*Jour. Ind. Soc. Ag. Statistics*
*51(2&3), 1998 : 331-342*

# Regression Adjustments for Nonresponse

Wayne A. Fuller and Anthony B. An[1]
*Iowa State University, Ames, IA 50010*

SUMMARY

Methods of using characteristics of nonrespondents and auxiliary data to make adjustments for nonresponse are investigated. Properties of ordinary regression estimators, and of regression estimators with initial weights based on estimated response probabilities are presented. The methods are applied to data from the United States Survey of Income and Program Participation.

*Keywords*: Sample surveys, Regression estimation, Missing data.

## *1. Introduction*

Essentially every survey conducted with respondents has some nonresponse. Therefore, there is an extensive literature on methods adjusting for nonresponse. References devoted to the general area include Madow *et al* [11], Kalton [7], Little and Rubin [10], and Lessler and Kalsbeek ([9], Ch. 7). There is a close relationship between samples with nonresponse and two-phase samples. See Sarndal and Swensson [13] and Kott [8]. We shall exploit this relationship and the theory of linear estimation in developing estimators for longitudinal studies with nonresponse.

## *2. Regression Estimation*

A general description of regression estimation is given in texts such as Cochran [1] and Sukhatme, *et al* [14]. Our development for nonresponse follows Fuller, Loughin, and Baker [14], and Isaki and Fuller [5]. Assume that a sample containing n units has been selected and that the probability of selecting unit i is $\pi_i$. Assume that a k-dimensional vector of population means, denoted by $\overline{\mathbf{X}} = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k)$ is known, that the vector $(y_i, x_{i1}, x_{i2}, \ldots, x_{ik})$ is observed for every unit in the sample and that an estimator of the mean of y is desired. The vector $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$ is sometimes called the vector of control variables. We assume that the first element of $\mathbf{x}_i$ is one for all i. Hence, the first element of $\overline{\mathbf{X}}$ is also one. A regression estimator of the mean of y is

1 SAS Institute, SAS Campus Drive, Cary, NC, 27513

$$\bar{y}_r = \bar{X}\hat{\beta} \tag{2.1}$$

where

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i' \pi_i^{-1} x_i\right) \sum_{i=1}^{n} x_i' \pi_i^{-1} y_i \tag{2.2}$$

and we have assumed $\sum x_i' \pi_i^{-1} x_i$ to be nonsingular. The estimator (2.1) can also be written as

$$\bar{y}_r = \sum_{i=1}^{n} w_i y_i \tag{2.3}$$

where

$$w_i = \bar{X}\left(\sum_{i=1}^{n} x_i' \pi_i^{-1} x_i\right)^{-1} x_i' \pi_i^{-1} \tag{2.4}$$

and the weights have the property

$$\sum_{i=1}^{n} w_i x_i = \bar{X} \tag{2.5}$$

Fuller [3] gave theory that covers the regression estimator of the mean and of the total. To construct an estimator of the variance of the regression estimator for a stratified two stage sample, replace the single subscript i with the triple $l$ jt. Then, omitting the finite correction term, a variance estimator is

$$\hat{V}\{\bar{y}_r\} = (n-k)^{-1} \sum_{l=1}^{L} (n_l - 1)^{-1} \sum_{j=1}^{n_l} (d_{lj.} - d_{l..})^2 \tag{2.6}$$

where

$$d_{lj.} = \sum_{t=1}^{m_{lj}} W_{ljt} (y_{ljt} - x_{ljt}\hat{\beta})$$

$$d_{l.} = n_l^{-1} \sum_{j=1}^{n_l} d_{lj.}$$

$n_l$ is the number of sample primary sampling units in stratum $l$, $m_{lj}$ is the number of sample elements in primary sampling unit j of stratum $l$, $\hat{\beta}$ is the vector of coefficients defined in (2.2), n is the total number of elements in the sample, and $w_{ljt}$ is the weight for element t in primary sampling unit j of stratum $l$.

The typical theoretical development for regression estimation in surveys, and that above, assumes the sample to be a probability sample from the population with known selection probabilities. However, given auxiliary information, regression estimation provides a method of reducing nonresponse bias for situations in which the response probabilities are unknown. The degree to which the bias is reduced depends upon the relationship between the control variables, the variables of interest, and the response probabilities.

