wave panel survey is dealt with for simplicity. However, it can easily be extended to more than three wave panel survey. We use data of such patterns, xxx, xxo, xoo, xox, only and discard data of the other patterns, oxx, oxo, oox, ooo (Type D and E). Figure 2.2 shows the above explanations schematically. The shaded part indicates data which are discarded.

In the first wave, we use only weighting adjustment method for nonresponse. In other words, no imputation is taken to compensate for nonresponse. The initial weights for survey units are replaced with the adjusted weights. Let N be the population size and let n sample size. Denote n_1 for the number of respondents in the wave 1. In this case $n - n_1$ nonrespondents are discarded from the sample. Thus the sample size for analysis is n_1.

Nonresponse in the wave 2, which is represented by B and C type in Figure 2.2, is filled with by imputation method. Denote n_2 for the number of respondents in the wave 2 among n_1 units which are respondents in the wave 1. Then n_2^* values are imputed by using the $n_1 - n_2$ responded data as doners in the wave 2. For nonresponse in the wave 3, which is represented by A and B type, some consideration is needed. While for type A it is simple to fill out, for type B there may be many choices. To fill nonresponse in the wave 3, using the imputed value

Response Pattern	Type
xxx xxx xxx ... xxx xxx xxx	three-wave response
xxo	A
xoo	B
xox	C
oxx oxo oox	D
ooo	E

FIGURE 2.2 *Response pattern of three-wave panel*

in the wave 2 may be a choice. Taking a new value through another imputation can be another choice. Here we take another imputation for nonresponse in the wave 3, which is independent of the imputation used in the wave 2, since that make matters easier. In this case, the same unit may take imputed values from different doners as wave changes.

3. Estimation

3.1. Notation

We will consider the case of simple random sampling of size n from the population of size N. Let y_{ij} be the value of a variable y for the j^{th} unit in the wave i ($i = 1, 2, 3$; $j = 1, 2, \ldots, N$). Denote n_1, n_2, n_3, the numbers of responses for wave 1, 2, 3, respectively and n_3' the number of respondents in the wave 3 who responded in the wave 2, and n_3'' that who did not respond in the wave 2. Then $n_3 = n_3' + n_3''$. Since nonresponded units in the wave 1 are to be discarded from the sample in subsequent wave, it is sure that both n_2 and n_3 are less than n_1. The

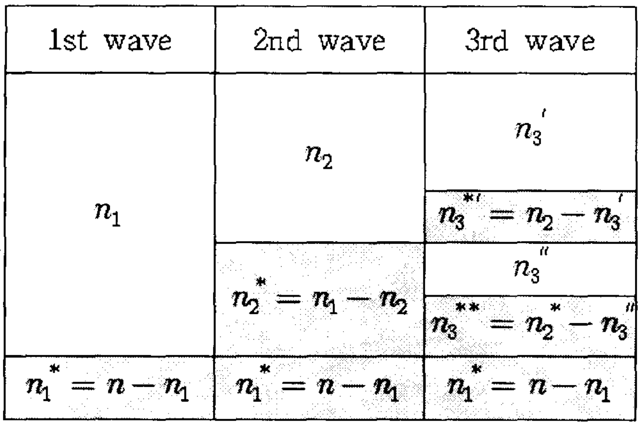

FIGURE 3.1 *The number of response at each wave*

following Figure 3.1 shows the number of response at each wave schematically.

Denote the number of nonresponse in the wave 1 as $n_1^* = n - n_1$. Similarly, let n_2^*, n_3^* be the number of nonresponse in the wave 2 and 3, then $n_2^* = n_1 - n_2$ and $n_3^* = n_1 - n_3 = n_3^{*\prime} + n_3^{**}$, where $n_3^{*\prime} = n_2 - n_3'$ and $n_3^{**} = n_2^* - n_3''$. For n_2^* nonrespondents in the wave 2, imputation is taken. We assume a single imputation class for simplicity. Denote y_{kj}^* as the imputed value for the j^{th} nonresponded unit ($j = 1, \ldots, n_i^*$) in the wave k ($k = 2, 3$). For several means we have the following definitions:

$$\bar{Y}_{(k)} = \frac{1}{N}\sum_{j=1}^{N} y_{kj}\text{: population mean in the wave k,}$$

$$\bar{y}_{(k)} = \frac{1}{n}\sum_{j=1}^{n} y_{kj}\text{: sample mean in the wave k,}$$

$$\bar{y}_{(k)r} = \frac{1}{n_k}\sum_{j=1}^{n_k} y_{kj}\text{: respondents' mean in the wave k.}$$

