SMALL AREA ESTIMATION ERRORS IN SAIPE USING GLM VERSUS FH MODELS

Eric V. Slud, Univ. of Maryland & Census Bureau
Mathematics Department, University of Maryland, College Park MD 20742

Key words: *AIC, Fay-Herriot model, generalized linear mixed model, internal and external standards, loss function, small area estimation.*

Abstract. The Small Area Income and Poverty Estimation (SAIPE) Program produces county-level estimates of child-poverty rates and counts of poor school-aged children, using census, CPS Annual Social and Economic Supplement (ASEC), and administrative-records data. The author has for several years studied the discrepancies, from both the 'internal' standard of CPS direct estimates and the 'external' standard of census-based estimates, of small-area estimators in this setting, based upon competing versions of the aggregate-level Fay-Herriot (FH) model used in SAIPE production or based upon Generalized Linear Mixed Model (GLMM) models. Slud (2003) compared the fit of various FH models tracked over multiple years of SAIPE data since the program's original model choices and evaluations were made (Citro and Kalton 2000), and found qualitative changes in CPS vs. census data relationships between 1990 and 2000. This paper presents an overall summary using multi-year data of the relative performance of the FH vs. GLMM county child poverty rate estimators, finding the quality of fit to the CPS versus census standard to be largely controlled by the county random effect variance parameter.

Acknowledgment. The author is grateful to Bill Bell, Don Malec, and Jerry Maples for helpful comments.

1. INTRODUCTION

Under the terms of Title I of the No Child Left Behind Act, more than $14 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 and Kalton 2000, p. 3) calling for the Census Bureau to supply 'updated estimates' of county-level child poverty for use in allocations to counties in 1997-98 and 1998-99, and thereafter to provide estimates at school-district level. Previously, 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) Annual Social and Economic Supplement (ASEC) and administrative-records data from IRS (income tax returns) and the Food Stamp program.

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 CPS samples and for the smallness of samples in many other counties. To augment sample sizes, SAIPE aggregates CPS data by county over three years, combining the samples for the year before and after with data for the year of interest in estimating numbers of poor and total school-age children related to sampled householders. Since the outcome of modeling is to estimate rates of poverty among of related poor school-age children in *all* counties, SAIPE may use only administrative-record predictor variables which are available in appropriately (weighted and) aggregated form for all counties nationally. The useful variables meeting these constraints have been found to be 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.

Since the SAIPE county model was designed to update the decennial-census survey estimates of child poverty using the much smaller CPS, and since the CPS and census estimates measure slightly different things, there is an unremovable ambiguity about the proper criterion of accuracy for the SAIPE model. The model's originators agree that the model aims to reflect the population quantities estimated by the CPS. Thus the primary standard of truth internal to every year's SAIPE dataset is the CPS direct weighted county estimate of number of poor related school-age children or of the rate of poverty among school-age related children, which are approximately unbiased estimators of the CPS estimand. In decennial-census years, the best available external standard of truth is provided by the child-poor or child-poverty-rate estimate from census data adjusted as closely as possible to the CPS-defined universe of school-age related children. The resulting 'census child-poverty rate' appears in following years, after log-transformation, as a predictor variable in the SAIPE (log-rate) model, and in decennial years, as an approximation to the CPS estimand.

The NRC Report (Citro and Kalton 2000) assessed competing SAIPE models based on 1990 and 1994 data. Later, Maiti and Slud (2002) reviewed these Fay-Herriot (1979) models, versus others of FH and of GLMM type, using the same years' data with a loss-function approach similar to that used in Appendix C of the the NRC Report. Slud (2003) reviewed several more FH models and weighted log-linear regression models, using multi-year SAIPE data with emphasis on the decennial years 1990 versus 2000. One model consistently among the best, although not the official SAIPE log-count model, had its county-level response variable equal to the logarithm of the ratio of the CPS weighted estimates of poor school-age related children over school-age related children, and had log-rate predictors (`LTAXRT, LSTMPRT, LFILRT, LCPRT` described below) closely related to those of the SAIPE log-count model. We refer to this model throughout the present paper as *the SAIPE log-rate model*.