To investigate estimation in the presence of nonresponse, let $p_i$ be the conditional probability of observing the unit i given that the unit is selected. Then the expectations of weighted sums of squares and products used in constructing the regression estimator are functions of the $p_i$. That is,

$$E\left\{\sum_{i=1}^{n} x_i' \pi_i^{-1} x_i \delta_i - \sum_{i=1}^{N} x_i' p_i x_i\right\} = 0 \tag{2.7}$$

and

$$E\left\{\sum_{i=1}^{n} x_i' \pi_i^{-1} y_i \delta_i - \sum_{i=1}^{N} x_i' p_i y_i\right\} = 0 \tag{2.8}$$

where

$$\begin{aligned} \delta_i &= 1 \text{ if element responds} \\ &= 0 \text{ otherwise} \end{aligned}$$

Therefore, under conditions such as those used by Fuller [3],

$$\operatorname*{p\,lim}_{n \to \infty} (\tilde{\beta} - \gamma) = 0$$

where

$$\gamma = \left(\sum_{i=1}^{N} x_i' p_i x_i\right)^{-1} \sum_{i=1}^{N} x_i' p_i y_i \tag{2.9}$$

and $\tilde{\beta}$ is the estimator of (2.2) constructed with the sums of squares and products of respondents defined in (2.7) and (2.8),

$$\tilde{\beta} = \left(\sum_{i=1}^{n} x_i' \pi_i^{-1} x_i \delta_i\right)^{-1} \sum_{i=1}^{n} x_i' \pi_i^{-1} y_i \delta_i \tag{2.10}$$

The population mean of Y can be expressed as

$$\bar{Y} = \bar{X}\gamma + \bar{A} \tag{2.11}$$

where $a_i = y_i - x_i \gamma$ and $\bar{A}$ is the population mean of the $a_i$. Thus the regression estimator

$$\bar{y}_{rr} = \bar{X}\tilde{\beta} \tag{2.12}$$

constructed with the coefficient vector (2.10) will be a consistent estimator of $\bar{Y}$ if $p \lim_{N \to \infty} \bar{A} = 0$. There are several ways for this to occur. The probability limit of $\bar{A}$ will be zero if the finite population is a random sample from an infinite population in which

$$y_i = x_i \beta + e_i$$

the $e_i$ are independent with bounded variance, and $E\{e_i \mid x_i\} = 0$.

The mean $\bar{A}$ is zero when $p_i$ is a constant for all i and the first element of $x_i$ is one for all i, because then

$$\gamma = \beta = \left(\sum_{i=1}^{N} x_i' x_i\right)^{-1} \sum_{i=1}^{N} x_i' y_i \tag{2.13}$$

and $\sum_{i=1}^{N} (y_i - x_i \beta) = 0$. A sufficient condition for $\bar{A}$ to be zero is the existence of a row vector c such that

$$c x_i' = p_i^{-1} \tag{2.14}$$

for $i = 1, 2, \ldots, N$. See Zyskind [16]. Thus, if the reciprocal of the response probability is a linear function of the control variables, the regression estimator is a consistent estimator of the mean of y. One way in which (2.14) can be satisfied is for the elements of $x_i$ to be dummy variables that define subgroups and for the response probabilities to be constant in each subgroup. This situation is sometimes described by saying that elements are missing at random in each subgroup.

Under reasonable assumptions, such as discussed in Fuller [3], the error in $\tilde{\beta}$ as an estimator of $\gamma$ can be approximated by

$$\tilde{\beta} - \gamma = G^{-1} T^{-1} \sum_{i=1}^{n} x_i' \pi_i^{-1} a_i$$

where $a_i$ is defined in (2.11)

$$T = \sum_{i=1}^{N} p_i \quad \text{and} \quad G = T^{-1} \sum_{i=1}^{N} x_i' p_i x_i$$

Because

$$(\hat{T}, \hat{G}) = \left( \sum_{i=1}^{n} \pi_i^{-1} \delta_i, \hat{T}^{-1} \sum_{i=1}^{n} x_i' \pi_i^{-1} x_i \delta_i \right)$$

are consistent estimators of T and G, the variance of the regression estimator can be estimated by using the estimated variance of $\sum_{i=1}^{n} x_i' \pi_i^{-1} a_i$. If we assume that the conditional probabilities of response in one primary sampling unit are independent of those in all other primary sampling units and that at least one observation unit is observed in each selected primary sampling unit, then (2.6) remains an appropriate estimator of the variance of the regression estimated mean of y, in the presence of nonresponse.

