Comparison of Small Area Models in SAIPE

Tapabrata Maiti and Eric V. Slud

May 6, 2003

Abstract. The ongoing Small Area Income and Poverty Estimates (SAIPE) project at the Census Bureau estimates numbers of poor school-age children by state, county, and ultimately school district, based upon Current Population Survey (CPS) and IRS data along with information from the latest decennial census. The current SAIPE county-level methodology relies on a Fay-Herriot (1979) model fitted to log-counts of related school-age children in CPS-sampled households, and discards data from those sampled counties with no sampled poor children. The present paper compares SAIPE small-area estimation by analogous Fay-Herriot models for logarithms of county child poverty *rates* with a unit- (i.e., individual-) level logistic regression model with county-level random effects (GLMM). This comparison is based upon several loss criteria applied to SAIPE datasets from 1994 and 1990, using CPS weighted estimates or (in 1990) decennial census data as standards of truth for the county-level child poverty rates being estimated. The GLMM is shown to fit the data better than the log-rate Fay-Herriot models, when judged by the internal evidence of the 1994 and 1990 CPS datasets. SAIPE's Fay-Herriot fitting method for 1990 log-rates performs excellently in matching to the 1990 Census log-rate in CPS-sampled counties, but worse than GLMM in counties with no CPS sample.

Key words: empirical best linear unbiased predictor (EBLUP), Fay-Herriot model, mean-squared errors, generalized linear mixed model, mixed effect logistic regression, loss function, small area estimation.

Acknowledgment. We thank the Census Bureau for research support during and after our ASA/Census Fellowships; William Bell for advice and comments; JASA editors and referees and Don Malec for suggestions which improved the exposition; and Robin Fisher for help with SAIPE data files. Conclusions and results are our own and are not endorsed by the Bureau.

1 Introduction: Historical Aspects of SAIPE

Under the terms of Title I of the Elementary and Secondary Education Act, more than $7 billion in compensatory education funds annually are allocated to counties and school districts using a formula involving child poverty-rate estimates. The SAIPE approach to county-level estimates was developed in response to legislation in 1994 (NRC Report of the National Academy of Sciences Panel of Estimates for Small Geographic Areas, Citro & Kalton 2000a, p. 3) calling for the Census bureau to supply 'updated estimates' of county-level child poverty for use in Title I allocations to counties in 1997-98 and 1998-99, and thereafter to provide estimates at school-district level. Prior to that time, decennial census data had been the source for such estimates (NRC Report 2000a, p. 16). 'Updated estimates' were to be based on models using census data plus data from other sources. The 'other sources' which have been chosen for this purpose are the annual Current Population Survey (CPS) (NRC Report 2000a, pp. 19-27) and administrative-records data (NRC Report 2000a, pp. 28-30) from IRS (income tax returns) and the Food Stamp program.

The CPS is the primary national survey measuring population and poverty each year, applying a rotating panel design to provide monthly data for clustered housing units sampled within a weighted probability sample of geographic units including about 1300 counties annually. It was chosen in the SAIPE program to provide the major indicators of national county-level changes in child poverty, in the form of sample-weighted estimates of numbers and proportion of poor children among children aged 5–17 related to primary householder (*poor related school-age children*). In SAIPE, and in this paper, the CPS data are analyzed as a two-stage sample with Primary Sampling Unit (PSU) equal to county. For improved stability of estimates, three years of CPS data (including that of the year before and the year after the income-year of interest) are combined to provide the response variable (logarithm of CPS weighted county estimate of number of poor related school-age children, or logarithm of ratio of CPS weighted county estimate of number of poor related school-age children over CPS estimate of related school-age children) in the models described below. Thus, in what we refer to as the *1994 SAIPE dataset*, the combined CPS (March income supplement) samples in the years 1992–1994, corresponding to 1488 distinct counties, were used to produce county-level child poverty estimates for the year 1993. Similarly, the *1990 SAIPE dataset* refers to CPS aggregates over the 1259 counties providing data in 1988–1990, for estimation of poverty in 1989.

The objective of the SAIPE county model is to use decennial-census and administrative predictor variables to express the similarity of child-poverty data across counties, thereby 'borrowing strength' (Ghosh and Rao 1994) from observed data to compensate for the absence of many counties from 3-year CPS samples and for the smallness of samples in many other counties. Since the outcome of modeling is to estimate numbers of related poor school-age children in *all* counties, SAIPE is constrained to use only administrative-record predictor variables which are available in appropriately aggregated form for all counties nationally. Since the IRS and CPS data, along with other national records, are by law confidential at the individual level, the constraint of uniform national coverage is coupled to further constraints regarding strong agency controls on the manner of release of data. The most useful variables which have been found to meet these constraints are the county numbers of child exemptions for families in poverty and of all child exemptions reported on tax returns, along with county numbers of households participating in the Food Stamp program.

Many exploratory analyses of SAIPE data with alternative models have been performed over the last decade (NRC Report 2000a, Chapter 5), in order to choose the best available model specification and small-area predictors from the variables derived from IRS county-aggregated data, CPS sample data, and long-form county aggregates from the most recent decennial census. The modeling framework chosen is that of Fay and Herriot (1979), as described in detail in the next S

ection. The four models which turned out to be the best (NRC Report 2000a, models (a)-(d) p. 56), primarily from the point of view of maximized likelihood and adequacy of fit to the CPS-derived response variables, included (logarithms of) the administrative-records predictors described above together with the logarithm of either the under-21 county population or the under-18 county population demographically updated from estimates of the previous decennial census. The under-18 population count was found clearly superior to the under-21 in the log-Number model which was judged the best, cited from now on as the *standard SAIPE model*, in which the response variable is the log number of related school-age poor children and the predictors are logs of numbers (rather than rates) from administrative records. Although the Report claims that the under-18 choice actually worsened the performance of the log-rate model, the difference is very slight, and we consider here only the under-18, not the under-21, population counts in defining regressors.

The current-year county-level predictor variables for the SAIPE *log-Rate* county model are:

`LTAXRT = logarithm of IRS-estimated Child Poverty Rate;`

`LSTMPRT = logarithm of Food-Stamp Participation rate;`

`LFILRT = logarithm of IRS Child Tax-Exemptions divided by Population Estimate;`

`LCPRT = logarithm of Poverty Rate for residents aged 5--17 from the latest decennial census.`

In either the log-Number or log-Rate form of the SAIPE Fay-Herriot (FH) model, the response variable cannot be computed in a PSU with sampled children when the sample contains no related *poor* school-age children. The data from such sampled counties are dropped when estimating model parameters. (In 1994, 304 such PSU's out of 1488 were dropped, and in 1990, 231 out of 1259.) The non-sampled counties and counties with no poor related school-age children in the sample are somewhat different from typical counties nationally: since the largest counties are always sampled, the non-sampled counties tend to be smaller and more rural, and that is obviously true also of the sampled but dropped counties. The regression model is supposed to hold equally well over all counties, but is fitted using only those with sampled related poor school-age children. Moreover, although the relative modeling effectiveness on sampled versus non-sampled counties, in the decennial year 1990 when external comparisons on all counties were possible, was not a primary concern of the National Academy of Sciences (NAS) panel, the NRC Report (2000a, p. 162) indicates that the standard SAIPE model over-predicts poverty in small counties. Both Reports (2000a, p. 162; 2000b, p. 6, 2nd bullet) urge as a research priority for SAIPE that estimation techniques such as GLMM which would not drop sampled counties should be developed. To remove unwanted time discrepancies between predictors and responses, and to reduce time lags between income-year and predictors, the NRC Report (2000a, p. 30, p. 162) urges that research move toward an analysis method which would operate on single-year CPS data if possible. Single-year samples would contain an even larger proportion of counties without sampled poor related children which would have to be dropped in fitting the SAIPE FH models.

