Estimation of Attrition Biases in SIPP

Eric V. Slud[1] and Leroy Bailey[2]
[1]Census Bureau, SRD, and Univ. of Maryland College Park
[2]Census Bureau, Statistical Research Division

Abstract

This paper studies estimators for the bias in estimated cross-sectional survey item totals due to attrition nonresponse weighting within a longitudinal survey. Adjustments for between-sample longitudinal nonresponse are made either by adjustment cells or by logistic regression. The bias estimators studied were first proposed by Bailey (2004) in connection with the Census Bureau's Survey of Income and Program Participation (SIPP), but are generalized here to include longitudinal survey items, and formulas and estimators for their design-based variances are given in terms of joint inclusion probabilities. In practice, variance estimates for the bias estimators are obtained, in the SIPP setting where PSU samples are drawn as balanced half-sample replicates, using either balanced replication methods or a formula of Ernst, Huggins and Grill (1986). The methods are illustrated using cross-sectional item data from the SIPP 1996 panel.

Keywords: Adjustment cells, Balanced replicates, Logistic regression, Longitudinal survey, Nonresponse, Variance estimation, VPLX.

1. Introduction

One of the persistent problems arising in large longitudinal surveys is to compensate through weighting schemes for nonresponse errors due to attrition. This problem is perennial for the Census Bureau's Survey of Income and Program Participation (SIPP), which has for many years been one of the largest national longitudinal surveys, measuring many variables related to family, employment status, insurance, amounts and sources of income, and indicators of participation in various government programs. SIPP's design follows panels of the order of 30,000–50,000 sampled households over a succession of staggered waves every four months for total durations typically of 3 years (SIPP User's Guide 2001). Attrition over the life of a panel is of the order of 30%: of the $94,444$ persons responding in Wave 1 of SIPP 96, 14% were lost by Wave 4, and 30% by the end of the panel, Wave 12.

While some aspects of the quality of reported data from a survey like SIPP can be judged only by comparing with related data from other sources, the problem of ascertaining and compensating for the biases due to attrition is largely a matter to be explored through internal consideration of the longitudinal data from the same survey. This is because the precise demographic and survey-design features and incentives associated with attrition are likely to be very different for another survey. At first sight, the possibility of assessing nonresponse biases by a purely internal statistical examination of survey data seems counterintuitive. That is especially so in a survey like SIPP, where the precise variables being measured — such as marriage and divorce, poverty or changes in employment status for one or more jobs between successive instances of followup questioning — will directly affect the chance of finding a subject at home and willing to respond to the survey's followup. However, by examining adjusted estimates of totals of *wave 1* cross-sectional variables, we separate the modeling of the propensity to respond from later changes in surveyed items.

Bailey (2004) studied the bias due to each of a number of attrition adjustment methods in SIPP by calculating for specific items the differences between the totals estimated at wave 1 versus the totals of the same wave-1 items estimated by nonresponse adjustment of the totals obtained from the wave-t subjects (who responded in wave 1). The approach of Bailey (2004), which we elaborate formally in Section 2, was restricted to single-wave (cross-sectional) questionnaire items and was purely descriptive in the sense that standard errors for the difference between the wave-1 and adjusted wave-t totals were not provided. The main methods of attrition nonresponse adjustment he considered were Horvitz-Thompson estimators (Särndal et al. 1992, pp. 42*ff.*) with weights derived either from a moderate number (of the order of 100, within a survey of more than 30,000 households) of adjustment cells or from a logistic regression model, specified in terms of Wave-1 items.

In this paper, we broaden Bailey's (2004) approach to provide formulas and estimators of the standard errors for the estimators of bias in estimating wave-1 cross-sectional item totals using adjusted wave-4 or wave-12 data. The estimating formulas given here for biases and standard errors generalize immediately to cover the case of bias estimation for longitudinal item totals based on early-wave versus adjusted later-wave data.

Exact variances of the bias estimators, and unbiased estimators for them, require knowledge of joint inclusion probabilities which are generally not available when, as in SIPP, the wave 1 inclusion weights take into account wave 1 nonresponse and are raked and *second-stage adjusted*. However, the SIPP 1990 redesign incorporated a feature of two balanced replicate samples within each sampled PSU. Using a theoretical formula of Ernst, Huggins and Grill (1986), we provide and implement closed-form and balanced repeated replicate (BRR) estimators of variance for the bias estimators we study.

The paper is organized as follows. Section 2 gives formulas from sample survey theory for estimating attrition nonresponse biases. Section 3 gives variance formulas, in terms of joint inclusion probabilities, of designed replicates, and of balanced repeated replicates as implemented in the Census Bureau's VPLX software. Section 4 illustrates and interprets the bias estimators and their estimated standard errors for the same eleven SIPP 96 (cross-sectional) items studied in Bailey's (2004) report, and the alternative estimators for standard errors are compared. Finally, tentative conclusions are drawn about the magnitudes and significance of SIPP biases due to attrition, and about important directions for further research.