3.2. Cross-sectional mean

In case of no missing data, the sample mean in the wave 1, $\bar{y}_{(1)} = (1/n) \sum_{j=1}^{n} y_{1j}$, is an unbiased estimator of the population mean, $\bar{Y}_{(1)} = (1/N)\sum_{j=1}^{N} y_{1j}$. However, if n_1^* $(= n - n_1)$ nonresponse occur in the wave 1, we use a weighting adjustment method to compensate for missing data as have mentioned in Section 2. Then $\bar{y}_{(1)}$ can be rewritten as

$$\bar{y}_{(1)} = \frac{1}{n}\sum_{j=1}^{n} y_{1j} = \frac{1}{n}\left[n_1\bar{y}_{(1)r} + n_1^*\bar{y}_{(1)nr}\right]. \tag{3.1}$$

Since there are no data available to calculate $\bar{y}_{(1)nr}$, some weighting adjustment is made to each responded unit. That is, after removing $(n_1^*/n)\bar{y}_{(1)nr}$ from (3.1), replace the weight $1/n$ for the j^{th} unit $(j = 1, \dots, n_i^*)$ with $(1/n)(n/n_1) = 1/n_1$. Suppose that the nonresponse happens at random, then an unbiased estimate of $\bar{Y}_{(1)}$ is provided by

$$\hat{\bar{Y}}_{(1)w} = \bar{y}_{(1)r} = \frac{1}{n_1}\sum_{j=1}^{n_1} y_{1j}. \tag{3.2}$$

The sampling variance of $\hat{\bar{Y}}_{(1)w}$ in (3.2) is given by

$$V(\hat{\bar{Y}}_{(1)w}) = \left(\frac{1}{n_1} - \frac{1}{N}\right)S_{(1)}^2, \tag{3.3}$$

where

$$S_{(1)}^2 = \frac{1}{N-1}\sum_{1}^{N}\left(y_i - \bar{Y}_{(1)}\right)^2.$$

By plugging the sample variance of responsdents, $s_{(1)r}^2$, into $S_{(1)}^2$, we can get an estimator of variance (3.3). For the population mean in the wave 2, 3, $\bar{Y}_{(k)} = (1/N)\sum_{j=1}^{N} y_{kj}$, $k = 2, 3$, the following estimator is unbiased in case of no missing data:

$$\bar{y}_{(k)} = \frac{1}{n_1}\sum_{j=1}^{n_1} y_{kj} = \frac{1}{n_1}\left[n_k\bar{y}_{(k)r} + n_k^*\bar{y}_{(k)nr}\right]. \tag{3.4}$$

Here, we use imputation method to compensate for $\bar{y}_{(k)nr}$. Denote y_{kj}^* as the imputed value for the missing value y_{kj}, $j \in A_{(k)mis}$, where $A_{(k)mis}$ represents the set of missing units in the wave k. Then unbiased estimate of $\bar{Y}_{(k)} = (1/N)\sum_{j=1}^{N} y_{kj}$ is given by

$$\hat{\bar{Y}}_{(k)I} = \frac{1}{n_1}\left[n_k\bar{y}_{(k)r} + n_k^*\bar{y}_{(k)nr}^*\right], \tag{3.5}$$

where $\bar{y}^*_{(k)nr} = (1/n^*_k)\sum_{j=1}^{n^*_k} y^*_{kj}$. Many kinds of variance estimators, such as adjusted jackknife variance estimator (Rao and Shao, 1992), balanced half sample variance estimation (Shao and Rao, 1996; Shao, Chen and Chen, 1997), bootstrap variance estimator (Shao and Sitter, 1998), model-assisted variance estimator (Särndal, 1992), are developed to estimate $V(\hat{\hat{Y}}_{(k)I})$. We use the following adjusted jackknife variance estimator in this paper. Let

