Estimation of Attrition Biases in SIPP

Eric V. Slud and Leroy Bailey

1Census Bureau, SRD, and Univ. of Maryland College Park
2Census 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.

This report is released to inform interested parties of research and encourage discussion. The views expressed on statistical and methodological issues are those of the authors and not necessarily those of the Census Bureau.

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. 42ff.) with weights derived either from a moderate number (of the order of 100, within a survey of more than 30,000 households) of ad-
Suppose that a sample \( S \) 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 \((x_i, y_i^B, r_i)\) are recorded for all \( i \in S \), where \( y_i^B \) is an item of interest, \( x_i \) 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 \( i \in S \) 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 \( S \), 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 \( p_i = P(r_i = 1 \mid S) \) are not known. The target parameter is

\[
\hat{\theta} = t_y^p - t_y^n = \sum_{i \in \mathcal{U}} (y_i^F - y_i^B),
\]

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 \( \hat{t}_z = \sum_{i \in S} z_i/\pi_i \).

In this setup, the reciprocal probabilities \( \pi_i^{-1} \) are assumed to be weights which adjust correctly for first-stage (Wave-1) nonresponse. The quantities \( y_i^F \) 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 \( \hat{\theta} = 0 \) known, and allowing us to evaluate the effectiveness of adjustments made for later-wave nonresponse.

Now if the conditional probabilities \( p_i = E(r_i | S) = P(r_i = 1 \mid S) \) were known, then the general unbiased estimator of \( \hat{\theta} = t_y^p - t_y^n \) would be \( \sum_{i \in S} (r_i y_i^F / p_i - y_i^B) / \pi_i \). The actual estimator \( \hat{\theta} \) 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 \( r_i, i \in \mathcal{U} \), are assumed to be conditionally independently given \( S \), with

\[
p_i = P(r_i = 1 \mid S) = g(x_i, \beta), \quad i \in S \tag{1}
\]

where \( x_i \) is the vector of auxiliary variables observed for each \( i \in S \), and \( \beta \) 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 \((r_i, i \in S)\), we obtain

\[
\text{Wave-change Estimator} = \sum_{i \in S} \frac{1}{\pi_i} (y_i^F r_i/\hat{p}_i - y_i^B) \tag{2}
\]

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 \( t_y^n \) and
the first-Wave Horvitz-Thompson estimator of the same quantity. Therefore we refer to it simply as the

\[
\text{Bias Estimator} = \sum_{i \in S} \frac{y_i^B}{\pi_i} \left( r_i - 1 \right)
\]  

(3)

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

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

\[
p_i = \Pr(r_i = 1 | i \in S) = \beta_j \text{ whenever } i \in C_j,
\]

where \( i \in U, j = 1, \ldots, K \). 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

\[
\hat{\beta}_j = \frac{\sum_{i \in S \cap C_j} r_i \pi_i^{-1}}{\sum_{i \in S \cap C_j} \pi_i^{-1}}
\]

(4)

**Logistic Regression Model.** Alternatively, for \( \beta \in \mathbb{R}^d \) 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 \( \hat{p}_i = (1 + e^{-x_i'\beta})^{-1} \), where \( \hat{\beta} \) is the unique solution of the weighted score equation

\[
\sum_{i \in S} x_i \left( 1 - \frac{e^{x_i'\hat{\beta}}}{1 + e^{x_i'\hat{\beta}}} \right) = 0
\]

(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 \( \beta \) and estimators \( \hat{p}_i = g(x_i, \hat{\beta}) \) close to \( p_i \) (uniformly over \( i \in S \) with high probability, when the sample size \(|S|\) 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 \( t_z - t_2 \) 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 \( \xi_{j,i} \) 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

\[
\hat{\vartheta}_A - \vartheta = \sum_{j=1}^{K} \left( \hat{\tau}_{r \xi_j y^r} \hat{\tau}_{\xi_j} - \tau_{r \xi_j y^r} \tau_{\xi_j} \right)
\]