## 3. *Response Probability and Regression Estimation*

The procedure of making a first adjustment to the selection probabilities for nonresponse is very common in survey sampling. The most frequently used procedure is to form adjustment cells and to ratio adjust the weights of respondents in the cell so that the sum of the weights is equal to the estimated (or known) total for the cell. See, for example, Little and Rubin ([10], p. 250). Frequently, this adjustment is followed by another estimation scheme such as ratio estimation or regression estimation. A modification of the procedure using an estimated response probability function is discussed by Folsom and Witt [2]. We consider the theory for such procedures in this section.

We assume that the inverse of the response probability for individual i is given by

$$p_i^{-1} = g(z_i; \theta^0) \tag{3.1}$$

where $\mathbf{z}_i$ is a vector of variables that can be observed for both respondents and nonrespondents, $\theta^0$ is the true value of $\theta$, $g(\mathbf{z}_i;\theta)$ is continuous in $\theta$ with continuous first and second derivatives in an open set containing $\theta^0$ for all $\mathbf{z}_i$. We also assume that $p_i$ is bounded below by a positive number. We assume a finite population of size N that is a sample from an infinite superpopulation. Let a sample of size n be selected from finite population. We begin the discussion under the assumption of a simple random nonreplacement sample. The response mechanism with probabilities $p_i$ produces a smaller sample of size m on which the vector $(Y, \mathbf{x}_i, \mathbf{z}_i)$ is observed. There may be, and usually will be, elements that appear in both $\mathbf{x}$ and $\mathbf{z}$.

Let $\delta_i$ be the indicator variable previously defined with $\delta_i = 1$ if a response is obtained and $\delta_i = 0$ if a response is not obtained. Using the vector $(\delta_i, \mathbf{z}_i)$, the parameter $\theta^0$ of the response probability function is estimated. Assume that $\hat{\theta} - \theta^0 = O_p(n^{-1/2})$ where $\hat{\theta}$ is the estimator of $\theta$. Let $\gamma = \beta$ defined in (2.13) denote the finite population regression vector and let $a_i = y_i - \mathbf{x}_i\gamma$, as defined in (2.11). We assume

$$\sum_{i=1}^{N} a_i = 0 \tag{3.2}$$

The sum of the $a_i$ is zero by construction when the vector $\mathbf{x}_i$ contains an element that is identically equal to one. Let

$$\hat{\gamma} = \left(\sum_{i=1}^{m} \mathbf{x}_i' \mathbf{x}_i \pi_i^{-1} p_i^{-1}\right)^{-1} \sum_{i=1}^{m} \mathbf{x}_i' y_i \pi_i^{-1} p_i^{-1} \tag{3.3}$$

$$\tilde{\gamma} = \left(\sum_{i=1}^{m} \mathbf{x}_i' \mathbf{x}_i \pi_i^{-1} \hat{p}_i^{-1}\right)^{-1} \sum_{i=1}^{m} \mathbf{x}_i' y_i \pi_i^{-1} \hat{p}_i^{-1} \tag{3.4}$$

where $\pi_i$ are the selection probabilities used to select the sample of n from N and $\hat{p}_i^{-1} = g(\mathbf{z}_i; \hat{\theta})$. The summations in (3.3) and (3.4) are over the m elements for which $\delta_i = 1$. We assume a sequence of finite populations and samples such that

$$\hat{\gamma} - \gamma = O_p(n^{-1/2}) \tag{3.5}$$

$$\mathbf{A}_n^{-1/2}(\hat{\gamma}-\gamma) \xrightarrow{L} N(\mathbf{0},\mathbf{I}) \tag{3.6}$$

where

$$\mathbf{A}_n = \mathbf{B}_{xx}^{-1}\hat{V}\left\{\sum_{i=1}^{m}\mathbf{x}_i' a_i \pi_i^{-1} p_i^{-1}\right\}\mathbf{B}_{xx}^{-1} \tag{3.7}$$

$$\mathbf{B}_{xx} = \sum_{i=1}^{m}\mathbf{x}_i'\mathbf{x}_i \pi_i^{-1} p_i^{-1}$$

and $\hat{V}\left\{\sum_{i=1}^{m}\mathbf{x}_i' a_i \pi_i^{-1} p_i^{-1}\right\}$ is the estimated variance computed by the appropriate sampling formula. Sufficient conditions for (3.5) and (3.6) are given in Fuller [3] and Isaki and Fuller [5].

