REGRESSION ADJUSTMENT FOR NONRESPONSE

Anthony B. An, SAS Institute and Wayne A. Fuller, Iowa State University
Anthony B. An, SAS Campus Drive, Cary, NC 27513

KEY WORDS: Nonresponse, SIPP, Regression, Consistency

1 INTRODUCTION

The Census Bureau designed the Survey of Income and Program Participation (SIPP) to provide improved information on participation in government programs. Characteristics of persons and households which may have impact on income and program participation are collected in the SIPP surveys.

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. Each month 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 30,766 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.

Longitudinal estimators are derived from the weights assigned to the people in the longitudinal sample. Many weighting procedures have been investigated for the longitudinal sample. The current weighting scheme at the U.S. Census Bureau is described by Waite (1990). The procedure makes two adjustments to the base weights, where the base weights are the reciprocals of the probabilities of selection. The adjustments attempt to compensate for nonresponse and undercoverage, using variables thought to be highly correlated with SIPP variables of interest. The first stage adjustment is of the post stratification type. The cells are defined by characteristics of people who were eligible in the Wave One sample. The second stage adjustment is a raking procedure, performed after the first adjustment, using data form the Current Population Survey as controls.

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.

We use Poisson Sampling to model the response behavior by assuming that the sample units in the phase III sample are selected with "response probabilities" and that response is independent from person to person. It can be shown that under mild conditions, incorporating the response probabilities into the regression will yield consistent estimators. We describe a procedure to estimate the response probabilities when they are unknown. We will compare the three-phase regression estimators using different sets of weights in the regression. One set of weights is the sampling weights. The second set of weights is the sampling weights adjusted by the estimated response probabilities. Estimated standard errors of the estimators using these two sets of weights are also compared.

2 NONRESPONSE AND POISSON SAMPLING

Given a selected sample, one model for response behavior is the Poisson sampling mechanism. Poisson sampling is a sampling procedure in which sample units are selected by independent Bernoulli trials. That is, if element i is selected in the sample, then element i will respond if a Bernoulli trial has a success outcome, with a success probability p_i. We call p_i the response probability. Poisson sampling is a rather restrictive model because it assumes the probability that element i responds, does not depend on the probability that element j responds.

Assume the finite population ξ_N contains N_c primary sampling units, called clusters, where the i-th cluster contains m_i elements. A probability sample s, which contains n_c clusters is selected from the finite population ξ_N. We use r_{ij} to indicate the response of the j-th element in the i-th cluster, when cluster i is selected,

$$r_{ij} = \begin{cases} 1 & \text{if element } j \text{ of cluster i responds} \\ & \text{when cluster } i \text{ is selected} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Using the Poisson sampling model, we have

$$p_{ij} = \Pr\left(r_{ij} = 1 | \text{cluster } i \text{ is selected}\right), \tag{2}$$

and

$$E\left(r_{ij} r_{i'j'} = 1\right) = \begin{cases} p_{ij} & \text{if } i = i',\ j = j' \\ p_{ij} p_{i'j'} & \text{otherwise.} \end{cases} \tag{3}$$

3 REGRESSION ESTIMATOR WITH NONRESPONSE ADJUSTMENT

Given a sample with nonresponse, estimating the population mean without adjusting for nonresponse, will introduce bias if the respondents are different from the nonrespondents. Using response probabilities to adjust the regression estimators is one way to reduce the bias due to the nonresponse.

Consider a regression estimator of the mean of a variable Y,

$$\hat{\mu}_{\text{Reg}} = \bar{\mathbf{X}}\hat{\gamma}, \tag{4}$$

where

$$\hat{\gamma} = \left(\sum_{i,j} \mathbf{x}'_{ij} \pi_i^{-1} r_{ij} \mathbf{x}_{ij}\right)^{-1} \left(\sum_{i,j} \mathbf{x}'_{ij} \pi_i^{-1} r_{ij} y_{ij}\right), \tag{5}$$

and $\mathbf{x}_{ij}$ is the vector of auxiliary variables, π_i is the inclusion probability for cluster i, r_{ij} is defined by (1), and $\bar{\mathbf{X}}$ is the population mean. We assume $\Sigma \mathbf{x}'_{ij} \pi_i^{-1} r_{ij} \mathbf{x}_{ij}$ to be nonsingular. In the case of full response, the regression vector $\hat{\gamma}$ in (5) is a consistent estimator of