2. Bias Estimation Formulas

Suppose that a sample $\mathcal{S}$ of subjects is drawn from sampling frame $\mathcal{U}$, with single inclusion-probabilities $\{\pi_i\}_{i\in\mathcal{U}}$ and joint inclusion probabilities $\{\pi_{ij}\}_{i,\,j\in\mathcal{U}}$. Suppose that 'baseline' observations $(\mathbf{x}_i,\, y_i^B,\, r_i)$ are recorded for all $i \in \mathcal{S}$, where y_i^B is an item of interest, $\mathbf{x}_i$ is a vector of auxiliary data, and r_i is the indicator of followup response. Suppose also that y_i^F denotes a followup measurement which is potentially defined for all population members but which is actually recorded only for $i \in \mathcal{S}$ such that $r_i = 1$. As is often done in survey theory, we treat estimated totals in a design-based framework except that the response indicators r_i are treated as random variables conditioned on $\mathcal{S}$, i.e., on all first-stage data. It can be assumed that $\pi_i,\, \pi_{ij}$ are known, but that $E(r_i) = P(r_i = 1)$ and the corresponding conditional probabilities $p_i = P(r_i = 1\,|\,\mathcal{S})$ are not known. The target parameter is

$$\vartheta = t_{y^F} - t_{y^B} = \sum_{i\in\mathcal{U}} (y_i^F - y_i^B),$$

where here and in what follows we adopt the standard notations that for any attribute $z_i,\, i \in \mathcal{U},$ the frame-population total is $t_z = \sum_{i\in\mathcal{U}} z_i$, and the corresponding Horvitz-Thompson estimator is $\hat{t}_z = \sum_{i\in\mathcal{S}} z_i/\pi_i$.

In this setup, the reciprocal probabilities π_i^{-1} are assumed to be weights which adjust correctly for first-stage (Wave-1) nonresponse. The quantities y_i^B are the first-stage measured survey variables of interest, usually first-stage cross-sectional survey data. In actual practice, y_i^F would be the corresponding variables which could be observed at a specified later stage (Wave) of the survey. The target parameter would then be the between-wave change in population total of the y attribute. However, to avoid confounding the true between-wave change with the errors we introduce due to sampling variability and imprecise modelling or estimation of the conditional probabilities p_i, we follow Bailey (2004) in the artificial choice $y_i^F = y_i^B$ for all i, making the target parameter $\vartheta = 0$ known, and allowing us to evaluate the effectiveness of adjustments made for later-wave nonresponse.

Now if the conditional probabilities $p_i = E(r_i|\mathcal{S}) = P(r_i = 1\,|\,\mathcal{S})$ *were* known, then the general unbiased estimator of $\vartheta = t_{y^F} - t_{y^B}$ would be $\sum_{i\in\mathcal{S}} (r_i y_i^F/p_i - y_i^B)/\pi_i$. The actual estimator $\hat{\vartheta}$ that we use depends on an estimator $\hat{p}_i$ for p_i under a parametric model (two specific examples of which will be discussed in detail below), within which the random variables $r_i,\, i \in U,$ are assumed to be conditionally independent given $\mathcal{S}$, with

$$p_i \;=\; P(r_i = 1\,|\,\mathcal{S}) \;=\; g(x_i, \beta) \quad , \qquad i \in \mathcal{S} \quad (1)$$

where x_i is the vector of auxiliary variables observed for each $i \in \mathcal{S}$, and β is a parameter of fixed dimension (much smaller than the frame population size $|\mathcal{U}|$). Under model (1), with $\hat{p}_i = g(x_i, \hat{\beta})$ generally derived from a weighted estimating equation in terms of the data $(r_i,\, i \in \mathcal{S})$, we obtain

$$\text{Wave-change Estimator} \;=\; \sum_{i\in\mathcal{S}} \frac{1}{\pi_i}\Big(y_i^F \,\frac{r_i}{\hat{p}_i} - y_i^B\Big) \quad (2)$$

Specializing further to the case of interest in this paper, we fix $y_i^F \equiv y_i^B$, in which case the estimator (2) can be viewed as the difference between the later-wave nonresponse-adjusted estimator of t_{y^B} and the first-Wave Horvitz-Thompson estimator of the same quantity. Therefore we refer to it simply as the

$$\text{Bias Estimator} \;=\; \sum_{i\in\mathcal{S}} \frac{y_i^B}{\pi_i}\Big(\frac{r_i}{\hat{p}_i} \,-\, 1\Big) \qquad (3)$$

We work with two different models of the form (1).