Now,

$$\begin{aligned}\hat{p}_i^{-1} - p_i^{-1} &= \frac{\partial g(\mathbf{z}_i;\theta^*)}{\partial\theta'}(\hat{\theta}-\theta^0)\\ &= \frac{\partial g(\mathbf{z}_i;\theta^0)}{\partial\theta'}(\hat{\theta}-\theta^0) + 0.5(\hat{\theta}-\theta^0)'\frac{\partial g(\mathbf{z}_i;\breve{\theta})}{\partial\theta\,\partial\theta'}(\hat{\theta}-\theta^0)\end{aligned} \tag{3.8}$$

where $\theta^*$ and $\breve{\theta}$ are on the line segment joining $\hat{\theta}$ and $\theta^0$. Let $\mathbf{h}_i$ denote the row vector of first derivatives of $g(\mathbf{z}_i;\theta)$ evaluated at $\theta = \theta^0$ and let $\mathbf{B}(\mathbf{z}_i;\breve{\theta})$ denote the matrix of second derivatives of $g(\mathbf{z}_i;\theta)$ evaluated at $\theta = \breve{\theta}$. Then we can write

$$\hat{p}_i^{-1} - p_i^{-1} = \mathbf{h}_i(\hat{\theta}-\theta^0) + 0.5(\hat{\theta}-\theta^0)'\mathbf{B}(\mathbf{z}_i;\breve{\theta})(\hat{\theta}-\theta^0) \tag{3.9}$$

We require the sample design to be such that sample moments converge to population moments. Therefore, we assume

$$N^{-1}\sum_{i=1}^{m}\pi_i^{-1}p_i^{-1}\mathbf{W}_i'\mathbf{W}_i - N^{-1}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{W}_i = O_p(n^{-1/2}) \tag{3.10}$$

where we assume one is an element of $\mathbf{W}_i$,

$$\mathbf{W}_i = (y_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{h}_i, (\text{vec}\,\mathbf{x}_i' a_i p_i \mathbf{h}_i)', (\text{vec}\,\mathbf{x}_i' p_i \mathbf{h}_i)')$$

and vec **A** is the column vector composed of the columns of the matrix **A**. Some of the moment assumptions can be weakened. We also assume that the sample mean squares of the vector of second derivatives, vec $\mathbf{B}(\mathbf{z}_i; \theta^0)$, converge to the corresponding population moments. We have

$$\tilde{\gamma} - \gamma = \mathbf{B}_{xx}^{-1} \sum_{i=1}^{m} \pi_i^{-1} p_i^{-1} \mathbf{x}_i' a_i [1 + p_i \mathbf{h}_i (\hat{\theta} - \theta^0)] + O_p(n^{-1}) \tag{3.11}$$

Under assumption (3.10), the matrix

$$N^{-1} \sum_{i=1}^{m} \pi_i^{-1} \mathbf{x}_i' a_i \mathbf{h}_i - N^{-1} \sum_{i=1}^{N} p_i \mathbf{x}_i' \mathbf{h}_i a_i = O_p(n^{-1/2}) \tag{3.12}$$

Therefore, $\tilde{\gamma} - \gamma = O_p(n^{-1/2})$ and

$$N^{-1} \sum_{i=1}^{m} Y_i - \hat{\mu}_y = O_p(n^{-1/2}) \tag{3.13}$$

If we assume

$$C_h = N^{-1} \sum_{i=1}^{N} p_i \mathbf{x}_i' \mathbf{h}_i a_i = O_p(1) \tag{3.14}$$

then

$$\tilde{\gamma} - \gamma = \left( \sum_{i=1}^{m} \mathbf{x}_i' \mathbf{x}_i \pi_i^{-1} p_i^{-1} \right) \sum_{i=1}^{m} \mathbf{x}_i' a_i \pi_i^{-1} p_i^{-1} + O_p(n^{-1/2}) \tag{3.15}$$

The matrix $C_h$ is $O_p(N^{-1/2})$ if the finite population is a sample from a superpopulation in which $a_i$ is independent of $(\mathbf{x}_i, z_i, \pi_i, p_i)$.