For these reasons, we compare the SAIPE FH models with Generalized Linear Mixed Models (GLMM's), specifically with a class of random-intercept logistic regression models. Since such models directly specify and estimate the *rate* of child poverty rather than the number of poor children, and since the differences in performance between the SAIPE standard and log-Rate models were not large (Report 2000a, Chapters 5-6 and Appendices B-C), we restrict attention to the log-Rate model, re-fitting it and related FH models along with the GLMM's.

1.1 Organization of the paper

This paper is organized as follows. Section 2 defines the models on which the existing and proposed Small Area Estimation (SAE) methods are based, with SAE formulas given in section 2.1; criteria for SAE comparisons defined in section 2.2; description and summary in section 2.3 of SAE biases in related simulations; and explanation in section 2.4 of how the methods take into account within- and between- PSU weighting in the CPS sample design. Section 3 details the data-analytic comparison of the models studied, proceeding from a descriptive analysis of CPS and census data in Section 3.1, to exact specification of all models in Section 3.2, and display of results respectively versus internal (CPS) and external (census) standard of truth in Sections 3.3 and 3.4. The results are pulled together in the discussion of Section 3.5 to yield conclusions and recommendations in Section 4.

2 Competing Methods for SAIPE Estimation

The mixed-effect linear models used and seriously considered in SAIPE are all of the following general Fay-Herriot (1979) model (FH) form. For each primary sampling unit (PSU), or county, indexed by $i = 1, \ldots, m$, sample-sizes n_i and p-dimensional vectors x_i of predictor variables are known (with n_i artificially re-set to 0 in the case of log-transformed responses when no *poor* children are sampled in PSU i). Response-variables satisfying

$$y_i = x_i^{\mathrm{tr}} \beta + u_i + e_i \quad , \qquad u_i \sim \mathcal{N}(0, \sigma_u^2) \quad , \qquad e_i \sim \mathcal{N}\left(0, \frac{v_e}{n_i}\right) \qquad (1)$$

are observed whenever $n_i > 0$. Here $\beta \in \mathbf{R}^p$ is a vector of unknown fixed-effect coefficients, and u_i, e_i are respectively PSU random effects and sampling errors, independent of each other within and across PSU's. Ordinarily, σ_u^2 is unknown and estimated, while v_e is known. In SAIPE practice as of 1995 (NRC Report 2000a, p. 37), v_e is unknown and the model error from fitting a regression model with the same predictors to previous decennial census data is treated as the known value of σ_u^2.

Small area estimates (SAE's) based on such FH models are statistics designed to estimate with small mean squared error (MSE) the parameters

$$\vartheta_i \, = \, x_i^{\rm tr} \beta \, + \, u_i \quad , \qquad i \, = \, 1, \, \ldots, \, m$$

In the SAIPE log-Rate FH models, y_i is the observed log child-poverty rate for the i'th PSU, with the rate itself defined by exponentiating:

$$\vartheta_i^* \, = \, \exp(\vartheta_i) \, \equiv \, \exp(x_i^{\rm tr} \beta \, + \, u_i) \tag{2}$$

Based upon the methods proposed in Slud (1999, 2000b) of estimating logistic GLMM's via maximization of an accurately calculated log-likelihood, we propose here a logistic GLMM which would make use of all available SAIPE data. If x_i are p-dimensional predictors and $u_i \sim \mathcal{N}(0, \sigma_u^2)$ are random PSU effects as in (1), and if y_i^0 denotes the number of sampled poor related school-age children, then the model assumes

$$y_i^0 \, \sim \, Binom(n_i, \, \pi_i) \quad , \quad \log\left(\frac{\pi_i}{1-\pi_i}\right) \, = \, x_i^{tr} \, b \, + \, u_i \tag{3}$$

with b now the unknown fixed-effect coefficients. The parameters $\pi_i = E(y_i^0/n_i \,|\, u_i)$ are themselves rates, and in that sense comparable to (2).

Remark 1 Note two distinctions between the rate-parameters in the models (1) and (3). First, the numbers n_i in (1) have historically been taken as numbers of sampled *households*, while n_i in (3) should logically be taken as the number of sampled related school-age children. Second, the rate-parameter ϑ_i^* in (2) is not quite the conditional expectation of the empirical child poverty rate: instead, $E(e^{y_i} \,|\, u_i) \, = \, \vartheta_i^* \, e^{v_e/(2n_i)}$. The rationale for comparing ϑ_i^* and π_i is that the same models are viewed as holding if y_i in (1) were replaced by the analogous log-rate Y_i over the *whole* of the PSU, n_i in (3) by the total number N_i of related school-age children, and y_i^0 by $Y_i^0 = N_i \, \exp(Y_i)$. □

2.1 SAE Formulas

In the FH setting, the estimators for ϑ_i based on the data $\{y_j, n_j : n_j > 0, \ 1 \le j \le m\}$ satisfying (1) are the standard EBLUP estimators (*cf.* Prasad and Rao 1990, Ghosh and Rao 1994)

$$\hat{\vartheta}_i \;=\; x_i^{\rm tr}\,\hat{\beta} \,+\, \hat{\gamma}_i\,(y_i - x_i^{\rm tr}\hat{\beta})$$

where $(\hat{\beta}, \hat{\sigma}^2_u)$ or $(\hat{\beta}, \hat{v}_e)$ are the maximum likelihood (ML) estimators in model (1), and $\hat{\gamma}_i = \hat{\sigma}^2_u/(\hat{\sigma}^2_u + v_e/n_i)$ or $\hat{\gamma}_i = \sigma^2_u/(\sigma^2_u + \hat{v}_e/n_i)$, depending on which variance component is being estimated. We adopt the convention that $\hat{\gamma}_i \equiv 0$ (so that $\hat{\vartheta}_i = x_i^{\rm tr}\,\hat{\beta}$) when $n_i = 0$.