$$v_{0I(RS)} = \frac{n_0 - 1}{n_0}\sum_{i=1}^{n_0}\left(y^{\bar{a}}_{oI}(-i) - \bar{y}_{oI}\right)^2, \tag{3.6}$$

where the i^{th} replicate is

$$y^{\bar{a}}_{oI}(-i) = \begin{cases} \frac{1}{n_0 - 1}\Big[n_{01}\bar{y}_{01} - y_{0i} + \sum_{j\in A_{mis}}\big(y^*_{oj} + \bar{y}_{o1}(-i) - \bar{y}_{01}\big)\Big], & \text{if } i \in A_{res}, \\ \frac{1}{n_0 - 1}\big[n_{01}\bar{y}_{01} - y^*_{0i}\big], & \text{if } i \in A_{mis} \end{cases}$$

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

3.3. Estimation of mean change between different waves

The key advantage of panel survey is its abilities to measure mean or mean change between different waves. Estimation of mean change between different waves is considered here. The mean change between the wave k and the wave k' is denoted as $\bar{Y}_{(k)} - \bar{Y}_{(k')}$. First, consider the mean change between wave k ($k = 2, 3$) and wave $k' = 1$. The following is an estimator of $\bar{Y}_{(k)} - \bar{Y}_{(1)}$:

$$\widehat{\bar{Y}_{(k)} - \bar{Y}_{(1)}} = \hat{\hat{Y}}_{(k)I} - \hat{\hat{Y}}_{(1)w}. \tag{3.7}$$

If the imputation is performed such that a simple random sample of size n^*_1 is selected with replacement from $A_{(2)res}$ in the wave 2, where $A_{(2)res}$ denotes the set of respondents, then the above estimator in (3.5) can be represented as

$$\hat{\hat{Y}}_{(2)I} = \frac{1}{n_1}(1 + d_{2i})y_{2i}, \tag{3.8}$$

where d_{2i} is the number of times that unit i is used as a doner, $i \in A_{(2)res}$. Then

$$E_I\big(\hat{\hat{Y}}_{(k)I} - \hat{\hat{Y}}_{(1)w}\big) \equiv E_I\big(\hat{\hat{Y}}_{(k)I} - \hat{\hat{Y}}_{(1)w}|R, D\big) = \bar{y}_{(k)r} - \bar{y}_{(1)},$$

$$V_I\big(\hat{\hat{Y}}_{(k)I} - \hat{\hat{Y}}_{(1)w}\big) \equiv V_I\big(\hat{\hat{Y}}_{(k)I} - \hat{\hat{Y}}_{(1)w}|R, D\big) = \frac{n^*_k}{n_1^2}\Big(1 - \frac{1}{n_k}\Big)S^2_{(k)r},$$

where I denotes imputation, R denotes response mechanism, and D denotes sampling mechanism. So,

$$
\begin{aligned}
&V\big(\hat{\bar{Y}}_{(k)I} - \hat{\bar{Y}}_{(1)w}\big) \\
&= V\big(\bar{y}_{(k)r} - \bar{y}_{(1)}\big) + E\left[\frac{n_k^*}{n_1^2}\Big(1 - \frac{1}{n_k}\Big)S_{(k)r}^2\right] \\
&= V_D E_R\big(\bar{y}_{(k)r} - \bar{y}_{(1)}\big) + E_D V_R\big(\bar{y}_{(k)r} - \bar{y}_{(1)}\big) \\
&\quad + E_D E_R\left[\frac{n_k^*}{n_1^2}\Big(1 - \frac{1}{n_k}\Big)S_{(k)r}^2\right] \\
&= V\big(\bar{y}_{(k)} - \bar{y}_{(1)}\big) + \Big(\frac{1}{n_k} - \frac{1}{n_1}\Big)S_{(k)}^2 + \frac{n_k^*}{n_1^2}\Big(1 - \frac{1}{n_k}\Big)S_{(k)}^2 \\
&= V\big(\bar{y}_{(k)}\big) + V\big(\bar{y}_{(1)}\big) - 2\,\mathrm{Cov}\big(\bar{y}_{(k)}, \bar{y}_{(1)}\big) + \Big(\frac{1}{n_k} - \frac{1}{n_1}\Big)S_{(k)}^2 \\
&\quad + \frac{n_k^*}{n_1^2}\Big(1 - \frac{1}{n_k}\Big)S_{(k)}^2.
\end{aligned}
\tag{3.9}
$$