## 4. *Application to SIPP*

Our study was motivated by an investigation of nonresponse for the Survey of Income and Program Participation (SIPP) conducted by the U.S. Census Bureau. See Jabine, King and Petroni [6] for a description of SIPP. The SIPP, is a multistage stratified (72 strata) cluster systematic sample of the noninstitutionalized resident population of the United States, where the cluster is a household. The sample is the sum of four equal sized rotation groups. During each month of the study, one rotation group was interviewed. One cycle of four interviews for the four groups is called a wave. Several waves which cover a period of time are called a panel. For example, Panel 1987, composed of seven waves, contains the SIPP-interviewed people from February 1987, through May 1989. The survey produces two kinds of estimates: cross-sectional and longitudinal. In order to be a part of the longitudinal sample, the respondent must provide data at each of seven interview periods. About 79% of those that responded at the first interview (Wave One) of Panel 1987 also responded at the remaining six interviews. A total of 39,766 people interviewed in Wave One were eligible for the 1987 panel longitudinal sample. A total of 24,429 individuals completed all seven interviews. Estimation for the longitudinal sample uses information from all Wave One respondents and also uses control information from the Current Population Survey. We compare alternative estimators that use the information in different ways.

We treat the Panel 1987 SIPP data as a three-phase sample, where the phase I sample is the Current Population Survey. In the analysis, we assume zero error in the estimates of the phase I sample. The phase II sample is the 1987 Wave One data. The phase II included all the people who were eligible and participated in the survey during Wave One. The phase III sample is defined as a subsample from the phase II which includes all people who participated in the survey from Wave One through Wave Seven unless they died or moved to an ineligible address. The phase III sample is also called the longitudinal sample of panel 1987.

The current estimation scheme makes an initial adjustment in the panel weights, equivalent to an initial estimate of response probabilities, based on 80 adjustment cells. See Waite [15] and Petroni, Singh, and Kaspryzk [12]. The adjustment cells are formed using variables such as income, race, education, type of income, type of assets, labor force status and employment status observed at the first interview. The second stage adjustment is a ranking procedure based on 97 variables with estimated means taken from the Current Population Survey. These 97 variables are based on gender, age, race, family type and household type.

To extend the estimator of the response probabilities beyond that based on adjustment cells, let $\hat{r}_{ijk}$ denote the estimated response probabilities based on the adjustment cells for individual k in cluster j of stratum i. The $\hat{r}_{ijk}$ are constant in a cell and are the ratio of weighted respondents to the total sample in a cell where the weights are the inverses of the selection probabilities. Let $\delta_{ijk}$ denote an indicator variable equal to one if the individual responds on all seven periods and zero otherwise. To reduce the number of variables to be included in the nonlinear estimation of a logistic model, we first computed the linear regression of $\delta_{ijk}$ on $\left(x_{ijk}, q_{ijk}\right)$ where $x_{ijk}$ is the vector of second phase adjustment variable and $q_{ijk}$ is the vector of indicator variables for the adjustment cells. Let $\hat{\delta}_{ijk}$ be the predicted values from this regression and let $\bar{r}$ be the mean response rate. Then $\theta = \left(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4\right)$ of the logistic model

$$\left(1 - p_{ijk}\right)^{-1} = 1 + \exp\left\{\theta_0 + \theta_1 \log\left[\hat{r}_{ijk}(1-\hat{r}_{ijk})^{-1}\right] + \theta_2(\hat{\delta}_{ijk} - \hat{r}_{ijk}) + \theta_3(\hat{\delta}_{ijk} - \hat{r}_{ijk})^2 + \theta_4(\hat{r}_{ijk} - \bar{r})(\hat{\delta}_{ijk} - \hat{r}_{ijk})\right\}$$

was estimated. The estimated vector is

$$\hat{\theta} = \begin{matrix}(0.035, & 1.032, & 6.158, & -5.280, & 6.576) \\ (0.052) & (0.036) & (0.316) & (1.794) & (2.506)\end{matrix}$$

If only the variable $\hat{r}_{ijk}(1-\hat{r}_{ijk})^{-1}$ is included in the model, $\hat{\theta}_1 = 1.000$. Recall that $\hat{r}_{ijk}$ is based on 80 cells and $\hat{\delta}_{ijk} - \hat{r}_{ijk}$ reflects the effect of 97 variables. Even after adjusting for the degrees of freedom hidden in the variables, the fit indicates that the adjustment cells model for response probability was improved by addition of the three parameters $\theta_2$, $\theta_3$, and $\theta_4$ to the model.

By the results of Section 2, if $p_{ijk}^{-1}$ is well approximated by a linear function of X, then the regression estimator is approximately unbiased. Therefore, improving estimates of the response probabilities does not necessarily lead to an improvement in the final estimator.