Adjustment Cell Model. Let the frame $\mathcal{U}$ be partitioned into cells $C_j\,,\; j=1,\ldots,K,$ with K fixed, and for the parameter $\beta\in(0,1)^K$, assume

$$p_i = \Pr(r_i=1\,|\,i\in\mathcal{S}) \;=\; \beta_j \qquad \text{whenever } \; i\in C_j,$$

where $i\in\mathcal{U},\; j=1,\ldots,K.$ Then the form of estimator for $p_i=\beta_j$ for $i\in C_j,$ which is unbiased for all inclusion-probabilities $\pi_i,$ is the *Cell Adjustment-Factor* given by

$$\hat{\beta}_j \;=\; \frac{\sum_{i\in\mathcal{S}\cap C_j}\, r_i\pi_i^{-1}}{\sum_{i\in\mathcal{S}\cap C_j}\, \pi_i^{-1}} \qquad (4)$$

Logistic Regression Model. Alternatively, for $\beta\in\mathbf{R}^d$ a coefficient parameter vector of the same dimension d as the covariate $x_i,$ we could assume that p_i depends on individual covariates through the formula $p_i=(1+e^{-x_i'\beta})^{-1}$ and then estimate $\hat{p}_i=(1+e^{-x_i'\hat{\beta}})^{-1},$ where $\hat{\beta}$ is the unique solution of the weighted score equation

$$\sum_{i\in\mathcal{S}} x_i\,\frac{1}{\pi_i}\Big(r_i \,-\, \frac{e^{x_i'\,\hat{\beta}}}{1+e^{x_i'\,\hat{\beta}}}\Big) \;=\; \mathbf{0} \qquad (5)$$

3. Estimating the Variance of Bias Estimators

Our next objective is to find general large-sample approximate expressions and unbiased estimators for the variance of the Wave-Change Estimator (2), and its specialization to the Bias Estimator (3). Under the parametric model (1), with consistent estimators $\hat{\beta}$ for β and estimators $\hat{p}_i=g(x_i,\hat{\beta})$ close to p_i (uniformly over $i\in\mathcal{S}$ with high probability, when the sample size $|\mathcal{S}|$ is large), a principal method for deriving variances is first to *linearize* the estimators of interest, i.e., to approximate the centered estimators by linear expressions in centered (Horvitz-Thompson) estimators $\hat{t}_z-t_z$ of survey-item totals t_z. We now proceed to do this for the estimator (2), denoted $\hat{\vartheta}$, under the two models considered in Section 2.

First, in the adjustment cell setting, define the "cell j attribute ξ_j" as the indicator which for individual i is $\xi_{j,i}=I_{[i\in C_j]}$. Then the adjustment-cell version of $\hat{\vartheta}-\vartheta,$ denoted $\hat{\vartheta}^A-\vartheta$ and defined from (2) with (4) substituted, is easily seen to have the form

$$\hat{\vartheta}^A-\vartheta \;=\; \sum_{j=1}^K \Big(\hat{t}_{r\,\xi_j\,y^F}\,\hat{t}_{\xi_j}/\hat{t}_{r\,\xi_j} \,-\, t_{r\,\xi_j\,y^F}\,t_{\xi_j}/t_{r\,\xi_j} \;-\hat{t}_{\xi_j\,y^B} \,+\, t_{\xi_j\,y^B}\Big) \qquad (6)$$

$$\approx \sum_{j=1}^K \Big(\frac{\hat{t}_{\xi_j}}{\hat{t}_{r\,\xi_j}}\,(\hat{t}_{r\,\xi_j\,y^F}-t_{r\,\xi_j\,y^F}) + \frac{\hat{t}_{r\,\xi_j\,y^F}}{\hat{t}_{r\,\xi_j}}\,(\hat{t}_{\xi_j}-t_{\xi_j}) \;-\frac{\hat{t}_{r\,\xi_j\,y^F}\,\hat{t}_{\xi_j}}{\hat{t}^2_{r\,\xi_j}}\,(\hat{t}_{r\,\xi_j}-t_{r\,\xi_j}) \,-\, \hat{t}_{\xi_j\,y^B} \,+\, t_{\xi_j\,y^B}\Big)$$

where $\approx$ means that the expressions on the left- and right-hand sides differ by an amount which is negligible in probability as $|\mathcal{S}|$ gets large, *with the population-size* $|\mathcal{U}|$ *much larger still*. But this last expression is approximately equal in the same sense to the difference $\hat{t}_{z^A}-t_{z^A},$ where the "attribute" z_i^A is defined whenever $i\in C_j$ by

$$z_i^A \;=\; \frac{\hat{t}_{\xi_j}}{\hat{t}_{r\,\xi_j}}\,r_i\,y_i^F + \frac{\hat{t}_{r\,\xi_j\,y^F}}{\hat{t}_{r\,\xi_j}} - \frac{\hat{t}_{r\,\xi_j\,y^F}\,\hat{t}_{\xi_j}}{\hat{t}^2_{r\,\xi_j}}\,r_i - y_i^B \quad (7)$$