In Slud (2003), the weighted linear regression model predictors for log child poverty rate in terms of the SAIPE log-rate model variables were shown (like the SAIPE log-rate model predictors) to be correlated in the range .92–.95 with census-based log child-poverty rates. However, the linear relationship between the CPS log-rate predictors and the census log-rates was shown to have changed between 1990, when its intercept was very close to 0 and its slope to 1, and 2000, when the log CPS rate versus log census rate had slope clearly less than 1 and intercept clearly greater than 0.

In this paper, we compare the fit of FH and random-intercept generalized-linear mixed (GLM) models for CPS log-rate with both the internal CPS and external census standard, for all available years of SAIPE data: 1990, 1994, 1996, 1998, 1999, and 2000. The GLMM's studied are logistic-regression models, except for a few comparisons with log-link models. The FH-type models are found to provide a remarkably close fit to the external census standard, while the GLM models fit more closely to the CPS data. The models and criteria for comparison are introduced in the next Section; and in subsequent sections, various FH models are assessed over the SAIPE model years (Sec. 3); next the GLM models are similarly assessed among themselves (Sec. 4); and then the best performers from the two classes of models are contrasted (Sec. 5). Conclusions are drawn in the final Section 6.

2. OUTLINE OF MODELS & CRITERIA FOR COMPARISON

We treat two broad classes of models. First, we consider Fay-Herriot (1979) or FH, models:

$$\log y_{it} = \beta_t^{tr}\, X_{it} \,+\, e_{it}^{tr}\, Z_{it} \,+\, u_{it} \qquad (1)$$

where y_{it} is the response (i.e., the direct weighted CPS estimate for child poverty rate, based on a 3-year aggregate centered on the survey year) in county i and year t; X_{it} are observed predictor variables; Z_{it} are observed variables (usually $1/\sqrt{n_{it}}$ where n_{it} is the number of households sampled by CPS) related to CPS sample size and selection probabilities in county i and year t; and u_{it}, e_{it} are unobserved independent normally distributed mean-0 errors, independent and identically distributed (*iid*) across i for each t, respectively associated with county-level random differences and with scaled sampling errors. The parameters β_t, $\mathrm{var}(u_{it}) = \sigma^2_{u,t}$, and $\mathrm{var}(e_{it}) = v_{e,t}$ are unknown and estimated separately for distinct t, by maximum likelihood. The sampling error terms e_{it} are dependent across t due to 3-year aggregation, and the possible dependence among u_{it} for fixed i is also of great interest.

The variables X_{it} are the SAIPE predictor variables by county for year t. We name and describe them, along with the numerators and denominators of the underlying rates, for later reference. The basic predictors for the SAIPE log-rate model are

```
LTAXRT  =  log(PoorXMP/ChXMPT)
   = log IRS-estimated child poverty rate

LSTMPRT  =  log(FSTMP/RPOPest)
   = log Food-Stamp participation rate

LFILRT  =  log(ChXMPT/RChPOP)
   = log IRS child tax exemption rate
```

```
LCPRT  =  log(CenPOOR/CPSuniv)
   = log poverty rate, residents age 5-17
```

where ChXMPT, PoorXMP are respectively the IRS numbers of child and poor-child exemptions for the survey year (which is the income year plus 1), RChPOP denotes the resident child population derived from the previous decennial census, FSTMP an adjusted number of food stamp recipients in July of the income year, RPOPest the total resident county population (from updated census and demographic estimates), and CenPOOR the number of poor children and CPSuniv the 'child poverty universe' (approximate number of children according to CPS definitions) from the previous decennial census. In one model, we augment the SAIPE log-rate predictor variables by the negative logarithm of

PrbSel = CPS county selection probability.

The second class of models we consider is the Generalized Linear Mixed Model specified by

$$n_{it}\, y_{it} \sim \text{Binom}(n_{it},\, g(\beta_t^{*tr}\, X_{it} \,+\, u_{it}^*)) \qquad (2)$$

where $g(x)$ is a known smooth increasing function (the *inverse link*, and X_{it}, n_{it} are as in (1), and β_t^* along with the variance $(\sigma^*)^2$ of the (mean-0, normally distributed) random PSU-effect u_{it}^* are the unknown parameters fitted by maximum likelihood to (2). (In fact, since the responses $n_{it}y_{it}$ are not integers because y_{it} are the CPS weighted estimates of rates, it would be more correct to say that our modeling approach uses and maximizes the same likelihood expression as (2)). In the most important and usual special case of (2), the function $g(x) = e^x/(1+e^x)$ is the logistic distribution function, yielding the *random-intercept logistic model*,

$$n_{it}\, y_{it} \sim \text{Binom}\Big(n_{it},\, \frac{\exp(\beta_t^{*tr}\, X_{it} \,+\, u_{it}^*)}{1+\exp(\beta_t^{*tr}\, X_{it} \,+\, u_{it}^*)}\Big) \quad (3)$$