Data-analytic results are presented below both with the usual FH model (cited as **LmA**) with v_e assumed known but σ^2_u unknown, and also (cited as **LmB**) with σ^2_u as known and v_e unknown in model (1). In either case, the (complete-data) estimators for the exponentiated small-area parameters $\vartheta^*_i = \exp(\vartheta_i)$ are given by the approximately bias-corrected formula

$$\hat{\vartheta}^*_i \;=\; \exp\left(x_i^{\rm tr}\,\hat{\beta} \,+\, \hat{\gamma}_i\,(y_i - x_i^{\rm tr}\hat{\beta}) + \frac{1}{2}\,\hat{\sigma}^2_u\,(1 - \hat{\gamma}_i)\right) \tag{4}$$

In GLMM analyses, the parameters $(b, \, \sigma^2_u)$ in (3) are estimated via Maximum Likelihood, using the accurate numerical log-likelihood approximations described in Slud (2000b), and the resulting SAE's take the form

$$\hat{\vartheta}^*_i \;=\; \int \frac{e^{(\eta+\sigma z)(y_i+1)}}{(1+e^{(\eta+\sigma z)})^{n_i+1}}\,\phi(z)\,dz \Big/ \int \frac{e^{(\gamma+\sigma z)\,y_i}}{(1+e^{(\gamma+\sigma z)})^{n_i}}\,\phi(z)\,dz \tag{5}$$

where values $\eta = x_i^{tr}\,\hat{\beta}$ and $\sigma = \hat{\sigma}_u$ are substituted; where ϕ is the standard normal density; and where the values $y_i, \, n_i$ are taken to be 0 in PSU's with no sample. These SAE's are as given by Slud (1999).

2.2 Criteria for Model-Comparisons

It is recognized that even in decennial years, census and CPS estimates of school-age child poverty rates differ for nonsampling reasons (2000a NRC Report, pp. 16-27); but the fit of different SAIPE models was generally assessed in the NRC Report (Citro & Kalton 2000a) by 'internal' comparisons with CPS weighted estimates. The corresponding 'external' comparator is the child poverty rate estimated from the most current decennial census, perhaps modified by updated demographic information. It seems from the historical notes on SAIPE in Chapters 1–3 of the NRC Report that adequacy of fit to CPS estimates has now replaced fit to census-based estimates as the primary criterion in determining the best SAIPE model, although the Report uses the pattern of model-based SAE's on regionally and demographically defined subgroups of counties to check the model results. However, the CPS estimates are highly variable in small counties, and to the extent that the internal and external comparators disagree, even in larger counties, there is currently no principled way to decide which should be primary.

We undertake model comparisons in this paper with respect to the 1990 and 1994 SAIPE datasets, based on *loss-functions* which measure discrepancies between model-based SAE's $\hat{\vartheta}^*_i$ and a county rate ϑ^{*0}_i (estimated from CPS or census) serving as standard of truth. The loss functions (*cf.* NRC Report 2000a, Appendix C) are:

$$SSQ \;=\; \sum_i \,(\hat{\vartheta}^*_i - \vartheta^{*0}_i)^2 \;\;, \quad WtSSQ \;=\; \sum_i \,\nu_i\,(\hat{\vartheta}^*_i - \vartheta^{*0}_i)^2$$

$$Abs \;=\; \sum_i \,|\hat{\vartheta}^*_i - \vartheta^{*0}_i| \;\;, \quad WtAbs \;=\; \sum_i \,\nu_i\,|\hat{\vartheta}^*_i - \vartheta^{*0}_i|$$

Here ν_i denotes a within-PSU sample-size indicator: for sampled PSU's, the CPS-estimated number n^*_i (defined below in Section 2.4) of sampled related school-age children in the i'th county; and for non-sampled counties the census population in 1990, or the demographic population estimate in 1994. For sampled PSU's, *WtAbs* is the total number of sampled children whose poverty status would be mis-predicted to be poor by each SAE method. Whenever SAE's are compared to census rates, which unlike CPS sample estimates are never 0, we supplement the other loss-criteria with

$$AbsRel \;=\; \sum_{i=1}^{m} \, |\hat{\vartheta}_i^* / \vartheta_i^{*0} \;-\; 1|$$

For the reasons sketched in the first paragraph of this subsection, we present results separately with *both* choices of truth-standard, internal (CPS) and external (decennial-census), and separately within the CPS sample, for counties which did and did not yield sampled poor related school-age children, and (in 1990 data, with decennial census as truth standard) for counties not sampled by CPS.

In addition, each model implicitly predicts an expected proportion of counties which will yield no poor children among their sample of children. Therefore, we also compare models with respect to the accuracy of agreement of these expected proportions with the observed ones. The calculations under (1) are as follows. In county i, with n_i^* sampled children and log-rate y_i, suppose that the observed rate is replaced by 0 rather than e^{y_i} whenever $n_i \, e^{y_i} < 1$. For fixed parameters $(\beta, \, \sigma_u^2, \, v_e)$ in (1), the expected proportion of counties which will have no sampled poor children is

$$m^{-1} \sum_{i=1}^{m} \Phi(- \frac{\log(n_i) + \beta^{tr} x_i}{\sqrt{\sigma_u^2 + \, v_e/n_i}}) \tag{6}$$

The expected proportion of counties without poor sampled children according to model (3) is

$$m^{-1} \sum_{i=1}^{m} \int (1 + \exp(x_i^{\rm tr} \beta + \sigma_u z))^{-n_i} \, \phi(z) \, dz \tag{7}$$

2.3 Background for comparisons

In order to compare GLMM-based SAE's in SAIPE to the currently used SAE's, Slud (2000, 2002) has previously conducted several simulation experiments. Briefly, those experiments found that the GLMM method had 10–30% lower empirical mean-squared error than the FH method when the true model was GLMM, and even when the true GLMM was assumed to have a logarithmic link — a setting more closely resembling the log-linear mean-rate specification of the standard SAIPE FH model— the MSE of the GLMM estimators was not more than 5–10% worse *and was often better* than that of the FH estimators. Part of the explanation for this was that even with a logarithmic link, the FH model was misspecified when applied only to sampled PSU's with nonzero counts of child poor. However, the simulation experiments showed a small systematic bias of FH-based SAE's, when the log-link GLMM holds. The bias of the SAE's under FH models (LmA or LmB) was roughly linear, as a function of the true rate $\vartheta_i \in [0, .27]$:

$$Bias \;\approx\; .00167 \;-\; .0667\,\vartheta_i$$

Under GLMM's with logarithmic link, the SAE biases gradually increase as a function of response probability ϑ_i up to 0.14, from $-.005$ to .005, and then trend sharply down, as far as $-.04$ or farther, near $\vartheta_i = .27$.

To study the effects in SAIPE of data-truncation due to dropping counties with no sampled poor children from analysis, we performed another set of unpublished simulations on the FH model (1), with true parameters equal to the SAIPE log-rate model parameters below, with values y_i treated as observed only when $n_i e^{y_i} \geq 1$. Then, in each simulation iteration, the parameters (β, σ_u^2) (**LmA** case) or (β, v_e) (**LmB** case) were estimated by ML from the observed data and then substituted into the SAE formulas (4), (6).