Next, under the logistic-regression model for later-wave nonresponse, we find by first-order Taylor expansion of (5) in $\hat{\beta}$ around the value $\beta=\beta_{\mathcal{U}}$ satisfying

$$\sum_{i\in\mathcal{U}} x_i\,\Big(r_i \,-\, \frac{e^{x_i'\,\beta_{\mathcal{U}}}}{1+e^{x_i'\,\beta_{\mathcal{U}}}}\Big) \;=\; \mathbf{0}$$

that

$$\hat{\beta}-\beta_{\mathcal{U}} \;\approx\; \hat{\mathcal{I}}^{-1}\,\mathbf{T}$$

where

$$\hat{\mathcal{I}} = \sum_{i\in\mathcal{S}} \frac{x_i^{\otimes 2}}{\pi_i}\,\frac{e^{x_i'\hat{\beta}}}{(1+e^{x_i'\hat{\beta}})^2}$$

and

$$\mathbf{T} = \sum_{i\in\mathcal{S}} \frac{x_i}{\pi_i}\,(r_i - \frac{e^{x_i'\beta_{\mathcal{U}}}}{1+e^{x_i'\beta_{\mathcal{U}}}})$$

and for any vector $\mathbf{v},$ we define the notation $\mathbf{v}^{\otimes 2}=\mathbf{v}\mathbf{v}'$. Then the centered wave-change estimator $\hat{\vartheta}^L-\vartheta,$ with $\hat{\vartheta}^L$ given by (2) after substituting $\hat{p}_i=(1+e^{-x_i'\hat{\beta}})^{-1}$ for $\hat{\beta}$ satisfying (5), is

$$\hat{\vartheta}^L-\vartheta \;\approx\; \hat{t}_{r\,y^F\,(1+e^{-x'\beta_{\mathcal{U}}})} - t_{r\,y^F\,(1+e^{-x'\beta_{\mathcal{U}}})} \;-\hat{t}_{y^B} \,+\, t_{y^B} \,-\, \hat{t}'_{r\,x\,y^F\,e^{-x'\hat{\beta}}}\,\hat{\mathcal{I}}^{-1}\,\mathbf{T}$$

which has asymptotically the same variance as the centered total-estimator $\hat{t}_{z^L} - t_{z^L}$ defined in terms of the 'attribute'

$$z_i^L \;=\; r_i\, y_i^F\,(1 + e^{-x_i'\hat{\beta}}) - y_i^B \qquad\qquad$$
$$- \hat{t}'_{r\,x\,y^F\,e^{-x'\hat{\beta}}}\,\hat{\mathcal{I}}^{-1}\,x_i\,(r_i - \hat{p}_i) \qquad (8)$$

Thus the variance $V(\hat{\vartheta})$ of $\hat{\vartheta}$ could be found approximately by the Horvitz-Thompson variance estimator

$$\hat{V}(\hat{\vartheta}) \;=\; \sum_{i,k \in \mathcal{S}} \frac{\pi_{ik} - \pi_i \pi_k}{\pi_i \pi_k} \cdot \frac{z_i}{\pi_i} \cdot \frac{z_k}{\pi_k} \qquad (9)$$

respectively for $z_i \equiv z_i^A$ given by (7) under the adjustment-cell model and for $z_i \equiv z_i^L$ given by (8) under the logistic regression response model. inclusion probabilities are not available at person-level for many large surveys like SIPP, because even the single inclusion probabilities π_i are constructed by an elaborate raking and trimming process. Therefore, the variance formulas (9) are not directly applicable. However, the linearization idea just described, which is now a standard method described in textbooks like that of Särndal et al. (1992), was shown by Woodruff and Causey (1976) to be applicable within any replicate-based variance estimation method. That is, they showed that in large samples, replicate-based variance estimators of totals of the artificial attributes which make the linearized estimators take a Horvitz-Thompson form would provide accurate approximations to the variances of the estimators being linearized. We next show how the SIPP design allows replicate-based estimation of $V(\hat{\vartheta})$ in terms of the pseudo-attributes $z_i = z_i^A$ or z_i^L respectively under the adjustment-cell or logistic regression response models.

We begin by describing the 1990 redesign of SIPP in terms similar to those of Rottach (2004), building on the standard SIPP documentation (2001). (A more detailed description along the same lines can also be found in Slud (2006).) SIPP's complex multistage sampling design incorporated paired replicate samples within Primary Sampling Units (PSU's). At the national level, the survey is based upon *strata* — geographic units nested within County and MCD — consisting in 1996 of 112 self-representing (SR) strata, subdivided into 372 artificial PSU's all of which were sampled, together with 105 nonself-representing (NSR) strata in each of which exactly two were sampled according to the Durbin method with a fixed system of PSU-level probability weights. Within each PSU, a systematic sample of (a fixed number of) persons was drawn, apparently using a single randomized starting index, although the sampled individuals are alternately indexed into two classes which Rottach (2004) calls *half-samples* but which here will be referred to as *half-PSU*'s. The variance estimation techniques applied to SIPP, beginning with Ernst, Huggins and Grill (1986) and Fay (1989), treat these half-PSU's as though they were two independently drawn samples, e.g., as though two independently randomized starting indices had been used. $h = 1, 2$ denote half-PSU. Where needed, let $\mathcal{S}_s$ denote the set of sampled PSU's within stratum s. Let j be the index for individuals within a specified combination (s, i, h), and $\mathcal{S}_{s,i,h}$ denote the set of individuals sampled within the half-PSU (s, i, h). For NSR stratum s, let $\pi_1^{(s)}$, $\pi_2^{(s)}$ respectively denote the inclusion probabilities for the sampled PSU's in $\mathcal{S}_s$ respectively labelled $i = 1$ and 2, and let $\pi_*^{(s)}$ denote the joint inclusion probability for both sampled PSU's. (There may be more than two PSU's in SR strata, but all of their inclusion probabilities are 1.) Now denote the individual inverse weight or single inclusion probability by $\pi_{(s,i,h,j)}$ to conform with Section 2, since individuals now have the quadruple indices (s, i, h, j). Then the within-PSU weights for sampled individuals in sampled half-PSU's (s, i, h) are $\pi_i^{(s)} / \pi_{(s,i,h,j)}$.