Formulas for small-area point estimators of the county-level parameters (respectively $E(y_{it}\,|\,u_{it})$ in (1) and $E(y_{it}\,|\,u_{it}^*)$ in (3)) can be found respectively in Citro and Kalton (2000) and in Maiti and Slud (2002) or Slud (2003).

In what follows, we compare the fit of various models of the type (1) and (2), to SAIPE data for the years 1990, 1994, 1996, 1998, 1999, and 2000, using several measures. We compare fit to CPS within model classes using the AIC (penalized maximized log-likelihood); fit to the most recent decennial census by correlating model small-area estimators with the log census rates; and fit to both standards of truth, broken down according to whether counties were in the CPS sample and if so whether there were sampled poor related children, using various 'loss-functions' measuring discrepancies between predicted and truth-standard child-poverty rates, either absolute or squared, differences or relative differences, unweighted or weighted by the demographically updated estimate of number of county school-age children. The various loss functions were calculated and gave highly consistent results; therefore we present results (Tables 7–9 below) only for the case of weighted sum of absolute differences.

Table 1: Variant FH models, by Name and number p of non-constant predictors; denominator den of sampling-error variance ($1/Z_i^2$ in (1)). Variance components σ_u^2, v_e are indicated as estimated (*ML*) or fixed. Each * entry is the same as the one immediately above, and # in columns 5-7 specifies the model used to fix a parameter or predictors.

#	Name	p	den	σ_u^2	v_e	Predictors
1	FHsaip	4	n_i	.015	*ML*	SAIPE log-rate
2	FHwtrg	4	*	#1	0	*
3	FHspA	4	*	*ML*	#1	*
4	FHspB	4	*	#3	*ML*	*
5	FHspO	4	*	*ML*	*ML*	*
6	FHrt	4	$\sqrt{n_i}$	#1	*ML*	*
7	FHsp1	5	$\sqrt{n_i}$	#1	*ML*	#1, plus log(RPOPest)
8	FHsp4	8	n_i	#1	*ML*	#7, plus log of ChXMPT, RChPOP and CenPrCPS
9	FHsp4B	8	*	#3	*ML*	#8
10	FHsp1C	5	*	#1	*ML*	#1, plus -log(PrbSel)
11	FHsp2	6	*	#1	*ML*	#7, plus $\log^2$(RPOPest)
12	FHsp2B	6	*	#1	*ML*	#11
13	FHrt2	6	$\sqrt{n_i}$	#1	*ML*	#11

3. COMPARISONS ACROSS YEARS AND FH MODELS

We begin by exploring a number of variant FH models, displayed in Table 1. These are named and further specified in terms of their predictor variables, of whether the Z_i^2 variables in (1) are $1/n_i$ or $1/\sqrt{n_i}$ (the latter idea due to Fisher and Asher 1999, 2000), and whether the unknown variance component is σ_u^2 or v_e . The values $\sigma_u^2 = \sigma_{u,t}^2$ fixed in FHsaip for the survey years 1990, '94, '96, '98, '99, 2000 were: $.0137, .014, .014, .015, .014, .016$. These are indicated in Table 1 as being all roughly .015.

It seems clear from the AIC model-comparisons in Table 2 that FHsaip fits the data slightly better than, but very similarly to, FHwtrg; that

FHspB, FHspO with their σ_u^2 values progressively larger (respectively .06–.08 and .24-.33 as compared to .015 for FHsaip) fit the CPS data better, with FHspO almost the best-fitting model. Augmented models which include the log-denominators of the SAIPE log-rate model predictors give only small improvements over the FHsaip model, but the models FHsp2 and FHsp2B that augment by log(RPOPest), $\log^2$(RPOPest) (variables found in the data analysis of Slud 2003) do represent a noticeable improvement. The best models, from the vantage point of AIC, are the FHrt models which include sampling-error variance terms of the form $v_e/\sqrt{n_i}$. Models like this for log-counts instead of log-rates, advocated by Robin Fisher, are currently used in SAIPE production estimates.