We found that the SAE estimators $\hat{\vartheta}_i^*$ had a non-negligible and systematically positive bias with median value 0.07, and the bias had correlation 0.27 with ϑ_i^*. Moreover, whether the analysis was done using the **LmA** or **LmB** model, the (untransformed) SAE biases ranged from $-.02$ to 0.30 (interquartile range was .052–.096 for **LmB** and approximately the same for **LmA**) and showed a definite negative slope with respect to log sample size (i.e., $\log(n_i)$), roughly following the equation $SAEbias \approx .095 - .006 \log(n_i)$. The large size and systematic tendency of the simulated SAE biases with respect to ϑ_i or n_i seem to be due to the large biases (of order 0.1 in coefficients, and 3 or 10% in v_e) in the estimated parameters based on truncated data. However, the FH model-fitting errors due to truncation may not seriously compromise the quality of the associated SAE's (*cf.* Section 3.1).

2.4 Accounting for Between- and Within-PSU Weighting

Since the data for SAIPE estimates are collected with the CPS's sample survey design, the method of analysis should reflect the weights (roughly, the inverse probabilities of sampling) of PSU's and of individuals within PSU's. For the FH models, the two types of weighting enter very differently. First, the differential weighting of individuals within each PSU is already taken into account in the weighted CPS ratio estimator of county child-poverty rate whose logarithm is the response variable for all FH models. Next, the form of the sampling-error term v_e/n_i in the FH model links the error structure to county size, which serves as a surrogate for PSU weight. That is, if the PSU random-effects u_i were known, then the log-likelihood $-\frac{1}{2}\sum_{i=1}^{m}\{\log(v_e/n_i) + n_i(y_i - x_i^{\mathrm{tr}}\beta - u_i)^2/v_e\}$ is the same as for a normal model with error-variance v_e in which the i'th PSU receives weight n_i.

However, the SAIPE log-rate FH model or the FH models fitted below do not otherwise take between-county CPS weights into account.

The definitions of models (1) and (3) immediately lead to different strategies for between-PSU weighting in their ML analysis. First, in the log-likelihood for (1), each county enters with equal weight, the only distinction between counties based on sample-size being made through the sampling-error variance term v_e/n_i. On the other hand, the **Glm** model (3), while judged as an aggregate PSU-level model, has the form of a *unit-level* model, which (due to the identical covariates used over each county) enters n_i log-likelihood terms for county i, thereby giving larger counties greater influence in determining the fixed-effect coefficients and variance components than is done in the FH model MLE's.

In (3), regarded as a unit-level model, n_i would be the number of related school-age children in county i, with y_i^0 the actual sampled number of poor children. However, fitting the model in that form does not allow the within-county weights to be reflected at all. Therefore, in estimating GLMM model parameters with SAIPE data, we take n_i instead to be the CPS weighted ratio estimate n_i^* equal to the county number of children sampled, defined as: the exact count of sampled HU's with population multiplied by the weighted estimate of total CPS school-age children divided by the weighted estimate of total CPS HU's with population. (Our CPS data did not include a separate weighted estimate of the sampled number of related school-age children.) The same type of sampling weight adjustment was previously adopted by Robert, Rao and Kumar (1987) in the case of a fixed effects logistic model. Similarly y_i^0 in the likelihood for (3) is replaced by $n_i^* \, e^{y_i}$. Since our purpose is to compare GLMM and FH analyses, we replace n_i by n_i^* in all FH models, other than the SAIPE under-18 log-rate model itself.

3 SAIPE Data Analyses: FH vs. GLMM

3.1 Preliminary Fitting of log-Rates

For all of the models compared here, the key determinant of model quality is the closeness of the relationship between the predictor variables and the CPS or census (log-rate) response variables. Therefore, we begin by summarizing the results of direct scatter-plotting and linear modelling of these responses in terms of the SAIPE (log-rate) predictors.

(i) The sample correlation between the log of county estimates of child-poverty rates from the 1980 and 1990 decennial censuses is high (0.82) but not extremely so.

(ii) The correlation, in 1990 CPS-sampled counties with poor sampled children, between the logs of the 1990 census-estimated child poverty rate and the CPS weighted-ratio estimated child poverty rate is not high (0.51).

(iii) The SAIPE log-rate predictors of the 1990 decennial census log child poverty rate provide an extremely good fit (correlation 0.96 between fitted and observed values over all 3130 counties). Similarly high correlations persist when the relationship is restricted to CPS-sampled counties or to counties with CPS-sampled poor children.

(iv) The correlation is very high (0.95) between the 1990 decennial census log rate (over CPS counties with poor sampled children) and the corresponding fitted values from the linear regression model for 1990 log CPS rate versus the SAIPE log-rate predictors (which incorporated census data only from 1980). The correlation between fitted values (from the same linear model) and log census rates on the set of all CPS-sampled counties is also 0.95, and the correlation over **all** counties is 0.925, still amazingly high.

(v) In a linear model with 1990 log CPS rate as response, and with the log-rate SAIPE predictors augmented by the logs of denominators (which are the census county population size, census population 5-17, and census county population recalculated using CPS definitions), the coefficients for the latter variables were quite insignificant, and the multiple R^2 has the same value .281 with and without the extra variables.

(vi) A linear model with 1990 log CPS rate as response, with log-rate SAIPE predictors augmented by 1990 log census-rate, leaves the coefficient of the log census-rate very insignificant.

One consequence of (i)-(iv) is that a linear model with SAIPE log-rate predictors already reproduce the log 1990 census rate remarkably well. SAIPE's legislative mandate to provide 'updated estimates' was apparently interpreted to mean that between decennial censuses, CPS direct child-poverty estimates in larger counties were regarded as more reliable than past census rates, and the EBLUP estimates $\hat{\vartheta}_i$ and (4) explicitly average both types of rates. Remark (v) further justifies using a log-rate model instead of the log-count model favored in the NRC Report (2000a). Remark (vi) suggests that the residuals of the log CPS rate from the model with SAIPE log-rate predictors are largely independent of the current census county rates. In the context of SAIPE *state* poverty estimates, Huang and Bell (2002) found that in decennial years a model which equates true census and CPS state poverty rates is best (in AIC sense), while models involving the other SAIPE predictors are best in income years at least 2 years after the previous census. In (vi), we find that a model for 1990 CPS rates can ignore the 1990 census rate if it retains the SAIPE log-rate predictors !

3.2 Exact Specification of Models for Comparison

In our data analyses below, we compare six models, five FH models and the GLMM model (3), ML analysis of which is labeled as **Glm**. All six models are aggregated at county level, with FH responses e^{y_i} and GLMM responses y_i^0/n_i^* equal to the CPS weighted ratio estimate of child poverty rate. The SAIPE under-18 log-Rate model, labeled **Saipe**, is the same as model (d) in the NRC Report (2000a): model (1) with y_i the logarithm of the CPS weighted ratio estimate of child poverty rate, n_i the total number of sampled HU's with population, and $\sigma_u^2 = 0.0140$ taken to be known from a separate regression using data from the previous decennial census. In all other models, the sample-size n_i is replaced by the quantity n_i^* defined in the previous paragraph.