(6)

\[
\approx \sum_{j=1}^{K} \left( \hat{\tau}_{r \xi_j y^r} \hat{\tau}_{\xi_j} - \tau_{r \xi_j y^r} \tau_{\xi_j} \right)
\]

where \( \approx \) means that the expressions on the left- and right-hand sides differ by an amount which is negligible in probability as \(|S|\) gets large, with the population-size \(|U| \) much larger still. But this last expression is approximately equal in the same sense to the difference \( \hat{\tau}_A - \tau_A \), where the “attribute” \( z_i^A \) is defined whenever \( i \in C_j \) by

\[
z_i^A = \frac{\hat{\tau}_{r \xi_j y^r} r_i}{\hat{\tau}_{r \xi_j} r_i} y_i^F - \frac{\hat{\tau}_{r \xi_j y^r} \tau_{\xi_j} - \tau_{r \xi_j y^r} \tau_{\xi_j}}{\hat{\tau}_{r \xi_j}^2} r_i - y_i^B
\]

(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 \( \beta = \beta_U \) satisfying

\[
\sum_{i \in U} x_i (r_i - \frac{e^{x_i'\beta_U}}{1 + e^{x_i'\beta_U}}) = 0
\]

that

\[
\hat{\beta} - \beta_U \approx \hat{\beta}' - T
\]

where

\[
\hat{\beta}' = \sum_{i \in S} \frac{x_i}{\pi_i} \frac{e^{x_i'\hat{\beta}}}{(1 + e^{x_i'\hat{\beta}})^2}
\]

and

\[
T = \sum_{i \in S} \frac{x_i}{\pi_i} (r_i - \frac{e^{x_i'\beta_U}}{1 + e^{x_i'\beta_U}})
\]

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

\[
\hat{\vartheta}_L - \vartheta \approx \hat{\tau}_{r y^r} (1 + e^{-x_i'\beta_U}) - \tau_{r y^r} (1 + e^{-x_i'\beta_U})
\]

\[
- \hat{\tau}_{y^r} - \tau_{y^r} - \hat{\vartheta}'_{r x y^r} e^{-x_i'\hat{\beta}_U} \hat{\beta}' - T
\]
which has asymptotically the same variance as the centered total-estimator \( t_{zL} - t_z \) defined in terms of the ‘attribute’

\[
z_i^L = r_i y_i^F (1 + e^{-x_i^F \beta}) - y_i^B - B_r x_i y_i e^{-x_i^F \beta} \hat{\mathcal{I}}^{-1} z_i (r_i - \hat{\beta}_i) \tag{8}
\]

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

\[
\hat{V}(\hat{\theta}) = \sum_{i,k \in S} \left( \frac{\pi_{ik} - \pi_i \pi_k}{\pi_i \pi_k} \right) \cdot \frac{z_i}{\pi_i} \cdot \frac{z_k}{\pi_k} \tag{9}
\]

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 \( \pi_i \) 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 (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-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 \( V(\hat{\theta}) \) in terms of the pseudo-attributes \( z_i \equiv z_i^A \) 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 \( s \) respectively labelled \( i = 1 \) and \( 2 \), and let \( \pi_4^{(s)} \) 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 \( \pi_1^{(s)} / \pi(s, i, h, j) \).

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 \( t_z \) for an attribute \( z(s, i, h, j) \), is

\[
\hat{V}_{EHG} = \sum_{s \in \mathcal{R}} b_{s,0} \left( \frac{Z_{s,1}}{\pi_1(s)} - \frac{Z_{s,2}}{\pi_2(s)} \right)^2 + \sum_{s \in \mathcal{S}} b_{s,1} \sum_{i \in \mathcal{S}_{s,i}} \left( \frac{Z_{s,i,1} - Z_{s,i,2}}{\pi_4^{(s)}} \right)^2 \tag{10}
\]

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

\[
Z_{s,i,h} = \sum_{j \in \mathcal{S}_{s,i,h}} \pi_4^{(s)} \frac{\pi(s, i, h, j)}{\pi(s, i, h, j)} \quad Z_{s,i} = 2 \sum_{h=1} Z_{s,i,h}
\]

and

\[
b_{s,0} = \frac{\pi_1^{(s)}}{\pi_4^{(s)}} - 1 \quad b_{s,1} = \max\{1 - b_{s,0}, 0\}
\]

Note that \( b_{s,0} \geq 0 \) always for the Durbin method, and \( b_{s,0} = 0, b_{s,1} = 1 \) whenever \( s \in \mathcal{R} \).