Another way to quantify the behavior of the various FH models is to correlate their county small-area predictors for child poverty rates with the corresponding rates from the most recent decennial census. Table 3 shows that while small-area estimates (SAE's) from almost all of the FH models maintain remarkably high correlations with census rates, the model FHspO which had very favorable AIC (a purely internal CPS measure of fit) and high correlations with census rates on non-CPS counties, has poor correlations with census rates on CPS counties. Among other models, FHrt shows remarkably good correlations with census rates on both CPS and non-CPS counties, in addition to its fine performance with respect to AIC.

One might ask whether FH-type models fitted over successive survey years could possibly reinforce or 'borrow strength' from each other. For such an approach to work, intuition suggests that the residuals from the best FH model-type fitted over successive years would have to show some correlation. In that light, Table 4 suggests that such time-sequence model structure might conceivably help: models at a remove of less than 3 years have residuals with correlations at least .34, but correlations at lags of four years or more are small. However, even the correlations at lags of one or two years are likely due mostly to correlated CPS ASEC sampling errors e_{it}.

4. COMPARISON ACROSS YEARS AND GLM MODELS

The main GLM model we consider is GLsaip, which is (3) with predictor variables exactly the same as the FHsaip model. In addition, we tried several variants: first, GLsp4 augments the set of predictors in (3) by the logarithms of the denominators in the log-rate predictors, exactly as in FHsp4; similarly, the model GLsp2 augments the basic set of four SAIPE log-rate predictors by log(RPOPest), $\log^2$(RPOPest) as in FHsp2; next, after finding that the fitted random-intercept variance σ^{*2} for GLsaip is large (roughly 1.0), we fitted models GL.2 and GL.04 as (3) with σ^{*2} fixed respectively at 0.2 and 0.04; and finally, the model GLexp is just like GLsaip except that the *logit* link in (3) is replaced by the *log* link, $g(x) = e^x$, in (2). Table 5 shows for GLM model AIC values, even more strikingly than among the FH models, the value of augmenting the predictor set by the log(RPOPest) and $\log^2$(RPOPest) variables, and GLsp2 seems clearly best. However, in all of these GLM models the AIC values are suspect because of one or more influential counties in which the *logLik* contribution was more than 1000.

Table 2: AIC values minus 1800, for the FH models in Table 1 with parameter dimensions d, by SAIPE survey year, based on data for CPS-sampled counties with ≥ 1 sampled poor (school-age, related) child. Values are comparable only within column; smallest two values in column are displayed in boldface.

Model	d	90	94	96	98	99	00
FHsaip	6	457	741	224	196	230	183
FHwtrg	6	478	783	245	216	248	196
FHspA	6	426	696	207	174	204	170
FHspB	6	396	657	192	152	180	157
FHspO	6	325	575	153	89	120	113
FHrt	6	**320**	**562**	**137**	**83**	**116**	**98**
FHsp1	7	435	696	210	181	222	176
FHsp4	10	431	685	214	186	226	181
FHsp4B	10	384	633	191	152	184	160
FHsp1C	7	450	736	225	185	228	180
FHsp2	8	428	679	198	173	209	170
FHsp2B	8	377	621	175	136	163	148
FHrt2	8	**308**	**536**	**125**	**71**	**100**	**91**

Although the next Section shows that the GLM models fit the CPS county data excellently, the correlations in Table 6 between GLM county-rate predictors and the most recent census rates are remarkably weak on CPS-sampled counties (except for GL.04), but still interestingly strong (almost as strong as those of FHsaip and other good FH models) on the non-CPS counties. This becomes understandable only with the remark that the GLMM's have large PSU-effect variances, ranging from 1.0 to 1.4, and as a result their SAE's give much greater weight to the CPS direct estimators at small sample sizes than do FHsaip and the related models with small σ^{*2}. Models GL.2 and GL.04 with σ^{*2} pa-

Table 3: Correlations between FH model SAE's and corresponding census rates, (A) over all CPS-sampled counties by survey year, and (B) over all non-CPS-sampled counties. Last column's entries are correlations with 2000 census, all other columns' are correlations with 1990 census estimates.