In terms of these notations, the estimator of Ernst, Huggins and Grill (1986) which they show to be slightly upwardly biased for the variance of the total-estimator $\hat{t}_z$ for an attribute $z_{(s,i,h,j)}$, is

$$\hat{V}_{EHG} \;=\; \sum_{s \in \mathtt{NSR}} b_{s0} \left(\frac{Z_{s,1}}{\pi_1^{(s)}} - \frac{Z_{s,2}}{\pi_2^{(s)}} \right)^2 \qquad (10)$$
$$+ \sum_{s \in \mathcal{S}} b_{s,1} \sum_{i \in \mathcal{S}_f} \frac{(Z_{s,i,1} - Z_{s,i,2})^2}{(\pi_i^{(s)})^2}$$

where weighted half-PSU and PSU- aggregated totals are given by

$$Z_{s,i,h} \;=\; \sum_{j \in \mathcal{S}_{sih}} \pi_i^{(s)} \frac{z_{(s,i,h,j)}}{\pi_{(s,i,h,j)}} \;, \quad Z_{s,i} \;=\; \sum_{h=1}^{2} Z_{s,i,H}$$

and

$$b_{s,0} \;=\; \frac{\pi_1^{(s)}\,\pi_2^{(s)}}{\pi_*^{(s)}} - 1 \;, \quad b_{s,1} = \max\{1 - b_{s,0},\, 0\}$$

Note that $b_{s,0} \geq 0$ always for the Durbin method, and $b_{s,0} = 0$, $b_{s,1} = 1$ whenever $s \in \mathtt{SR}$.

In settings like SIPP, where replicate samples are available within PSU's, the Census Bureau often uses a variance estimation method of Fay (1989) implemented in Fay's VPLX software. This method generates a set of distinct multiplicative replicate weight-factors $f_{s,i,h,r}$ where $r = 1, \ldots, R$ denotes an index ($R = 160$ for SIPP 96). The variance $V(\hat{\vartheta})$ for a possibly nonlinear estimator $\hat{\tau}$ is then estimated as follows. For each replicate r, the estimator is recalculated with first-order inclusion weights $\pi^{-1}_{(s,i,h,j)}$ replaced by $f_{s,i,h,r}/\pi_{(s,i,h,j)}$ and the result denoted $\hat{\tau}^{(r)}$. In particular, if the estimator of interest was $\hat{t}_z$, then the r'th replicate is

$$\hat{\tau}^{(r)} = \sum_{h=1}^{2} f_{s,i,h,r} \, \frac{Z_{s,i,H}}{\pi_i^{(s)}} \qquad (11)$$

The Fay-method (1989) balanced repeated replicate (BRR) estimator for the variance of $\hat{\tau}$ is

$$\hat{V}_{Fay} \; = \; \frac{4}{R} \sum_{r=1}^{R} \left(\hat{\tau}^{(r)} - \frac{1}{R} \sum_{m=1}^{R} \hat{\tau}^{(m)} \right)^2 \qquad (12)$$

As shown by Fay (1989) and described in greater detail by Rottach (2004) and Slud (2006), if the estimator of interest is a Horvitz-Thompson total ($\hat{\tau} = \hat{t}_z$), then the Fay method variance estimator approximates $\hat{V}_{EHG}$ and is algebraically identical to it if the number R of replicates is large enough. (That sufficiently large number, as justified in Slud 2006, is 3 times the number of NSR strata minus twice number of NSR strata with $b_{s,1} = 0$, plus the number of SR PSU's, or 667 in SIPP 96. The number R of replicates used in SIPP 96 was 160.) In a small simulation study in Slud (2006), $\hat{V}_{Fay}$ and $\hat{V}_{EHG}$ were checked to be uniformly close to one another in cases where $\hat{\tau} = \hat{t}_z$ and, in all models except those with greatest imbalance between half-PSU's, also to be close to the empirical variance.