Various estimators for $V(\bar{y}_{(k)})$ and $V(\bar{y}_{(1)})$ have been suggested by many authors. Therefore, it needs only to find a reasonable estimator of $\mathrm{Cov}\big(\bar{y}_{(k)}, \bar{y}_{(1)}\big)$. Since there are n_k units which are respondents both in the wave k and in the wave 1, use only the data on n_k respondents to estimate $\mathrm{Cov}\big(\bar{y}_{(k)}, \bar{y}_{(1)}\big)$. Then, the following (3.10) can be an estimator of $\mathrm{Cov}\big(\bar{y}_{(k)}, \bar{y}_{(1)}\big)$:

$$
\widehat{\mathrm{Cov}}\big(\bar{y}_{(k)}, \bar{y}_{(1)}\big) = \frac{s_{(1)k}^{*2}}{n_k}, \tag{3.10}
$$

where

$$
\begin{aligned}
s_{(1)k}^{*2} &= \frac{1}{n_2 - 1}\sum_1^{n_2}\big(y_{1j} - \bar{y}_{(1)}^*\big)\big(y_{kj} - \bar{y}_{(k)}^*\big), \\
\bar{y}_{(1)}^* &= \frac{1}{n_2}\sum_1^{n_2} y_{1j}, \quad \bar{y}_{(k)}^* = \frac{1}{n_2}\sum_1^{n_2} y_{kj}.
\end{aligned}
$$

Next, consider the mean change between the wave 3 and the wave 2. A natural estimator of the mean change, $\bar{Y}_{(3)} - \bar{Y}_{(2)}$, is

$$
\widehat{\bar{Y}_{(3)} - \bar{Y}_{(2)}} = \hat{\bar{Y}}_{(3)I} - \hat{\bar{Y}}_{(2)I}. \tag{3.11}
$$

Since both $\hat{\bar{Y}}_{(3)I}$ and $\hat{\bar{Y}}_{(2)I}$ contain imputed values, it is very complicate to derive the variance of the above estimator. Thus, to make matters easier, consider a modified estimator of $\bar{Y}_{(3)} - \bar{Y}_{(2)}$ such that

$$\widehat{\bar{Y}_{(3)} - \bar{Y}_{(2)}} = \hat{\bar{Y}}^*_{(3)I} - \bar{y}_{(2)r}, \tag{3.12}$$

where

$$\hat{\bar{Y}}^*_{(3)I} = \frac{1}{n_2}\Big(\sum_1^{n_3'} y_{3j} + \sum_{n_3'+1}^{n_2} y^*_{3j}\Big).$$

Note that only n_2 units, respondents in the wave 2, are used. It is straightforward to show the unbiasedness of (3.12). The following is the variance of the estimator (3.12).

$$\begin{aligned} V\big(\hat{\bar{Y}}^*_{(3)I} - \bar{y}_{(2)r}\big) &= V\big(\tilde{y}_{(3)}\big) + V\big(\bar{y}_{(2)}\big) - 2\mathrm{Cov}\big(\tilde{y}_{(3)}, \bar{y}_{(2)}\big) \\ &\quad + \Big(\frac{1}{n_3'} - \frac{1}{n_2}\Big)S'^2_{(3)} + \frac{n_3^{*\prime}}{n_2^2}\Big(1 - \frac{1}{n_3'}\Big)S^{*\prime 2}_{(3)}, \end{aligned} \tag{3.13}$$

where

$$\begin{aligned} S^{*\prime 2}_{(3)} &= \frac{1}{n_2 - 1}\sum_1^{n_2}\big(y_{3j} - \tilde{y}_{(3)}\big)^2, \\ \tilde{y}_{(3)} &= \frac{1}{n_2}\sum_1^{n_2} y_{3j}. \end{aligned}$$

The above estimator can be derived easily with similar calculation done in (3.8).