Five other FH models (1) are considered, all but the last of which have the same four predictors as **Saipe**. The first, labeled **FH.LmB**, is refitted by ML, with $\sigma^2 = .014$ fixed as though known to have exactly the same value as for **Saipe**, and v_e is estimated. Second, the model **FH.LmA** fixes the v_e estimate from **FH.LmB** as though *that* were known, and re-estimates parameters β and σ_u^2 in (1) by ML. Third, the model **FH.GB** fixes the σ_u^2 estimated in **Glm** below as though known, and fits v_e as unknown variance-parameter in (1). Next, we fit a joint ML model (1), labeled **FH.jt**, in which both σ_u^2, v_e are estimated as unknown parameters. (This can be done with existing software by alternately and repeatedly fixing v_e and estimating a FH model of **LmA** form, then fixing the estimated σ_u^2 and estimating v_e as unknown in the FH model of **LmB** form, until convergence.) Finally, the predictors from both the SAIPE log-Count model and the log-Rate model are nested in those for a larger model with the

SAIPE log-*Count* predictors plus three additional predictors equal to the logarithms respectively of (demographically updated Census estimates of) the total under-18 county population and the total county population, along with the log 5-17 population according to CPS definitions, calculated from the previous decennial census. The re-fitted FH model, with seven predictors plus intercept and *both* σ_u^2, v_e treated as unknown, is labelled **FH.jt2**.

3.3 Internal Evidence from 1994 & 1990 SAIPE Data

We begin our summary of the 1994 SAIPE data-analysis by showing in Table 1 the estimators of fixed-effect coefficients and variance components for all of the 4-predictor log-rate models described above. Results for the 7-predictor **FH.jt2** model are not displayed, because the coefficients of the first four predictors are all within 0.02 of the corresponding **FH.jt** values; the coefficients of the last three log-size predictors in **FH.jt2** are non-significant, and the log-likelihood for **FH.jt2** is only 1.7 greater than for **FH.jt**, a deviance of 3.4 which is moderate for a χ_3^2 likelihood-ratio test statistic. The log-sizes are similarly non-significant in the 1990 Saipe data. Thus, in the best-specified FH models, the log-Count is no better than versus the log-Rate form of the model. Therefore, we do not consider **FH.jt2** further.

`[Table 1 here.]`

The log-likelihoods in Table 1 all correspond to FH models with the same response variable and predictors, but cannot be meaningfully compared to the **Glm** log-likelihood. In any case, likelihood is a criterion of model-fit and is quite distinct from model-comparisons based on SAE behavior. Of the five FH models, **Saipe** is clearly the worst fitting. **FH.LmB** is just about as bad, and is not considered further. **FH.LmA** is better than **Saipe** with respect to log-likelihood and somewhat better but very similar under most of the loss-functions calculated below. Therefore **FH.LmA** can effectively serve as a surrogate for **Saipe** in all Tables after Table 2. The FH models show a wide variety of (σ_u^2, v_e) parameter combinations, connected by the remark that σ_u^2 and v_e generally vary inversely. However, since **Saipe** or **FH.LmA** have by far the smallest σ_u^2 values, and therefore give greatest weight to the predictors rather than the direct estimates y_i in (4), it is to be expected in light of Section 3.1 that SAE's from these models would agree particularly closely in 1990 to census rates.

Examination of the fit of the **Saipe** 1994 model reveals the rather different behavior of residuals for child-poverty rates for CPS-sampled counties with and without child poor. (Residuals are calculated in terms of rates, not log-rates, so for sampled counties without child poor, the residuals are the expressions (4)). This applies, unsurprisingly, to all five FH models, which are fitted using only the data from counties with sampled child poor. But there is a systematic bias between the fitted child poverty rates in the **Glm** analysis and those provided in the truncated-data FH models, clearly distinguishing those sampled counties with and without sampled child poor. First, in Figure 1 all estimated rates are necessarily over-estimates in counties with no sampled child-poor, but it was not easily predictable that relatively few counties with child poor would have estimated rates which are too large. (The same pattern obtains in the 1990 SAIPE dataset.) A more meaningful distinction between predictability of child poverty rates by county is given by Figure 2. There we plot versus the 1990 Census child-poverty rate the county-by-county differences between the *logit* of estimated 1994 child-poverty rates by **Glm** and by **FH.LmA**. The y-ordinate for the i'th county is $\log((\hat{\vartheta}_i^*)^{Glm}/(1-(\hat{\vartheta}_i^*)^{Glm})) - \log((\hat{\vartheta}_i^*)^{FH.LmA}/(1-(\hat{\vartheta}_i^*)^{FH.LmA}))$. Figure 2 shows that there are systematic differences between the way in which the **Glm** and **FH.LmA** models predict within sampled counties which have no sampled poor (related, school-age) children as compared with those that did have sampled poor children. While the fitted **Glm** and **FH.LmA** child-poverty rates had average logits respectively of -1.51 and -1.57 among counties with poor children (slightly greater than the counties' average logit child-poverty rate of -1.65 found in the 1990 decennial census), the **Glm** had average logit rate of -2.36 (much lower than the 1990 census value of -1.97) among the counties which in 1993-5 had CPS sample but no sampled poor children, and the **FH.LmA** average logit estimated rate had the much higher value of -1.87 (not far from the **FH.LmA** and **Glm** average logit rate on counties with sampled poor children). For application of model predictions to non-sampled counties, it is much more plausible to use a method which automatically gives lower child-poverty-rate predictions to the sampled counties without sampled poor children. On this account, we would expect **Glm** SAE's to outperform those from **FH** models on the non-sampled counties.

[Figure 1 about here.]

[Figure 2 about here.]

The figures analogous to Figures 1 and 2, based on other FH fitting-methods in place of **FH.LmA** or on the 1990 SAIPE dataset in place of the 1994 dataset, look very similar to these, and are not shown. A further indication of the same behavior cited above, is the scatter-plot in Figure 3 of logit **Glm** estimated child-poverty rates versus the logit **FH.LmA** estimated rates for CPS-sampled (1988-1990) counties, both fitted on the 1990 SAIPE data. In this picture, the counties without sampled poor children are plotted as solid diamonds, and those with poor children are plotted as hollow diamonds. The estimated rates under **Glm** tend to exceed those under **FH.LmA** for counties with no sampled poor children, while the estimated **FH.LmA** rates tend to be larger in counties with poor sampled children.

[Figure 3 about here.]

[Tables 2 and 3 about here.]

Tables 2 and 3 respectively show the loss-values for the 1994 and 1990 SAIPE datasets, for the internal (CPS) standard of truth. (For comparison, we show also in Table 3 the loss-function values arising from 'estimating' the CPS rates by the census 1990 rates.) The immediate conclusion, by all loss-criteria, is that the **Glm** method performs better overall than the FH methods, and indeed, clearly outperforms all FH methods also on the sets of counties with *and* without child poor, except that **FH.GB** gives the best results among all methods on the counties with sampled poor children. Among the other FH methods, **Saipe** and **FH.LmA** are very similar, with **FH.LmA** usually the better of the two, except that on counties with no sampled poor children the very slight differences are in favor of **Saipe**. This pattern persists throughout: we do not continue to tabulate separate results for **Saipe**, since they are extremely close to those for **FH.LmA**.