$$\gamma = \left(\sum_{i=1}^{N_c} \sum_{j=1}^{m_i} \mathbf{x}'_{ij} \mathbf{x}_{ij}\right)^{-1} \left(\sum_{i=1}^{N_c} \sum_{j=1}^{m_i} \mathbf{x}'_{ij} y_{ij}\right), \tag{6}$$

under mild conditions. See Fuller (1975). If $\hat{\gamma}$ is consistent for γ, a sufficient condition for $\hat{\mu}_{\text{Reg}}$ in (4) to be consistent for the true population mean is that

$$\bar{A} = \bar{Y} - \bar{\mathbf{X}}\gamma = 0. \tag{7}$$

However, in the presence of nonresponse, $\hat{\gamma}$ need not converge to γ. Let $\overset{**}{\gamma}$ be an estimator and assume

$$plim\ \overset{**}{A} = plim\left(\bar{Y} - \bar{\mathbf{X}}\ \overset{**}{\gamma}\right) = 0. \tag{8}$$

Then a consistent estimator of the mean of Y is

$$\overset{**}{\mu} = \bar{\mathbf{X}}\ \overset{**}{\gamma}. \tag{9}$$

One estimator $\overset{**}{\gamma}$ is obtained by including the response probabilities in the weighted regression. This can be done by constructing regression weights using response probabilities, if we know the response probabilities. For example, let

$$\overset{**}{\gamma} = \left(\sum_{i,j} \mathbf{x}'_{ij} \omega_{ij} r_{ij} \mathbf{x}_{ij}\right)^{-1} \left(\sum_{i,j} \mathbf{x}'_{ij} \omega_{ij} r_{ij} y_{ij}\right), \tag{10}$$

where $\omega_{ij} = \pi_i^{-1} p_{ij}^{-1}$, then the regression estimator in (9) will be consistent for γ. However, in most cases we do not know the response probabilities p_{ij}, so we replace p_{ij} in (10) by their estimated values, $\widehat{p}_{ij}$. An (1995) proved that if estimators $\widehat{p}_{ij}$ are consistent for p_{ij}, then under conditions similar to those in Fuller(1975), the regression estimator in (9) will be a consistent estimator.

4 REGRESSION WEIGHTS FOR THREE-PHASE ESTIMATOR

In this section, we describe the construction of the three phase estimator using different sets of initial weights with and without adjustment by estimated response probabilities.

Let $\mathbf{x}_{ijk}$ be the vector of observations on the x-variables for the k-th individual in the j -th cluster of stratum i , where

$$\mathbf{x}_{ijk} = \left(x_{ijk1},\ x_{ijk2},\ \ldots,\ x_{ijkp}\right), \tag{11}$$

$i = 1, 2, ..., L$ is the stratum identification, $j = 1, ..., n_i$ is the cluster within stratum identification, $k = 1, 2, ..., m_{ij}$ is the individual within cluster identification, and x_{ijkl} is the observation on the l-th variable for individual ijk, where $l = 1, 2, ..., p$. Characteristics in different samples are identified by I, II, or III according to the sampling phase. In sample $\tau = I, II, III$, we define the data matrices

$$\left(\mathbf{X}^{(\tau)}, \mathbf{Y}^{(\tau)}, \mathbf{Z}^{(\tau)}\right) = (\mathbf{x}_{ijk}, \mathbf{y}_{ijk}, \mathbf{z}_{ijk}) \qquad (12)$$

and define the total number of individuals in sample τ by $n^{(\tau)} = \sum_{i=1}^{L} \sum_{j=1}^{n_i^{(\tau)}} m_{ij}^{(\tau)}$, where $n_i^{(\tau)}$ is the number of clusters in stratum i , $m_{ij}^{(\tau)}$ is the number of individuals in cluster j of stratum i.

The X variables are control variables for phase I, the Y variables are control variables for phase II, and the Z variables are the variables of interest. We assume that in the phase I sample, only X variables are observed and that the vector of sample totals of the X variables, denoted by $\mathbf{X}_I$, is available. In the phase II sample, we observe Y and X , and in the phase III sample, we observe X, Y, and Z .