Since there are n_3' units which are respondents both in the wave 2 and in the wave 3, use only the data on n_3' respondents to estimate $\mathrm{Cov}\big(\tilde{y}_{(3)}, \bar{y}_{(2)}\big)$. Then, an estimator of $\mathrm{Cov}\big(\tilde{y}_{(3)}, \bar{y}_{(2)}\big)$ is as follows:

$$\widehat{\mathrm{Cov}}\big(\tilde{y}_{(3)}, \bar{y}_{(2)}\big) = \frac{s^{**2}_{(23)}}{n}, \tag{3.14}$$

where

$$\begin{aligned} s^{**2}_{(23)} &= \frac{1}{n_3' - 1}\sum_1^{n_3'}\big(y_{2j} - \bar{y}^{**}_{(2)}\big)\big(y_{3j} - \bar{y}^{**}_{(3)}\big), \\ \bar{y}^*_{(2)} &= \frac{1}{n_3'}\sum_1^{n_3'} y_{2j}, \bar{y}^{**}_{(3)} = \frac{1}{n_3'}\sum_1^{n_3'} y_{3j}. \end{aligned}$$

4. Numerical Simulation

In order to test the performance of the proposed method, we conduct a limited simulation study employing a real data from the Korean Household Income and Expenditure Survey (KHIES). The size of the KHIES sample (Seoul), n, is 1025. We consider only three waves and use amount of expenditure as the study variable. Since all the data are complete response data, we simulate different missing situations.

In the first wave, we consider two missing types: One is 10% missing (type A), the other is 20% missing (type B). We select a subsample at random from the sample and make them nonresponse. Nonresponse in the first wave is deleted from the sample. In the second wave, only 95% of the respondents in the wave 1 give response and the other is missing. However, nonresponse in the second wave is not deleted. 95% of the units which belong to xx response pattern at the first two waves give response in the wave 3, and 5% of the units of xo response pattern also give response. We employ a weighting adjustment method to compensate for the nonrespondents in the wave 1, and use a hot deck imputation method for the nonrespondents in the wave 2 and 3. Table 4.1 shows the simulated response patterns. Since only sample data are available, it is impossible to calculate the

TABLE 4.1 *The simulated response patterns*

Response Pattern	*missing type A*	*missing type B*
xxx	829 (80.9%)	731 (71.3%)
xxo	45 (4.4%)	34 (3.3%)
xoo	4 (0.4%)	2 (0.2%)
xox	43 (4.2%)	53 (5.2%)
ooo	104 (10.1%)	205 (20.0%)
	1,025 (100%)	1,025 (100%)

bias and the MSE of each estimator. Hence, we try to compare the estimates suggested in Section 2 with the sample mean calculated with complete response data. The values of the mean estimates and their corresponding standard errors are calculated in Table 4.1. We get three different estimates for each missing type A and B, respectively: The first one is the sample mean of responded data only, the second one is the estimate after imputation under single imputation class, the last one is the estimate after imputation under two imputation classes.

Table 4.2 shows that there are no significant differences among several mean estimates. Since random nonresponse mechanism is assumed and all the estimators used is unbiased, that result seems natural. If nonresponse is nonignorable, we may divide the sample into many imputation classes by use of some auxiliary information. Even in that case random nonresponse mechanism within an imputation class may be employed, so the results in this simulation study make senses. The values in the parentheses are standard error of each estimator which is calculated by use of (3.3), (3.6) in Section 3. To examine the relative efficiency

TABLE 4.2 *Cross-sectional mean estimates and standard errors*

missing type		*1st wave*	*2nd wave*	*3rd wave*
completely responded	mean estimate	3899041.0 (98426.2)	4524057.1 (160554.5)	4114893.2 (134019.9)
missing type A	estimate from responded data	3816923.1 (104687.6)	4435878.2 (179956.6)	3978215.7 (149507.3)
	estimate from one imp'n class	3827807.4 (102574.4)	4426959.0 (161241.1)	3977133.7 (146114.2)
	estimate from two imp'n class	3897913.6 (101378.9)	452064.3 (160557.2)	4084361.3 (142061.1)
missing type B	estimate from responded data	387437.2 (108594.4)	4401375.2 (184792.3)	3860809.6 (155285.5)
	estimate from one imp'n class	3891373.4 (110031.4)	4418441.3 (185898.2)	399262.8 (154971.2)
	estimate from two imp'n class	3900137.5 (106549.5)	4476338.2 (178294.6)	3976372.5 (148055.6)