Among the $m = 1488$ counties with CPS-sampled related school-age children in 1993-95, there are no poor children in 304, or 20.4%. With the fitted parameters of model **Saipe** substituted, the expected proportion (6) is calculated to have the much smaller value 0.089. By contrast, models **FH.GB**, **FH.LmA**, and **FH.jt** respectively yield values 0.153, 0.162, 0.147 for the proportion (6), still too small but not outrageously so. This can already be interpreted as a serious and somewhat unanticipated *modeling* deficiency of

Saipe. The calculated estimated proportion (7) of counties with no poor sampled children from **Glm** is 0.142. From this comparison, we see that none of the models considered can accurately reproduce the proportion of sampled counties with no sampled child poor: **FH.LmA** and **Glm** models underestimate slightly, and **Saipe** badly. A similar pattern is repeated for the 1990 SAIPE data, in which $0.183 = 231/1259$ is the true proportion of all CPS-sampled counties with no sampled poor children, while the respective model predictions are: for **FH.jt** 0.111, for **FH.LmA** 0.131, for **FH.LmB** 0.127, for **Saipe** 0.068, for **FH.GB** 0.112, and for **Glm** 0.113.

3.4 Model-based estimators versus 1990 Census estimates

We turn now to external comparisons between the SAE's, with reference to the 1990 SAIPE dataset with the 1990 decennial census county child poverty rates as standard of truth. Consider first the counties in the 1990 SAIPE dataset, i.e., those sampled by CPS from 1988-1990. Table 4 displays five loss-function values for each of five SAE methods, for the 1990 census rates as standard. Now **Glm** performs better than **FH.GB** in counties with sampled poor children, and worse in counties without poor children, a pattern opposite to that in Table 3. The **Saipe**-surrogate method **FH.LmA** is now clearly the best-performing method, by all loss-criteria. **Glm** and **FH.jt** are substantially equivalent (not shown in the Table).

`[Tables 4 and 5 here.]`

Table 5 summarizes the performance of the various SAE methods on the counties which were not sampled by CPS in the period 1988-90, using the 1990 census as standard. Here as in Table 4, **FH.LmA** is the best of the FH methods. But now, **Glm** shows a clear advantage over **FH.LmA**: 25% by the SSQ and WtSSQ loss-criteria, and 8–12% by the other criteria.

As mentioned in Section 3.1, the excellent performance of **Saipe** and **FH.LmA** SAE's with respect to a decennial-census standard is an expected consequence of the extremely high correlation in the decennial year 1990 between the census-estimated child-poverty rate and the linear predictor of the log CPS rate in terms of SAIPE variables. In Table 1, the **Saipe** and **FH.LmA** models have the smallest σ_u^2 by far, so that their SAE's display the greatest reliance on the fixed-effect predictors; and at the same time, these models have the largest coefficient for the predictor LCPRT, the log child-poverty rate derived from the previous decennial census which is the variable most closely correlated with the 1990 decennial-census rate.

3.5 Which Criterion is Primary for SAIPE ?

The comparison between the FH and **Glm** model performance in SAIPE is fairly clear. The GLMM does a much better job of fitting to CPS rate data, and the quality of its SAE's in reproducing census rates is also good. By contrast, although the SAIPE log-rate models fit the CPS data inadequately by several criteria, their associated SAE's reproduce the decennial census rates in decennial income years remarkably well. What are the implications of these contrasting findings for SAIPE ?

SAIPE's legislative mandate seems to require updated estimates between decennial years which differ from decennial census rate-estimates whenever current direct CPS estimates are reliable but disagree with census estimates. In other words, CPS estimates in larger counties in mid-decade must be allowed (collectively, through a model) to perturb estimates primarily based on the past census. By the same logic, a modeled pattern in the current CPS direct estimates in smaller or moderate-sized counties which reflects a change from the past decennial-census values should also be allowed to affect SAE's.

However, there are at present no reliable 'ground-truth' measurements of county child-poverty rates taken between decennial censuses.

With this logic in mind, we compared the performance of the same models and SAE's as in Tables 4 and 5 upon the smaller set of CPS-sampled counties with estimated population size at least 11000. These are the counties with above-median population: they have a median of 38 CPS-sampled school-age related children. The SAE comparisons for these counties are very similar to those displayed in the Tables. Thus, even in counties with large enough size (and usually, sample of HU's with related school-age children) to make the direct estimate informative about the CPS log-rate, there is an excellent fit by **Glm**, and **Glm** also does well with respect to the census standard, although not nearly as well as **FH.LmA** or **Saipe**. On unsampled large counties, **Glm** does *exactly* as well as Saipe, which is satisfactory.

4 Conclusions and Recommendations

The main conclusions of this research are as follows:

(A) In the SAIPE setting, SAE biases within the Fay-Herriot log-rate model due to analysis with truncated datasets are small but systematic. While the magnitudes of errors are not conclusively larger with FH than with GLMM models, their distribution is noticeably different on CPS-sampled than on non-sampled counties.

(B) The SAE's based on the GLMM method **Glm** outperform all FH competitors with respect to internal (CPS) loss-function criteria, on the sets of all CPS-sampled counties, those with sampled child-poor, and those without child poor. The only exception is the method **FH.GB**, in which a large county random-effect is chosen to resemble that of **Glm**: this method slightly outperforms **Glm** on the counties with child poor, but is much worse than **Glm** in counties without child poor.

(C) The **Saipe** or **FH.LmA** SAE methods agree much better with the external census standard than other methods, including **Glm**, in CPS-sampled counties, both in counties with and without sampled poor children. However, in counties which were *not* sampled by CPS, **Glm** agrees more closely with census rates than the FH methods studied.

(D) Of all the SAE methods compared, the only ones which clearly outperform **Glm** by some criteria (**FH.GB** with respect to CPS standard on sampled counties with sampled poor children, and **FH.LmA** with respect to census standard on CPS-sampled counties) are resoundingly worse by other criteria. The only one of the FH methods which is often good and never terrible is the one based on **FH.jt** (jointly maximized with respect to both variance-component parameters). But it is dominated by **Glm**. Thus, our recommended method is **Glm**.

(E) The current **Saipe** method (log-rate, under-18 form) in essence produces conditional estimates for child poverty given at least one sampled poor child in a PSU. It performs extremely well with respect to the decennial census standard, but by log-likelihood and loss-function measures, does not fit SAE's to the CPS data well. Its poorer fit to CPS is not attributable to differences from the log-count FH models judged best in the NAS Panel reports, and is corroborated by its very inaccurate model-based predictions (from numbers of CPS-sampled related school-age children) of numbers of counties without sampled poor children. **Saipe**'s good fit to the census and poor fit to CPS rates are both largely due to the very small imposed (not fitted) PSU-effect variance σ_u^2.

(F) Since the **Glm** method performs just as well, by comparison with **Saipe** and **FH.LmA**, when restricted to larger counties, the SAIPE mandate to reflect current 'updated' (CPS and IRS) data in producing current Small Area Estimates appears to be more adequately fulfilled by **Glm** than by the Fay-Herriot methods.

If a GLMM approach to county-level SAE's were adopted in SAIPE, the county-level results would be raked to state totals, since the current linear-model approach will likely continue to be used at the state level.