(A) Sampled

Model	90	94	96	98	99	00	00new
FHsaip	.93	.97	.90	.93	.95	.96	.95
FHwtrg	.93	.97	.90	.93	.95	.96	.95
FHspA	.94	.95	.89	.92	.93	.94	.95
FHspB	.93	.94	.90	.91	.92	.94	.94
FHspO	.74	.78	.79	.71	.74	.74	.76
FHrt	.95	.97	.93	.94	.95	.95	.95
FHsp1	.95	.95	.93	.92	.94	.95	.95
FHsp4	.94	.94	.89	.92	.93	.95	.95
FHsp4B	.93	.92	.91	.90	.91	.93	.94
FHsp1C	.93	.96	.90	.92	.94	.95	.95
FHsp2	.95	.96	.94	.93	.94	.95	.95
FHsp2B	.94	.94	.93	.90	.91	.93	.93
FHrt2	.95	.95	.94	.92	.92	.94	.94

(B) Non-sampled

Model	90	94	96	98	99	00	00new
FHsaip	.89	.94	.89	.89	.92	.93	.90
FHwtrg	.88	.94	.88	.88	.92	.92	.89
FHspA	.89	.94	.90	.89	.92	.93	.91
FHspB	.90	.94	.90	.90	.92	.93	.91
FHspO	.91	.95	.92	.90	.92	.93	.90
FHrt	.90	.94	.90	.90	.92	.93	.91
FHsp1	.90	.94	.91	.89	.92	.93	.90
FHsp4	.89	.93	.89	.88	.91	.93	.90
FHsp4B	.89	.93	.91	.89	.91	.93	.90
FHsp1C	.88	.95	.89	.89	.92	.93	.90
FHsp2	.91	.95	.92	.89	.90	.93	.90
FHsp2B	.90	.93	.92	.88	.87	.91	.89
FHrt2	.90	.92	.90	.86	.85	.90	.88

Table 4: Correlation matrix among `FHsaip` model residuals (only for CPS-sampled counties with sampled poor children) across model years.

	90	94	96	98	99	00
90	1.000	0.244	0.143	0.030	0.015	0.066
94	0.244	1.000	0.454	0.076	0.040	0.033
96	0.143	0.454	1.000	0.341	0.113	0.051
98	0.030	0.076	0.341	1.000	0.740	0.459
99	0.015	0.040	0.113	0.740	1.000	0.778
00	0.066	0.033	0.051	0.459	0.778	1.000

Table 5: AIC values minus *Const* × 1000, for GLM models in each survey year, based on all CPS-sampled counties, with *Const* values displayed in bottom row. AIC values are comparable only within columns. Numbers of predictor variables for the six models are respectively 4, 8, 6, 4, 4, and 4. As discussed in text, AIC's are suspect because all models have some large single-county log-likelihood contributions.

Model	90	94	96	98	99	00
GLsaip	916	1011	515	135	688	766
GLsp4	895	1000	507	127	685	748
GLsp2	886	987	492	114	667	740
GL.2	1706	1900	1276	1055	1823	1931
GL.04	3824	3968	3183	3126	4223	4448
GLexp	1016	1042	564	202	732	801
Const	144	152	135	126	134	141

Table 6: Correlations between GLM model SAE's and corresponding census child-poverty estimators, (A) over all CPS-sampled counties by survey year, and (B) over all non-CPS-sampled counties. Last column's entries are correlations with 2000 census, all other columns' with 1990 census estimates.