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 \((R = 160 \text{ for SIPP 96})\). The variance \( V(\hat{\theta}) \) 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 \( \pi_{(s,i,h,j)} \) replaced by \( \hat{\pi}_{(s,i,h,j)} \) and the result denoted \( \hat{\tau}(r) \). In particular, if the estimator of interest was \( \hat{\tau}_z \), then the \( r \)th replicate is

\[
\hat{\tau}(r) = \frac{2}{R} \sum_{h=1}^{R} z_{s,i,h,r} \hat{\pi}_{(s,i,h)} \tag{11}
\]

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

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

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 \( \hat{\tau} = \hat{\tau}_z \), 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 \( b_{s,i} = 0 \), 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 \( \hat{\tau} = \hat{\tau}_z \) 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 \( \hat{\tau}^A \), \( \hat{\tau}^L \):

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 \( z^A \) or \( z^L \) 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 \( z^A \) or \( z^L \) — turns out in both adjustment models to give algebraically the same value as the ordinary Fay method (1), because of the identities \( \hat{\tau}^A = \sum_{i \in S} z^A_i \) and \( \hat{\tau}^L = \sum_{i \in S} z^L_i \) 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 \( y^B = y^L \) 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 (Mdcid), 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 (NFLF), (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 significant 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 \( \hat{V}_{Fay} \) column shows the frame population totals \( \hat{\tau}_y \) for
shows a significantly large bias in adjustment. The pat-
several items for which one method but not the other
sign whenever both are significantly large; but there are
the biases for both adjustment methods have the same
methods. At least it is generally true in Table 1 that
for the two Tables, but not quite for the two adjustment
The patterns in relative and studentized bias are similar
in some instances where relative biases are small, but in
Wave 12 (Table 2) virtually all the studentized biases are
highly significant.

The indicated cross-sectional items measured in Wave 1.
(Recall that the Wave 1 'sample' includes only respon-
ders, with Wave 1 nonresponse taken into account in
the inclusion probabilities \( \pi_i \).) Fay-method BRR vari-
ance 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 lay-
out 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 adjust-
ments, 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 \( \text{Dev}_C \) and \( \text{Dev}_L \),
which should be compared with standard normal per-
centage 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 pat-
tern 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 vari-
able, and the logistic model with an intercept ensures that
the regressor-weighted summed deviations between re-
sponse indicators and their model predictors are 0 over
the whole population.

We should mention that the adjusted estimated totals
\[ \sum_{i \in S} \hat{y}_i^C / \langle \hat{p}_i \rangle_1 \] 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 Bai-
ley (2006). The differences are due to modifications of
\( \pi_i \) made in Bailey (2006) to reflect second stage adjust-
ments, i.e., the raking of summed weights over certain
subdomains to demographic (updated census) totals and
trimming of resultant weights. Such second stage adjust-
ments are in fact made in SIPP production estimates of
cross-sectional population attributes in later waves.