5 REFERENCES

Citro, C. and Kalton, G., eds. (2000a) **Small-Area Estimates of School-Age Children in Poverty**, *Evaluation of Current Methodology* (National Research Council), Washington DC: Nat. Acad. Press.

Citro, C. and Kalton, G., eds. (2000b) **Small-Area Estimates of School-Age Children in Poverty**, *Priorities for 2000 and Beyond* (National Research Council), Washington DC: Nat. Acad. Press.

Fay, R. and Herriot, R. (1979) Estimates of income for small places: an application of James-Stein procedures to census data. *JASA* **74**, 341-53.

Ghosh, M., and Rao, J.N.K. (1994) Small Area Estimation: an appraisal. *Statist. Sci.* **9**, 55-93.

Huang, E., and Bell, W. (2002) Comparing alternative models for using decennial census data in SAIPE state poverty estimates. *ASA Proc. Surv. Res. Meth.*, to appear.

Prasad, N. and Rao, J.N.K (1990) The estimation of mean-squared errors of small-area estimators. *Jour. Amer. Statist. Asssoc.* **85**, 163-171.

Robert, G., Rao, J.N.K., and Kumar, S. (1987). Logistic regression analysis of sample survey data. *Biometrika* **74**, 1-12.

Slud, E. (1999) Models for simulation and comparison of SAIPE analyses. Census Bureau preprint, posted at SAIPE web-site (see below).

Slud, E. (2000) Comparison of aggregate versus unit-level models for small-area estimation. *ASA Proc. Surv. Res. Meth.*, 563-8.

Slud, E. (2000b) Accurate calculation and maximization of log-likelihood for mixed logistic regression. Census Bureau preprint, posted at SAIPE web-site: **ftp.census.gov/hhes/www/saipe/tecrep.html**

Slud, E. (2002) Simulation Studies of Bias in Small-Area Point and MSE Estimates in SAIPE. Census Bureau preprint.

Table 1: Coefficient and variance estimates in 1994 SAIPE dataset for 6 models discussed in the text. The model labeled **Saipe** is the log-rate under-18 model fitted to the '94 SAIPE data. Its coefficients, given to only two decimal places for reasons of nondisclosure, are as in the NRC report except for an unexplained discrepancy of 0.02 in the LFILRT coefficient.

Param.	Glm	FH.GB	FH.LmA	FH.LmB	Saipe	FH.jt
Intercept	.726	-.104	-.014	.051	.04	-.115
LTAXRT	.390	.394	.294	.277	.26	.362
LSTMPRT	.406	.210	.288	.293	.30	.225
LFILRT	-.318	-.359	-.343	-.390	-.44	-.334
LCPRT	.441	.294	.349	.378	.37	.313
σ_u^2	.550	.550	.071	.014	.014	.309
v_e	.000	.755	17.27	17.27	33.625	3.306
logLik	*	-1226.8	-1249.2	-1268.4	-1272.8	-1193.1

Table 2: Loss-criterion values for 4 SAE methods versus CPS (weighted sampled-based) estimates, from 1994 SAIPE data, for all 1488 sampled counties, for all 1184 counties with sampled poor children, and for all 304 counties without sampled poor children.

	Glm	**FH.GB**	**Saipe**	**FH.LmA**
All counties				
SSQ	15.850	18.023	41.346	38.830
WtSSQ	91.124	158.55	722.64	566.02
WtAbs	1221.5	855.29	5219.4	4111.5
Abs	79.958	75.371	165.54	155.248
With child poor				
SSQ	11.908	3.954	33.065	30.533
WtSSQ	66.078	9.029	636.01	478.95
WtAbs	948.98	155.52	4695.3	3583.9
Abs	49.784	15.937	120.75	110.22
No child poor				
SSQ	3.942	14.069	8.280	8.297
WtSSQ	25.046	149.52	86.632	87.073
WtAbs	272.48	699.77	524.10	527.60
Abs	30.174	59.435	44.791	45.031

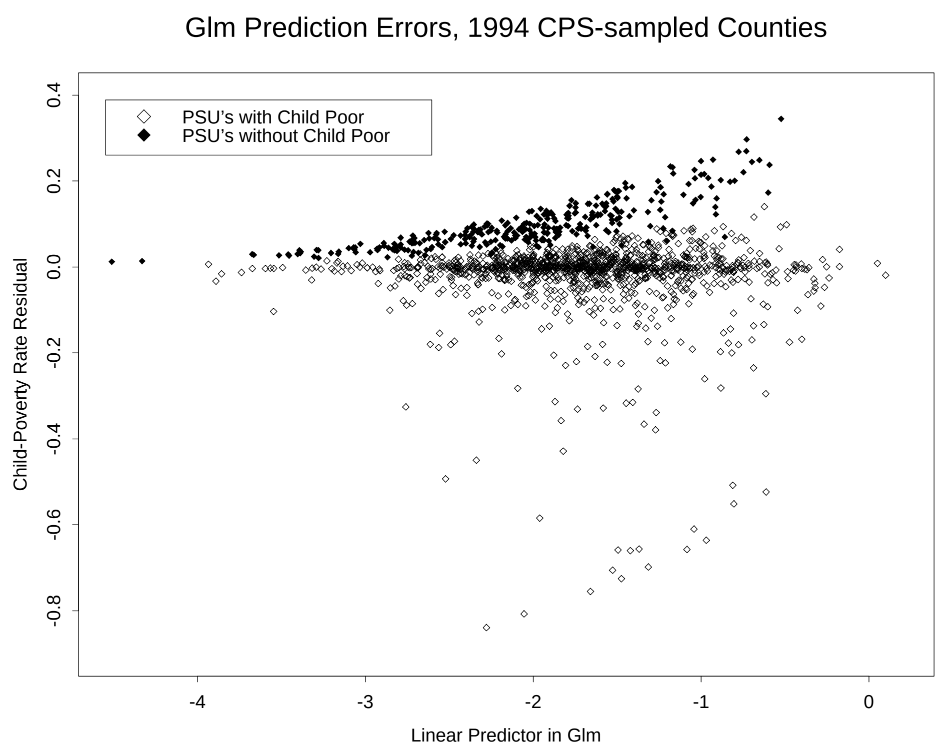

Figure 1: Scatterplot of residuals between SAE's and CPS-estimated child poverty rates based on 1994 SAIPE county data, where counties *with* sampled poor children are plotted with hollow triangles, and counties *without* sampled poor children are plotted with solid triangles.