Since joint person-level inclusion probabilities are not available for SIPP, we have two available replicate-based methods for estimating the variances of the nonlinear wave-change or bias estimators $\hat{\vartheta}^A$, $\hat{\vartheta}^L$:

(1) The Fay method variance $\hat{V}_{Fay}$ for the wave-change estimator $\hat{\tau}$ given directly by (2).

(2) The Ernst, Huggins and Grill estimator $\hat{V}_{EHG}$ applied to the linearized attributes z_i^A or z_i^L given respectively by (7) or (8).

A third method which initially seems to be also feasible — the Fay method applied to the linearized total estimators based on attributes z_i^A or z_i^L — turns out in both adjustment models to give algebraically the same value as the ordinary Fay method (1), because of the identities $\hat{\vartheta}^A = \sum_{i \in \mathcal{S}} z_i^A$ and $\hat{\vartheta}^L = \sum_{i \in \mathcal{S}} z_i^L$ for each set of replicate weights.

The two estimators $\hat{V}_{Fay}$ and $\hat{V}_{EHG}$ as in (1)-(2) above will be calculated and compared in the next Section for various choices of attribute $y_i^B = y_i^F$ in SIPP 96. For other types of survey estimators, not the bias estimators studied here, linearized and BRR variances have previously been compared by Sae-Ung et al. (2004) within the SIPP context.

4. Results for SIPP 96 Data

We summarize the SIPP 96 computed results on wave-1 minus wave-4 or wave-12 weighted totals, along with estimated standard errors. We consider totals and differences based only on the 94,444 individuals in SIPP 96 who had positive wave-1 weights, which were the second-stage weights (`bwgt`) recoded by Julie Tsay and used in the report of Bailey (2004). Throughout, we follow Bailey (2004) in studying the following 11 SIPP cross-sectional items: indicators that the individual living in a Household receives (i) Food Stamps (`Foodst`), or (ii) Aid to Families with Dependent Children (`AFDC`); indicators that the individual receives (iii) Medicaid (`Mdcd`), or (iv) Social Security (`SocSec`); and indicators that the individual (v) has health insurance (`Heins`), (vi) is in poverty (`Pov`), (vii) is employed (`Emp`), (viii) is unemployed (`UnEmp`), (ix) is not in the labor force (`NILF`), (x) is married (`MAR`), or (xi) is divorced (`DIV`). The definition of later-wave response used in most longitudinal SIPP studies and also in the present paper is: participation in all waves cumulatively up to the later wave of interest.

The adjustment-cell nonresponse model considered here is standard for SIPP (Tupek 2002), consisting of 149 cells defined in terms of region and Wave-1 measures of education, income-level, employment status, race, ethnicity, asset types, and numbers of imputed items. All cells contained sufficiently many individuals for Waves 4 and 12, and weights within a reasonable range, so that pooling was not necessary. The logistic regression response model used the Wave-1 predictor indicator variables for: Poverty, White Not Hispanic, Black, Renter, College educated, household 'reference person', and two pairwise interactions (Renter$*$College, Black$*$College). The model is the same one used in Bailey (2004), with the addition of a Poverty indicator. Poverty was included in the model because it was highly signifcant in exploratory model fitting, because poverty is used in defining the standard SIPP set of adjustment cells, and most of all because we want to note the effect on population-wide bias estimates for a variable which is also a survey item whose total is regularly reported.

Table 1 presents the estimators for Wave 1 item totals and also the biases in those totals resulting from adjustment between waves 1 and 4, with either the Adjustment-Cell (`BiasC`) or Logistic Regression (`BiasL`) method. The `Wav1` column shows the frame population totals $\hat{t}_y$ for

Table 1: Wave 1 SIPP 96 item totals and Wave 4 vs. 1 bias estimators, given in thousands. Standard errors have been used in forming *Studentized* bias deviates Dev_C, Dev_L. Here **C** indicates Wave 4 adjustment by Cells, **L** by Logistic. The asterisk by `Poverty` recalls its use in Logistic regression model.

Item	Wav1	BiasC	BiasL	Dev_C	Dev_L
Foodst	27268	-85.8	494.8	-0.51	2.77
AFDC	14030	-55.3	338.4	-0.36	2.10
Mdcd	28173	153.1	778.6	1.13	5.29
SocSec	37087	699.3	413.6	5.62	2.88
Heins	194591	1629.7	1233.5	7.16	5.22
*Pov	41796	-770.3	29.5	-4.29	3.50
Emp	191201	189.4	216.7	1.44	1.32
UnEmp	6406	-336.7	-375.9	-6.02	-6.50
NILF	66647	147.3	163.5	1.18	1.01
MAR	114367	1253.2	95.2	6.46	0.52
DIV	18463	-206.0	-357.4	-2.03	-3.70

Table 2: Wave 1 SIPP 96 item totals and Wave 12 vs. 1 bias estimators, given in thousands. Standard errors were used to form *Studentized* deviates Dev_C, Dev_L. Cell-based adjustment using Wave 12 responses is indicated **C**, Logistic model-based adjustment **L** .