We calculated the EHG estimators \( \hat{V}_F^{1/2} \) of standard
errors of for Bias estimators (3), corresponding to all of
the (Fay-method BRR) standard errors \( \hat{V}_F^{1/2} \) 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 remark-
ablely 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>
<thead>
<tr>
<th>Item</th>
<th>Wave 1</th>
<th>BiasC</th>
<th>BiasL</th>
<th>Dev_C</th>
<th>Dev_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foodst</td>
<td>27268</td>
<td>-85.8</td>
<td>494.8</td>
<td>-0.51</td>
<td>2.77</td>
</tr>
<tr>
<td>AFDC</td>
<td>14030</td>
<td>-55.3</td>
<td>338.4</td>
<td>-0.36</td>
<td>2.10</td>
</tr>
<tr>
<td>Mdcd</td>
<td>28173</td>
<td>153.1</td>
<td>778.6</td>
<td>1.13</td>
<td>5.29</td>
</tr>
<tr>
<td>SocSec</td>
<td>37087</td>
<td>699.3</td>
<td>413.6</td>
<td>5.62</td>
<td>2.88</td>
</tr>
<tr>
<td>Heins</td>
<td>194591</td>
<td>1629.7</td>
<td>1233.5</td>
<td>7.16</td>
<td>5.22</td>
</tr>
<tr>
<td>*Pov</td>
<td>41796</td>
<td>-770.3</td>
<td>29.5</td>
<td>-4.29</td>
<td>3.50</td>
</tr>
<tr>
<td>Emp</td>
<td>191201</td>
<td>189.4</td>
<td>216.7</td>
<td>1.44</td>
<td>1.32</td>
</tr>
<tr>
<td>UnEmp</td>
<td>6406</td>
<td>-336.7</td>
<td>-375.9</td>
<td>-6.02</td>
<td>-6.50</td>
</tr>
<tr>
<td>NILF</td>
<td>66647</td>
<td>147.3</td>
<td>363.5</td>
<td>1.18</td>
<td>1.01</td>
</tr>
<tr>
<td>MAR</td>
<td>114367</td>
<td>1253.2</td>
<td>95.2</td>
<td>6.46</td>
<td>0.52</td>
</tr>
<tr>
<td>DIV</td>
<td>18463</td>
<td>-206.0</td>
<td>-357.4</td>
<td>-2.03</td>
<td>-3.70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Wave 1</th>
<th>BiasC</th>
<th>BiasL</th>
<th>Dev_C</th>
<th>Dev_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foodst</td>
<td>27268</td>
<td>-1179</td>
<td>261</td>
<td>-3.46</td>
<td>0.63</td>
</tr>
<tr>
<td>AFDC</td>
<td>14030</td>
<td>-1452</td>
<td>-459</td>
<td>-4.98</td>
<td>-1.38</td>
</tr>
<tr>
<td>Mdcd</td>
<td>28173</td>
<td>-397</td>
<td>1169</td>
<td>13.53</td>
<td></td>
</tr>
<tr>
<td>SocSec</td>
<td>37087</td>
<td>4142</td>
<td>3844</td>
<td>16.38</td>
<td></td>
</tr>
<tr>
<td>Heins</td>
<td>194591</td>
<td>3792</td>
<td>2527</td>
<td>8.30</td>
<td></td>
</tr>
<tr>
<td>*Pov</td>
<td>41796</td>
<td>-1528</td>
<td>245</td>
<td>-4.06</td>
<td>3.14</td>
</tr>
<tr>
<td>Emp</td>
<td>191201</td>
<td>-1449</td>
<td>-2242</td>
<td>-5.08</td>
<td>-6.14</td>
</tr>
<tr>
<td>UnEmp</td>
<td>6406</td>
<td>-744</td>
<td>-811</td>
<td>-5.66</td>
<td>-6.28</td>
</tr>
<tr>
<td>NILF</td>
<td>66647</td>
<td>2193</td>
<td>3058</td>
<td>8.35</td>
<td></td>
</tr>
<tr>
<td>MAR</td>
<td>114367</td>
<td>5287</td>
<td>2551</td>
<td>13.85</td>
<td></td>
</tr>
<tr>
<td>DIV</td>
<td>18463</td>
<td>-381</td>
<td>-689</td>
<td>-1.82</td>
<td>-3.48</td>
</tr>
</tbody>
</table>

ASA Section on Survey Research Methods

Table 1: Wave 1 SIPP 96 item totals and Wave 4 vs. 1 bias estimators, given in thousands. Standard errors have
been in forming Studentized bias deviates \( \text{Dev}_C \) and
\( \text{Dev}_L \). Here \( C \) indicates Wave 4 adjustment by Cells,
\( L \) by Logistic. The asterisk by Poverty recalls its use in
Logistic regression model.

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 \( \text{Dev}_C \), \( \text{Dev}_L \). Cell-
based adjustment using Wave 12 responses is indicated
\( C \), Logistic model-based adjustment \( L \).

3718
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.

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

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