The matrix of initial weights in the phase II sample is denoted by

$$\mathbf{W}^{(II)} = \operatorname{diag}\left(w_{ijk}^{(0,II)}\right) \qquad n^{(II)} \times n^{(II)}. \qquad (13)$$

In the phase II sample of SIPP, the initial weights $w_{ijk}^{(0,II)}$ used in this study are the inverse of inclusion probabilities adjusted for control variables: age, gender and race, such that the weighted sum of X variables, using $w_{ijk}^{(0,II)}$ as weights, will yield the population values for these control variables. Since the phase III sample is drawn from the phase II sample, the initial weights in the phase III sample are obtained by adjusting the initial weights $w_{ijk}^{(0,II)}$ using control variables Y. In the SIPP data, Y variables are indicator variables for the noninterview adjustment cells. These noninterview adjustment cells are formed using auxiliary variables that were believed to be correlated with response. The initial weights in the phase III sample are adjusted within each cell. That is, for each element (i, j, k) in the phase III sample, let ℓ be the cell to which (i, j, k) belongs, and the adjustment ration

$$\lambda_\ell = \frac{\sum_{(i',j',k')\in\ell,(i',j',k')\in \text{phase}II} w_{i'j'k'}^{(0,II)}}{\sum_{(i',j',k')\in\ell,(i',j',k')\in \text{phase}III} w_{i'j'k'}^{(0,II)}},$$

then the initial weight for (i, j, k) in the phase III sample is

$$w_{ijk}^{*(0,III)} = \lambda_\ell w_{ijk}^{(0,II)}. \qquad (14)$$

The sum of the weights $w_{ijk}^{*(0,III)}$ equals the sum of $w_{ijk}^{(0,II)}$ within each noninterview adjustment cell ℓ,

$$\sum_{(i,j,k)\in\ell} w_{ijk}^{*(0,III)} = \sum_{(i,j,k)\in\ell} w_{ijk}^{(0,II)}. \qquad (15)$$

A second set of initial weights is also used in our analysis. These weights are the product of the initial weight $w_{ijk}^{*(0,III)}$ and the inverse of the estimated response probability $\hat{p}_{ijk}^{-1}$,

$$w_{ijk}^{(0,III)} = w_{ijk}^{*(0,III)} \hat{p}_{ijk}^{-1}, \qquad (16)$$

where the $\hat{p}_{ijk}$ are estimated from the phase II sample. We give the details of estimating $\hat{p}_{ijk}$ in Section 5. Let

$$\mathbf{W}^{(III)} = \operatorname{diag}\left(w_{ijk}^{(0,III)}\right). \qquad (17)$$

The total number of individuals in the population is denoted by N and the population means of the variables are denoted by μ. A subscript indicating the phase of the sample is applied to estimated totals. For example,

$$\hat{\mathbf{X}}_{II} = \sum_{i=1}^{L} \sum_{j=1}^{n_i^{(II)}} \sum_{k=1}^{m_{ij}^{(II)}} w_{ijk}^{(0,II)} \mathbf{x}_{ijk} \qquad (18)$$

is the estimated total for X computed from the phase II sample using the initial weights $w_{ijk}^{(0,II)}$ where m_{ij} is the number of elements in the ij-th cluster.

We outline the procedure for calculating a three-phase estimator using the weights in (16).

Three-Phase Estimator

(1) Calculate cluster totals using initial weights within each cluster,

$$(\mathbf{x}_{ij.}, \mathbf{y}_{ij.}, \mathbf{z}_{ij.}) = \sum_{k=1}^{m_{ij}} w_{ijk}^{(0,III)} (\mathbf{x}_{ijk}, \mathbf{y}_{ijk}, \mathbf{z}_{ij.}). \quad (19)$$

These cluster totals will be used to perform the regression.

(2) In the phase II sample, estimate the mean of Y variables, by regressing Y on X,

$$\hat{\mu}_Y^{(1)} = \bar{Y}_{II} + \left(\bar{\mathbf{X}}_I - \hat{\bar{\mathbf{X}}}_{II}\right) \hat{\beta}_{Y \cdot X}^{(II)} \qquad (20)$$

where $\hat{\beta}_{Y \cdot X}^{(II)}$ is the estimated regression coefficient matrix.