Item	Wav1	BiasC	BiasL	Dev_C	Dev_L
Foodst	27268	-1179	261	-3.46	0.63
AFDC	14030	-1452	-459	-4.98	-1.38
Mdcd	28173	-397	1169	-1.34	3.26
SocSec	37087	4142	3844	16.38	13.53
Heins	194591	3792	2527	8.30	5.03
*Pov	41796	-1528	245	-4.06	3.14
Emp	191201	-1449	-2242	-5.08	-6.14
UnEmp	6406	-744	-811	-5.66	-6.28
NILF	66647	2193	3058	7.93	8.35
MAR	114367	5287	2551	13.85	6.61
DIV	18463	-381	-689	-1.82	-3.48

the indicated cross-sectional items measured in Wave 1. (Recall that the Wave 1 'sample' includes only responders, with Wave 1 nonresponse taken into account in the inclusion probabilities π_i.) Fay-method BRR variance estimators for these Bias estimators were used to create Standard Errors, and the studentized estimators (`Bias/SE`) are given in the final two columns. The layout for Table 2 is completely analogous.

In Table 1, estimated bias tends to be small, no more than 2% of the Wave 1 total, except that `UnEmp` bias is 5% to 6% downward both in C and L columns, and the logistic-model biases for `AFDC, Mdcd` are roughly 3%. The Poverty bias for the logistic model is particularly small. The biases for Wave 12 versus Wave 1 adjustments, summarized in Table 2, are generally larger than for Wave 4. Items `AFDC, SocSec, UnEmp, NILF` all have relative Wave 12 adjustment biases more than 3% by both methods, with several greater than 10%.

The studentized bias-estimator ratios Dev_C and Dev_L, which should be compared with standard normal percentage points, are significant in Wave 4 (Table 1) even in some instances where relative biases are small, but in Wave 12 (Table 2) virtually all the studentized biases are highly significant.

The patterns in relative and studentized bias are similar for the two Tables, but not quite for the two adjustment methods. At least it is generally true in Table 1 that the biases for both adjustment methods have the same sign whenever both are significantly large; but there are several items for which one method but not the other shows a significantly large bias in adjustment. The pattern is not easy to interpret, because it evidently depends strongly on the exact choice of the adjustment-cell and logistic regression models used. We can see this most clearly in the tiny logistic-model bias for `Pov` in Table 1, since Poverty was explicitly chosen as a predictor variable, and the logistic model with an intercept ensures that the regressor-weighted summed deviations between response indicators and their model predictors are 0 over the whole population.

We should mention that the adjusted estimated totals $\sum_{i \in \mathcal{S}} r_i \, y_i^F / (\hat{p}_i \, \pi_i)$ of Wave 1 items for Wave 4 and Wave 12 responders in Tables 1 and 2 differ slightly from the totals reported in these Proceedings by Bailey (2006). The differences are due to modifications of π_i made in Bailey (2006) to reflect second stage adjustments, i.e., the raking of summed weights over certain subdomains to demographic (updated census) totals and trimming of resultant weights. Such second stage adjustments are in fact made in SIPP production estimates of cross-sectional population attributes in later waves.

We calculated the EHG estimators $\hat{V}_{EHG}^{1/2}$ of standard errors of for Bias estimators (3), corresponding to all of the (Fay-method BRR) standard errors $\hat{V}_{Fay}^{1/2}$ used in Tables 1 and 2. The results, which we display in Table 3 for Wave 4 and in Table 4 for Wave 12, were very interesting. Within Table 3, the Adjustment-Cell variances (`EHG.C, Fay.C`) are generally close, mostly within one or two percent, except for a 7% discrepancy in `Pov` and 19% in `UnEmp`. However, the differences were remarkably greater between the Fay and EHG standard errors of the bias in the logistic-regression-based adjustments, never less than 15-20%, and the ratios in the `Heins` and

Table 3: Standard errors calculated for Wave 4 versus Wave 1 item adjustment biases, respectively via the linearized EHG versus Fay-BRR methods (EHG vs. Fay prefix), for the Cell-based versus Logistic regression adjustment methods (C vs. L suffix). All standard errors given in thousands.

Item	EHG.C	Fay.C	EHG.L	Fay.L
Foodst	160.6	167.6	260.3	203.0
AFDC	151.1	155.3	190.0	165.9
Mdcd	138.8	135.6	210.5	160.4
SocSec	127.0	124.3	162.2	143.5
Heins	227.0	227.5	574.8	264.5
Pov	192.7	179.5	300.7	224.2
Emp	139.7	131.9	567.9	171.4
UnEmp	69.5	56.0	75.2	58.5
NILF	126.3	124.7	230.6	166.9
MAR	191.7	194.1	356.6	188.7
DIV	101.7	101.7	116.0	96.6

Table 4: Standard errors calculated for Wave 12 versus Wave 1 item adjustment biases, respectively via the linearized EHG versus Fay-BRR methods (EHG vs. Fay prefix), for the Cell-based versus Logistic regression adjustment methods (C vs. L suffix). All standard errors given in thousands.