of each estimator, we compare the variance estimates under each missing type with that of complete response data. Table 4.3 presents the relative efficiencies of several estimates. If missing data are deleted, it causes decrease of sample size. Looking at the first row in each missing type, some inefficiency is caused due to the decrease of sample size. After using imputation method under one imputation class to compensate for missing data instead of deleting, the gain in efficiency is only a little under missing type A and is almost the same under missing type B. However, in case of using two imputation class, the gain increases a little. Table 4.4 contains the results of longitudinal analysis. Unfortunately, the estimates of correlation coefficients between different waves lie in (0.00, 0.30). Hence the coefficients of variation for the mean change estimators are too large. In spite of that all, we can note that the standard errors calculated by (3.9) and (3.13) are smaller than those calculated under the assumption that the sample in a wave is independent of other samples in different waves. This result comes from the fact that the data from different waves are positively correlated.

TABLE 4.3 *The comparison of relative efficiency*

missing type		*1st wave*	*2nd wave*	*3rd wave*
A	est. from responded data	1.13	1.25	1.25
	est. from one imp'n class	1.08	1.17	1.18
	est. from two imp'n class	1.06	1.00	1.12
B	est. from responded data	1.20	1.32	1.36
	est. from one imp'n class	1.39	1.34	1.35
	est. from two imp'n class	1.17	1.21	1.20

TABLE 4.4 *Standard errors of mean change estimates between different waves*

missing type	*difference*	*s.e. of the suggested est.*	*s.e. of the indep. sample*
A	$\bar{Y}_{(2)} - \bar{Y}_{(1)}$	182366.2	189418.4
	$\bar{Y}_{(3)} - \bar{Y}_{(1)}$	178805.9	190981.8
	$\bar{Y}_{(3)} - \bar{Y}_{(2)}$	224888.5	226517.9
B	$\bar{Y}_{(2)} - \bar{Y}_{(1)}$	202551.7	207477.7
	$\bar{Y}_{(3)} - \bar{Y}_{(1)}$	174176.5	186555.5
	$\bar{Y}_{(3)} - \bar{Y}_{(2)}$	226623.4	231595.1

5. Concluding Remarks

It is a very complicated problem to treat wave nonresponse in panel surveys. To compensate for wave nonresponse, both weighting adjustment approach and imputation approach can be considered. Since each method has in some sense complementary strengths and drawbacks, it is natural to consider combinations of the two approaches. Therefore, we suggested a new combined method, in which we can get simplicity by deleting less frequent nonresponse patterns. The results of our simulation study indicate that the proposed method is quite simple to use and may give a reasonable performance.

REFERENCES

Folsom, R. E. and Witt, M. B. (1994). "Testing a new attrition nonresponse adjustment method for SIPP", *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 428–433.

Kalton, G. and Miller, M. (1986). "Effects of adjustments for wave nonresponse on panel survey estimates", *Proceedings of the Section on Survey Research*

Methods, American Statistical Association, 194–199.

Lepkowski, J. M. (1989). "Treatment of wave nonresponse in panel surveys", In *Panel Surveys*, John Wiley & Sons, 348–374.

Rao, J. N. K. and Shao, J. (1992). "Jackknife variance estimation under hot deck imputation", *Biometrika*, **79**, 811–822.

Rao, J. N. K. and Shao, J. (1996). "On balanced half-sample variance estimation in stratified sampling", *Journal of the American Statistical Association*, **91**, 343–348.

Rizzo, L., Kalton, G., Brick, M. and Petroni, R. (1994). "Adjusting for panel nonresponse in the survey of income and program participation", *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 422–427.

Shao, J., Chen, Y. and Chen, Y. (1998). "Balanced repeated replications for stratified multistage survey data under imputation", *Journal of the American Statistical Association*, **93**, 819–831.

Shao, J. and Sitter, R. R. (1996). "Bootstrap for imputed survey data", *Journal of the American Statistical Association*, **91**, 1278–1288.