Table 1 contains estimated standard errors of three three-phase procedures expressed relative to the estimated standard error of the Census procedure. The

**Table 1.** Estimated standard errors for alternative estimation procedures divided by estimated standard error of Census Procedures

| Variable | Procedure | | |
|---|---|---|---|
| | 3-Phase Element[1] | 3-Phase Cluster[1] | 3-Phase Cluster[2] $\hat{p}_{ijk}$ |
| Jan 87 Personal Income | 0.99 | 0.96 | 0.95 |
| Jan 89 Personal Income | 0.99 | 0.99 | 0.98 |
| Jan 87 Household Income (10's) | 1.01 | 0.97 | 0.97 |
| Jan 89 Household Income (10's) | 1.01 | 0.98 | 0.98 |
| Jan 87 Labor Force (%) | 0.98 | 0.94 | 0.96 |
| Jan 89 Labor Force (%) | 0.99 | 0.98 | 1.01 |

[1]Regressions computed using selection probability weights

[2]Regressions computed using estimated response probability weights

procedure called "3-phase element" is a regession estimation procedure that uses the 97 phase I (CPS) variables and the 79 phase II variables in a three-phase estimator. The regressions at each stage are computed using sums of squares and products based on individuals. The weights used in the regression computations are the initial sampling weights. The "3-phase cluster" procedure is the same estimator with regression coefficients based on cluster totals. The procedure "3-phase cluster, $\hat{p}_{ijk}$" uses the estimated weights from the logistic function in the calculation of the regression coefficients. There are modest differences in the standard errors. It is not surprising that some estimated standard errors are larger for the procedure that uses estimated response probabilities. If the response probabilities vary, this produces a wider range of weights which can increase the variance.

It is felt that the primary effect of using estimated probabilities will be on the bias of the estimators. In the case of SIPP, few significant difference in the final estimators were observed. One exception was labor force. The estimator using response probabilities produced estimates of the fraction of individuals in the labor force that were larger than those using the initial sampling weights.

ACKNOWLEDGEMENT

This research was partly supported by Cooperative Agreement 43- 3AEU-3-80088 with the National Agricultural Statistics Service and the U.S. Bureau of the Census.

REFERENCES

[1] Cochran, W.G., 1977. *Sampling Techniques* ($3^{rd}$ ed.). John Wiley and Sons, New York.

[2] Folsom, R.E. and Witt, M.B., 1944. Testing a new attrition nonresponse adjustment method for SIPP. *Technical report. Research Triangle Institute, Research Triangle Park, North Carolina.*

[3] Fuller, W.A., 1975. Regression analysis for sample survey. *Sankhya,* **C37,** 117-132.

[4] Fuller, W.A., Loughin, M.M., and Baker, H.D., 1994. Regression weighting for the 1987-88 National Food Consumption Survey. *Survey Methodology,* **20** 75-85.

[5] Isaki, C.T. and Fuller, W.A., 1982. Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.,* **77,** 89-96.

[6] Jabine, T., King K., and Petroni, 1990. *Survey of Income and Program Participation: Quality Profile*. Bureau of the Census, Washington, D.C., U.S. Department of Commerce.

[7] Kalton, G., 1983. *Compensating for Missing Survey Data*. Institute for Social Research, The University of Michigan, Ann Arbor, Michigan.

[8] Kott, P.S., 1994. A note on handling nonresponse in sample surveys. *J. Amer. Statist. Assoc.*, **89**, 693-696.

[9] Lessler, J.T. and Kalsbeek, W.D., 1992. *Nonsampling Error in Surveys*. John Wiley and Sons, New York.

[10] Little, R.J.A. and Rubin, D.B., 1987. *Statistical Analysis with Missing Data*. John Wiley and Sons, New York.

[11] Madow, W.G., Nisselson, N., and Olkin, I. (eds.), 1983. *Incomplete Data in Sample Surveys*. Academic Press, New York.

[12] Petroni, R.J., Singh, R.P., and Kasprzyk, D., 1992. Longitudinal weighting issues and associated research for the SIPP. *Proceedings of the Survey Research Methodology Section, American Statistical Association*, 548-553.

[13] Sarndal, C.E. and Swensson, B., 1987. A general view of estimation for two phases of selection with application to two-phase sampling and nonresponse. *International Statistical Review*, **55**, 279-294.

[14] Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S., and Asok, C., 1984. *Sampling Theory of Surveys with Applications*. Iowa State University Press, Ames, Iowa.

[15] Waite, P.J., 1990. SIPP 1987: Specifications for panel file longitudinal weighting of persons. *Internal Census Bureau memorandum from Waite to Courtland*. June 1, 1990.

[16] Zyskind, G., 1967. On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.*, **38**, 1092-1109.