(3) In the phase III sample, using (20) as the controls, regress Z on X and Y to estimate μ_Z:

$$\hat{\mu}_{\text{3-phase}} = \hat{\bar{\mathbf{Z}}}_{III} + \left[\bar{\mathbf{X}}_I - \hat{\bar{\mathbf{X}}}_{III},\ \hat{\mu}_Y^{(1)} - \hat{\bar{\mathbf{Y}}}_{III}\right] \hat{\beta}_{Z \cdot XY}^{(III)}, \qquad (21)$$

where $\hat{\boldsymbol{\beta}}_{Z\cdot XY}$ is the estimated regression coefficient matrix based on the cluster totals calculated in (19).

The variance of the three-phase estimator can be estimated by Taylor expansion (An, Breidt and Fuller (1994)).

5 ESTIMATION OF RESPONSE PROBABILITIES

In this section, we describe the method we used to estimate the response probabilities. We assume that phase II sample is the full sample for the SIPP data, and the phase II sample consists of the respondents who responded on all seven interviews. We denote the response probability associated with the individual $(i,\ j,\ k)$ by p_{ijk} and let the indicator variable for response be r_{ijk}. The estimation procedure for p_{ijk} is composed of the following steps.

(1) In the phase II sample, regress the indicator variable for response on Y, and on both X and Y, respectively, and calculate the predicted value from the regressions,

$$\mathbf{r}_{\text{Reg. on } Y} \equiv (\hat{r}_{ijk}) = \mathbf{Y}^{(II)}\left(\mathbf{Y}^{(II)\prime}\mathbf{Y}^{(II)}\right)^{-1}\mathbf{Y}^{(II)\prime}\mathbf{r}, \tag{22}$$

and

$$\mathbf{r}_{\text{Reg. on } X \text{ and } Y} = \boldsymbol{\Omega}^{(II)}\left(\boldsymbol{\Omega}^{(II)\prime}\boldsymbol{\Omega}^{(II)}\right)^{-1}\boldsymbol{\Omega}^{(II)\prime}\mathbf{r}, \tag{23}$$

where $\boldsymbol{\Omega}^{(II)} = \left(\mathbf{X}^{(II)},\ \mathbf{Y}^{(II)}\right)$, $\mathbf{r} = (r_{ijk})$ is the $n^{(II)}$-dimensional column vector. We denote the difference of the predicted values from the two regressions by

$$\mathbf{diff} = \hat{\mathbf{r}}_{\text{Reg. on } Y} - \hat{\mathbf{r}}_{\text{Reg. on } X \text{ and } Y} \equiv (\text{diff}_{ijk}) \tag{24}$$

and denote the sample mean of $\hat{\mathbf{r}}_{\text{Reg. on } Y}$ by

$$\bar{\hat{\mathbf{r}}}_{\text{Reg. on } Y} = n^{(II)^{-1}}\mathbf{J}'\hat{\mathbf{r}}_{\text{Reg. on } Y}, \tag{25}$$

where $\mathbf{J}$ is a vector whose elements are one.

(2) Estimate the parameter vector, $\boldsymbol{\theta} = (\theta_1,\ \theta_2,\ \theta_3,\ \theta_4)'$, of a logit model,

$$\begin{aligned} 1 - p_{ijk} \;=\; & \Big(1 + \exp\Big\{\theta_1 \log\left[\hat{r}_{ijk}\left(1-\hat{r}_{ijk}\right)^{-1}\right] \\ & +\theta_2\text{diff}_{ijk} + \theta_3\text{diff}^2_{ijk} + \\ & +\theta_4\left(\hat{r}_{ijk} - \bar{\hat{\mathbf{r}}}_{\text{reg. on } Y}\right)\text{diff}_{ijk}\Big\}\Big)^{-1} \end{aligned} \tag{26}$$

Denote the estimates of $\boldsymbol{\theta}$ from (26) by $\hat{\boldsymbol{\theta}}$, and calculate the estimates for p_{ijk} by

$$\begin{aligned} \hat{p}_{ijk} \;=\; & 1 - \Big(1 + \exp\Big\{\hat{\theta}_1 \log\left[\hat{r}_{ijk}\left(1-\hat{r}_{ijk}\right)^{-1}\right] \\ & +\hat{\theta}_2\text{diff}_{ijk} + \hat{\theta}_3\text{diff}^2_{ijk} \\ & +\hat{\theta}_4\text{diff}_{ijk}\left(\hat{r}_{ijk} - \bar{\hat{\mathbf{r}}}_{\text{reg. on } Y}\right)\Big\}\Big)^{-1}. \end{aligned} \tag{27}$$

The estimated $\hat{p}_{ijk}$ in (27) will be used as the estimated response probability for individual $(i,\ j,\ k)$ in the mean estimators.