(A) Sampled

Model	90	94	96	98	99	00	00new
GLsaip	.58	.56	.55	.54	.50	.48	.52
GLsp4	.58	.56	.55	.54	.50	.48	.53
GLsp2	.58	.56	.55	.54	.50	.49	.53
GL.2	.72	.72	.70	.67	.65	.61	.66
GL.04	.89	.90	.87	.84	.83	.82	.85
GLexp	.57	.52	.52	.52	.46	.46	.50

(B) Non-sampled

Model	90	94	96	98	99	00	00new
GLsaip	.92	.93	.88	.86	.86	.86	.88
GLsp4	.91	.93	.89	.85	.85	.85	.87
GLsp2	.92	.93	.90	.86	.83	.84	.87
GL.2	.92	.94	.90	.87	.89	.88	.90
GL.04	.92	.94	.91	.87	.90	.89	.90
GLexp	.92	.94	.89	.87	.87	.87	.90

Table 7: Weighted sum of absolute differences loss-function, calculated for county child-poverty rate predictors from indicated models and SAIPE data years, with respect to CPS truth-standard, over CPS counties (A) with and (B) without sampled poor children. `Hybrid` model, discussed in text, averages SAE's for `FHsaip` and `GLsaip`.

(A) With poor

	90	94	96	98	99	00
FHsaip	923	1048	1005	1042	1035	961
FHwtrg	1017	1207	1132	1159	1144	1053
FHspO	167	255	311	206	236	242
FHsp2	896	985	956	1004	998	940
FHrt	1035	1185	1145	1149	1121	1059
GLsaip	70	101	87	74	65	59
GLsp2	70	102	87	74	65	59
GL.04	481	587	554	560	531	495
GLexp	80	105	92	85	73	78
Hybrid	495	573	544	556	548	508

(B) No poor

	90	94	96	98	99	00
FHsaip	124	178	170	176	182	168
FHwtrg	124	181	171	177	183	168
FHspO	145	200	184	207	210	192
FHsp2	114	162	156	165	172	158
FHrt	124	177	168	178	182	167
GLsaip	36	57	51	45	42	39
GLsp2	34	55	49	44	41	38
GL.04	106	155	144	146	150	149
GLexp	41	66	58	53	49	47
Hybrid	80	118	110	111	112	104

rameters fixed at the smaller values .2 and .04 have progressively better correlations between SAE's and census rates, and the loss-function performance of `GL.04` is roughly comparable with that of the better FH models and of `Hybrid`, against both CPS and census standards.

The model `GLexp` was found to behave similarly to `GLsaip`. Tables 5 and 7 indicate slightly worse behavior for the log-link model on CPS counties with respect to the CPS standard; while Tables 6(B), 8, and 9 indicate a slight preference for `GLexp` on all counties with respect to the census standard.

5. COMPARISONS BETWEEN FH AND GLM MODELS

We can already contrast the FH and GLM models through the correlations in Tables 3 and 6 of their county predicted rates with those of the most recent decennial census. By this comparison with the external (census) truth-standard, the FH models seem preferable. We now compare the models, versus both internal and external standards, using the weighted sum of absolute deviations loss-function mentioned at the end of Section 2.

The weighted absolute difference between a SAE ($\hat{\vartheta}_i$) and a standard ($\vartheta_{i,std}$) child-poverty rate has a useful operational interpretation. With the weight w_i equal to the county child population, the term $w_i\,|\hat{\vartheta}_i - \vartheta_{i,std}|$ is equal to the absolute difference between the actual number of poor children according to the standard and the predicted number $w_i\,\hat{\vartheta}_i$. Alternatively, this term can be viewed as one-half the number of children misclassified as poor by the model producing the SAE's $\hat{\vartheta}_i$, if the classifications were made in the best possible way within the county subject to the total number $w_i\hat{\vartheta}_i$ classified poor. In the tables below, the county weights w_i are actually taken to be the county child population updated demographically from the most recent decennial census, divided by 2000. Thus the loss function numbers in Tables 7-9 should be viewed roughly as 1/4000 times the total numbers of children (over the indicated sets of CPS-sampled counties with sampled poor children, CPS counties with no sampled poor children, for non-CPS counties) that a given SAE method would misclassify as poor. For the respective survey years '90, '94, '96, '98, '99, '00, the totals of the weights w_i over the (3-year aggregated) CPS-sampled counties with sampled poor children are 17352, 19297, 19218, 18870, 19018, and 19920; and the totals over the CPS counties with no sampled poor children are 1259, 1404, 1364, 1712, 1757, and 1694. The corresponding totals over non-CPS counties are 4053, 3445, 4260, 4658, 4648, 4730. The displayed losses in the With-poor portions of Tables 7 and 8 should be viewed as proportions of the first (CPS, with sampled poor) set of totals; those in the No-poor portions of Tables 7 and 8 as proportions of the second set; and the losses in Table 9 as proportions of the third (non-CPS) set of totals. In one immediate sense, all of the models considered can be viewed as very successful: the losses in Tables 7(A) and 8(A) are never more than about 6% of the applicable totals, while those in Table 9 are never more than about 3% of the non-CPS total of weights w_i.