Item	EHG.C	Fay.C	EHG.L	Fay.L
Foodst	324.6	341.2	566.4	459.1
AFDC	303.0	291.5	408.1	351.6
Mdcd	300.6	295.7	517.7	411.8
SocSec	266.2	252.8	370.6	284.3
Heins	469.5	456.3	142.7	584.3
Pov	384.2	376.6	653.4	472.2
Emp	283.2	285.4	1446.5	378.6
UnEmp	138.2	131.4	147.1	130.8
NILF	268.5	276.6	571.0	375.0
MAR	383.8	418.0	901.9	391.0
DIV	200.1	209.3	222.4	197.4

Emp cases were greater than 2 and 3 respectively. Overall, the Fay-method standard errors (Fay.L) are systematically smaller than those (EHG.L) estimated from linearized attributes z_i^L by the Ernst, Huggins, and Grill (1986) formula (10). The pattern of discrepancy between the two standard error estimators is very similar in Table 4, involving Wave 12 adjustment-biases, to that in Table 3.

5. Conclusions and Further Research Directions

Our results suggest that the quality of longitudinal nonresponse adjustment in SIPP 96 is particularly problematic for late-wave adjustments, with adjustments to Heins and UnEmp particularly biased in Wave 4, and Heins, SocSec, UnEmp, NILF, and Mar especially biased in Wave 12. However the assessed magnitudes of bias are highly dependent on the specific adjustment method used. Individual item biases can likely be made small, like that for Poverty in the logistic method in Table 1, when considered only for the whole population. This artificial effect of model-choice disappears when bias summaries are made over selected subdomains. For this reason, we propose in future work to devise combined or composite bias measures over subdomains and survey response variables, and to search for effective model-based adjustment methods by studying the behavior of such bias measures on actual SIPP data.

As part of our development of variance estimators for this bias evaluation study, we compared variance estimators based on a formula of Ernst, Huggins, and Grill (1986) using linearized estimators, with the BRR method of Fay (1989) which can be used directly on estimators which are not linear combinations of Horvitz-Thompson weighted totals. Since asymptotic statistical theory based on linearization is ultimately the mathematical justification for both methods (Woodruff and Causey 1976, Krewski and Rao 1981), it is very interesting and slightly disturbing to find that the results from the two methods disagree meaningfully for many of the SIPP 96 survey variables and disagree drastically for a few of them. We propose to study further the reasons for these differences.

Acknowledgments

We gratefully acknowledge the help of Mahdi Sundukchi and John Boies of DSMD and Julie Tsay of SRD in file acquisition. Rick Valliant made some helpful comments about an earlier writeup on SIPP variance estimation. Most helpful of all, Reid Rottach, formerly of DSMD, gave the first author several extremely clear and patient explanations of his 2004 memo and its relation to the VPLX data-structure in the SIPP 96 context.

References

Bailey, L. (2004) “Weighting alternatives to compensate for longitudinal nonresponse in the Survey of Income and Program Participation”, Census Bureau internal report, Nov. 16, 2004.

Bailey, L. (2006), “Assessing the effectiveness of weighting cell adjustments for longitudinal nonresponse”, *ASA Proc. Survey Res. Meth.*, Seattle, WA.

Ernst, L., Huggins, V., and Grill, D. (1986), "Two new variance estimation techniques", *ASA Proc. Survey Res. Meth.*, Washington DC, pp. 400-405.

Fay, R. (1989), "Theory and application of replicate weighting for variance calculations", *ASA Proc. Survey Res. Meth.*, Washington DC, pp. 212-217.

Krewski, D. and Rao, J.N.K. (1981), "Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods", *Ann. Statist.* **9**, 1010-1019.

Rottach, R. (2004), "Internal-use replicate weighting system for SIPP (VAR-1)", *Census Bureau Memorandum, DSMD,* March 22, 2004.

Sae-Ung, S., Hall, D., and Sissell, C. (2004), "Evaluation of variance estimation methods based on a SPD longitudinal file", *ASA Proc. Survey Res. Meth.*, Toronto, Canada, pp. 4296-4303.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992) *Model Assisted Survey Sampling*, Springer-Verlag, New York.

Slud, E. (2006), "Variance estimation and approximation in the setting of SIPP 96", Census Bureau Report, Statist. Res. Div.

Survey of Income and Program Participation User's Guide (Supplement to the Technical Documentation) (2001), Third Edition, Washington DC.

Tupek, A. (2002), "SIPP 96: specifications for the longitudinal weighting of sample people", Internal Census Bureau memorandum.

Woodruff, R. and Causey, B. (1976), "Computerized method for approximating the variance of a complicated estimate", *Jour. Amer. Statist. Assoc.* **71**, 315-321.