If we assume that the response probability for individual $(i,\ j,\ k)$ is p_{ijk}, and the respondents form a Poisson sample, then the expected value of the total number of respondents in the phase II sample is

$$\sum_{i,j,k} p_{ijk}. \tag{28}$$

The estimated response probabilities in (27) are such that

$$\sum_{i,j,k} \hat{p}_{ijk} = n^{(III)}, \tag{29}$$

which is the sample size of the phase III sample. The estimated response probabilities $\hat{p}_{ijk}$ in (27) are used in constructing the initial weights for the phase III sample described in (16).

To investigate the goodness of fit of the function, we compare the estimates with the realization. We divide the phase II sample into eight categories in Table 1. Each individual belongs to one category according to its estimated response probability. For example, for individual $(i,\ j,\ k)$, if

$$0.44 < \hat{p}_{ijk} \leq 0.55, \tag{30}$$

then this individual is classified into the category which corresponds to "$0.44 < \hat{p}_{ijk} \leq 0.55$" in the column "Est. Response probability" of Table 1. In Table 1, the column "Total Observations" contains the total number of individuals in the phase II sample that fell into the corresponding category. The column "Mean of $\hat{p}_{ijk}$" shows the mean value of $\hat{p}_{ijk}$ within each category. The column "Response Rate" is the percentage of the respondents within the category in the phase II sample, that is, in category ℓ,

$$\text{Response Rate} = \begin{pmatrix}\text{Total} \\ \text{Observations}\end{pmatrix}^{-1} \sum_{(i,j,k)\in\ell} r_{ijk}. \tag{31}$$

The column "Deviation" is the difference between the mean of the estimated response probabilities $\hat{p}_{ijk}$ and the response rate within each category,

$$\text{deviation} = (\text{mean of } \hat{p}_{ijk}) - (\text{response rate}). \tag{32}$$

These differences are small in absolute value, but the deviation for the (0.65, 0.75) cell is about two binomial standard errors. All estimated response probabilities exceed 25%, and the category $0.75 < \hat{p}_{ijk} \leq 0.85$ contains 46% of the individuals in the phase II sample.

Table 1 Summary of Estimated Response Probabilities

Est. Response Probability	Total Observations	Deviation (%)	Mean of $\hat{p}_{ijk}$ (%)	Response Rate (%)	Min $\hat{p}_{ijk}$ (%)	Max $\hat{p}_{ijk}$ (%)
$0 \leq \hat{p}_{ijk} < .25$	0	–	–	–	–	–
$.25 \leq \hat{p}_{ijk} < .35$	9	0.44	33.78	33.33	31.91	34.95
$.35 \leq \hat{p}_{ijk} < .45$	246	-1.86	41.64	43.50	35.30	44.91
$.45 \leq \hat{p}_{ijk} < .55$	654	1.37	50.76	49.39	45.10	55.00
$.55 \leq \hat{p}_{ijk} < .65$	1647	0.09	60.81	60.72	55.04	64.99
$.65 \leq \hat{p}_{ijk} < .75$	4645	-1.36	70.70	72.06	64.00	74.99
$.75 \leq \hat{p}_{ijk} < .85$	14081	0.49	81.03	80.54	75.00	85.00
$.85 \leq \hat{p}_{ijk} \leq .100$	9484	-0.13	87.47	87.60	85.00	95.39

6 APPLICATION TO THE SIPP DATA

We apply our methods to the Panel 1987 data from SIPP. The phase I sample is the Current Population Survey. The phase II sample is the Panel 1987 Wave One sample. The sample size of the phase II sample is 30,766 individuals in 11,660 households. The phase III sample is the Panel 1987 longitudinal sample. The sample size of the phase III sample is 24,429 individuals in 9,776 households.

The regression variables are based on the non-interview adjustment cells and on the Current Population Survey variables used by the Census Bureau to construct weights for the Panel 1987 longitudinal sample. The X-variables are the variables associated with the second-stage adjustment used by the Census Bureau. The second-stage adjustment variables are based on gender, age, race, family type, and household type. There are 97 X variables in our analysis. The Y variables are indicator variables for the non-interview adjustment cells in the first stage adjustment procedure described in Waite (1990). The non-interview adjustment cells are formed using variables such as level of income, race, education, type of income, type of assets, labor force status, and employment status. There are 79 Y variables in our analysis. The Z variables used in our analysis are Personal Income, Personal Earnings, Family Income, Family Earnings, Family Property Income, Family Means Tested Transfers, Family Other Income, Household Earnings, Household Property Income, Household Means Tested Transfers, and Household Other Income. All variables are recorded for January 1987 and for January 1989. For example, family income for January 1989 is the total income of the family with which the interviewed person lived in January 1989. A household may have more than one family.