The glaringly clear result from the tables of loss-function values collected in Table 7 is that, across all years of data, the GLM class of models give smaller losses by far with respect to the CPS truth-standard than the FH class of (small σ_u^2) models on the CPS counties, with or without sampled poor children. Similarly, the clear result from Table 8 is that with respect to the census truth-standard, the (small-σ_u^2) FH models give far smaller losses than the (large-σ^{*2}) GLMM's on CPS counties. Finally, in Table 9 the losses on non-CPS counties (necessarily with respect to census standard) are very similar between the FH and GLM models. In both Tables 7(A) and 8(A), the FH model `FHsaip0` with large σ_u^2 and GLMM `GLM.04` with small σ^{*2} have losses falling between the other FH and GLM models.

Tables 7–9 continue to show that among the FH models, `FHsaip, FHsp2`, and `FHrt` are consistently best: `FHsp0` looks good only within the with-poor subtable of Table 7. In CPS counties with either truth-standard, and also in non-CPS counties, the losses from `GLsp2` are hardly different from those of `GLsaip`. The GLM models give mostly smaller losses in 1990 and 1994 than the FH models on non-CPS counties, but the reverse is true in later years.

It is striking that no single model or method of small-area estimation seems to perform well in all three of the settings explored here: CPS counties with respect to CPS truth-standard, CPS counties with respect to census standard, and non-CPS counties. Indeed, due to differences between census and CPS rates, no single method could perform comparably to `FHsaip` with respect to census standard and also to `GLsaip` on CPS counties. Maiti and Slud (2002) experimented with GLM-type models which up-weight the log-likelihood contributions from larger counties, and such re-weighted likelihood analyses yield losses like those from GLM models but slightly closer to those from FH models. But these methods have little to recommend them.

To create county-level estimates which are much better than the worse class of models for each of the truth standards on CPS counties, no better method is available than a straight county-by-county average of small-area point estimates from `FHsaip` and `GLsaip`: we denote this method as `Hybrid` in Tables 7 to 9. As expected, the method generally splits the difference between the losses of its two underlying models. However, in a few years ('90, '96, '00), `Hybrid` does as well on non-CPS counties as the better of its two components.

In future work, it seems promising to study whether fixed-variance forms of the exponential-link model (2), or variants with other link functions g^{-1}, could provide a more successful, principled approach to splitting the difference between the performance of estimators based on `FHsaip` and `GLsaip`. In a different direction, it would be interesting to see whether the results of Table 4 correlating the residuals from SAIPE small-area predictions, would persist after correcting for multi-year aggregation of the CPS sample, leading to a multi-year unified linear model incorporating the log-rate SAIPE model `FHsaip`.

Table 8: Weighted sum of absolute differences loss-function, calculated for county child-poverty rate predictors from indicated models and SAIPE data years, with respect to census truth-standard, over CPS counties (A) with and (B) without sampled poor children.

(A) With poor

	90	94	96	98	99	00
FHsaip	278	413	392	387	357	255
FHwtrg	283	347	345	341	317	290
FHspO	823	1096	946	1033	939	733
FHsp2	328	610	519	475	422	253
FHrt	338	314	329	380	381	348
GLsaip	921	1235	1155	1161	1108	914
GLsp2	919	1238	1157	1162	1108	912
GL.04	529	793	734	722	686	495
GLexp	917	1236	1155	1157	1110	904
Hybrid	519	786	727	708	662	500

(B) No poor

	90	94	96	98	99	00
FHsaip	17	21	24	31	27	22
FHwtrg	18	22	25	31	27	22
FHspO	26	37	29	40	33	36
FHsp2	17	18	21	34	33	21
FHrt	17	22	23	31	29	22
GLsaip	87	108	112	134	153	120
GLsp2	89	109	113	136	154	121
GL.04	23	24	29	43	49	29
GLexp	82	99	105	127	146	112
Hybrid	44	48	56	70	83	56