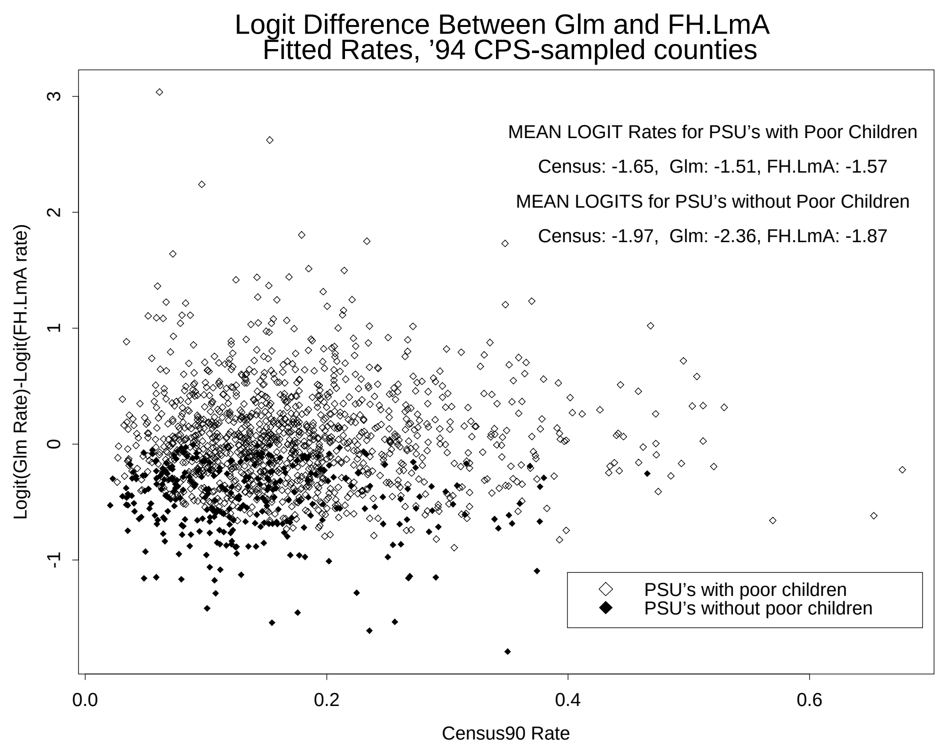

Figure 2: Scatterplot of differences between SAE's of 1994 child-poverty rates from **Glm** and from **FH.LmA**, plotted for each county versus county child-poverty rate from the 1990 census. Counties *with* sampled poor children are plotted with hollow triangles, and counties *without* sampled poor children are plotted with solid triangles.

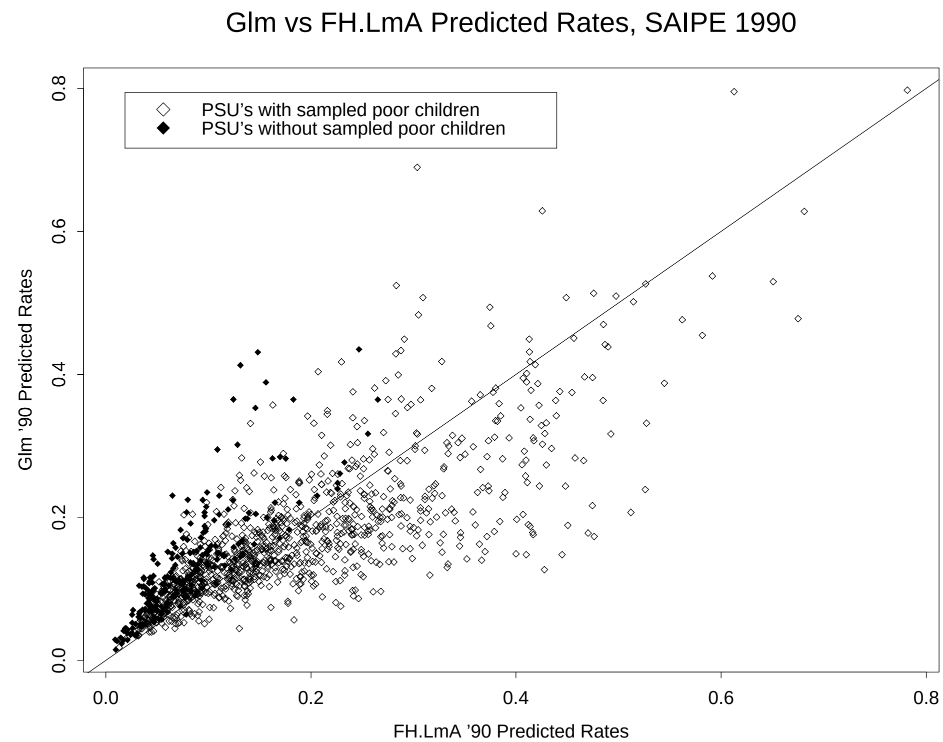

Figure 3: Scatterplot of Child Poverty-rate Predictors for **Glm** versus **FH.LmA** based on 1990 SAIPE dataset. Counties *with* sampled poor children are plotted with hollow triangles, and counties *without* sampled poor children are plotted with solid triangles.

Table 3: Loss-criterion values for 3 SAE methods and Census 1990 versus CPS (weighted sampled-based) estimates, from 1990 SAIPE data, for all 1259 sampled counties, for all 1028 counties with sampled poor children, and for all 231 counties without sampled poor children.

	Glm	**FH.GB**	**FH.LmA**	**Census**
All counties				
SSQ	5.641	8.923	20.751	23.03
WtSSQ	58.327	117.65	491.89	665.62
WtAbs	1108.5	757.90	4212.6	5462.9
Abs	47.722	45.565	112.37	123.67
With child poor				
SSQ	3.655	0.785	15.396	18.08
WtSSQ	40.670	3.498	419.98	598.21
WtAbs	872.02	123.80	3730.5	4994.2
Abs	29.460	6.743	82.096	94.47
No child poor				
SSQ	1.985	8.138	5.355	4.95
WtSSQ	17.657	114.15	71.909	67.41
WtAbs	236.47	634.10	482.06	468.7
Abs	18.262	38.822	30.272	29.20

Table 4: Loss-criterion values for 3 SAE methods, from 1990 SAIPE data, with respect to Census 1990 estimates: for all 1259 sampled counties, for the 1028 counties with and the 231 counties without sampled poor children.

	Glm	**FH.GB**	**FH.LmA**
All counties			
SSQ	8.630	14.648	1.392
WtSSQ	409.06	562.09	74.139
WtAbs	4397.4	5046.9	1833.3
Abs	77.649	98.005	28.820
AbsRel	509.50	686.74	191.75
With child poor			
SSQ	7.650	14.004	1.126
WtSSQ	389.62	552.96	71.04
WtAbs	4163.5	4878.2	1765.1
Abs	66.164	87.985	23.944
AbsRel	422.96	584.38	149.67
No child poor			
SSQ	0.980	0.644	0.266
WtSSQ	19.430	9.132	3.104
WtAbs	233.89	168.75	68.215
Abs	11.485	10.019	4.876
AbsRel	86.530	102.36	42.077

Table 5: Loss-criterion values for 3 SAE methods, fitted from 1990 SAIPE data, with respect to Census 1990 estimates for all 1870 counties not sampled by CPS in 1988–90.

	Glm	**FH.GB**	**FH.LmA**
SSQ	3.256	7.898	4.498
WtSSQ	7683.2	24003.1	10220.7
WtAbs	183377	368782	201046
Abs	55.502	98.295	62.039
AbsRel	310.31	589.08	331.95