The results for three-phase estimators with and without nonresponse probability adjustment are compared in Table 2. The column "Mean" shows the three-phase estimates for characteristics using initial weights with the nonresponse probability adjustment, and the column "t-test" gives the t-statistics for testing the effects of nonresponse probability adjustment on the mean estimators. The adjustment is significant for household income and related variables. The effects for other characteristics are not significant. This may due to the fact that the regression variables have produced adjustments equivalent to the probability adjustment. Table 2 also presents the estimated standard errors using two sets of initial weights. The column "s.e. with ad." gives the estimated standard errors for the three-phase estimator using the initial weights adjusted by the nonresponse probability, and the column of "s.e. without ad." are estimated standard errors using initial weights without the nonresponse adjustment. There is very little difference between the two estimated standard errors.

ACKNOWLEDGMENT

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

An, A. B. (1995). Regression estimation for finite population means in the presences of nonresponse. *Ph.D. dissertation*, Iowa State University, Ames, Iowa.

An, A. B., Breidt, F. J., and Fuller, W. A. (1994). Regression weighting methods for SIPP data. *Proceedings of the Survey Research Methodology Section, American Statistical Association,* 434-439.

Breidt, F. J. and Fuller, W. A. (1993). Regression weighting for multiphase samples. *Sankhyā Series* **B, 55**, 297-309.

Table 2 Comparison Between Three-Phase Estimators with and without Nonresponse Probabiliy Adjustment to the Intitial Weights

Characteristics	Mean($)	t-test	s.e. with ad.	s.e. without ad.
Jan 87 Personal Income	986.0	2.67	7.66	7.49
Jan 89 Personal Income	1043.6	2.26	7.45	7.42
Jan 87 Personal Earnings	761.4	1.55	7.03	6.90
Jan 89 Personal Earnings	799.7	1.04	6.55	6.54
Jan 87 Family Income	2708.1	-0.89	23.22	23.11
Jan 89 Family Income	2824.6	0.65	22.98	22.96
Jan 87 Family Earning	2217.4	-1.46	22.58	22.48
Jan 89 Family Earning	2296.6	-0.05	20.77	20.74
Jan 87 Family Property Income	148.8	0.88	5.24	5.18
Jan 89 Family Property Income	152.4	1.00	5.09	5.06
Jan 87 Family MTT*	31.6	-0.43	1.67	1.67
Jan 89 Family MTT	29.3	-0.39	1.69	1.66
Jan 87 Family Other Income	310.3	0.51	5.52	5.50
Jan 89 Family Other Income	346.3	0.55	8.46	8.48
Jan 87 HH**Income	2775.9	-3.10	23.52	23.45
Jan 89 HH Income	2896.2	-0.32	23.05	23.06
Jan 87 HH Earnings	2275.6	-3.33	22.87	22.82
Jan 89 HH Earnings	2345.9	-0.94	20.96	20.98
Jan 87 HH Property Income	150.5	0.81	5.25	5.19
Jan 89 HH Property Income	153.9	0.93	5.11	5.08
Jan 87 HH MTT	32.8	-0.67	1.81	1.83
Jan 89 HH MTT	30.3	-0.50	1.72	1.70
Jan 87 HH Other Incom	316.9	0.00	5.73	5.74
Jan 89 HH Other Incom	353.5	0.44	8.53	8.55
Jan 87 Labor Force (%)	46.5	-0.11	0.22	0.22
Jan 89 Labor Force (%)	47.4	-0.41	0.24	0.24

* MTT = Means Tested Transfers, ** HH = Household

Folsom, R. E. and Witt, M. B. (1994). Testing a new a attrition nonresponse adjustment method for SIPP. *Technical report. Research Triangle Institute, Research Triangle Park, North Carolina.*

Fuller, W. A. (1975). Regression analysis for sample survey. *Sankhyā Series* **C**, **37**, 117-132.

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.

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.