Remark 1 Both of the classes of models (1) and (3) have unresolved methodological issues. So far, no accurate estimators of mean-squared error of the small-area estimators from either type of model are known. In the FH case, 'second-order corrected' MSE estimators for small-area estimates $\vartheta_{it} = E(y_{it} \,|\, u_{it})$ have recently been derived by Slud and Maiti (2004), in the setting where y_{it} is a nonlinear transformation of a response variable satisfying a FH model. However, the log transformation in the FH model (1) results in counties with no sampled poor children being dropped from the estimation of population-wide unknown parameters, a form of truncation or biased sampling for which no cor-

Table 9: Weighted sum of absolute differences loss-function, calculated for county child-poverty rate predictors from indicated models and SAIPE data years, with respect to census truth-standard, over non-CPS counties.

	90	94	96	98	99	00
FHsaip	111	67	103	130	125	102
FHwtrg	116	76	109	129	120	103
FHspO	107	106	104	132	120	131
FHsp2	113	73	111	160	156	110
FHrt	106	66	107	141	143	107
GLsaip	95	88	123	147	153	119
GLsp2	101	79	132	155	167	115
GL.04	95	78	106	140	139	106
GLexp	88	94	117	145	148	108
Hybrid	95	76	109	137	136	104

rection is yet available. This was a large part of the motivation for studying the models (3) in the SAIPE context. These models also lead to small-area estimators for which no good mean-squared error estimators are known; moreover, these small-area estimators are biased due in part to the downward bias of the maximum likelihood estimators of σ^2, and no good bias-correction is known, either. □

Despite the methodological issues just remarked, the small-area estimators from the best variants of both the FH and GLM models perform very well on the SAIPE data, and it seems likely that the deficiencies of these estimators in Sections 3 and 4 above stem largely from the discrepancies between the CPS (internal) and census (external) truth-standards.

6. CONCLUSIONS

This paper has provided as comprehensive a survey as is currently possible of the behavior of different Fay-Herriot and Generalized Linear Mixed models for the SAIPE county-level child-poverty data. The best of a broad array of FH models compared has been found to be the SAIPE program's formulation, discussed in the NAS panel Report (Citro and Kalton 2000) of a log child-poverty rate model with four log-rate predictors, possibly with a sampling error with variance inversely proportional to the square root of sample size in place of sample-size itself. The best choice among the GLM's (random-intercept logistic regression models) studied, is one with either the same set of log-rate predictors as the SAIPE log-rate model, or one with these predictors augmented by the log and log-square of the demographically estimated population. A GLM with *log* link has comparable performance, and may even be a little better on non-CPS counties.

The besetting problem in SAIPE, that estimators must be judged both by an internal (CPS ASEC) and external (census) standard of truth, has been seen to yield Fay-Herriot type estimators which conform remarkably well to the census standard, and GLM type estimators which fit the CPS sampled data better by all reasonable loss-function measures. Both the Fay-Herriot and random-intercept logistic models (with essentially the same predictor variables) are good, and both perform comparably well on the non-CPS-sampled counties. A Hybrid model which averages SAE's from the FH and GLM models has performance which roughly splits the difference in losses between the separate SAE's, as does a GLMM with a fixed small PSU effect variance.

7. REFERENCES

Citro, C. and Kalton, G., eds. (2000) **Small-Area Estimates of School-Age Children in Poverty**, *Evaluation of Current Methodology* (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. *Jour. Amer. Statist. Assoc.* **74**, 341-53.

Fisher, R. and Asher J. (1999) Alternate CPS sampling variance structures for constrained and unconstrained county models. Tech. Rep. #1, **http://www.census.gov/hhes/www/saipe/publications.html#tech**.

Fisher, R. and Asher J. (2000) Alternate CPS sampling variance structures for constrained and unconstrained county models: Additional research. Tech. Rep. #4, same **SAIPE** web-site as previous ref.

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

Maiti, T. and Slud, E. (2002) Comparison of Small Area Models in SAIPE. Census Bureau preprint.

Slud, E. (2003) Assessing fit of SAIPE models to Census and CPS county child-poverty rates. Federal Committee on Statistical Methodology Research Conference Proceedings, Nov. 2003, Arlington, VA.

Slud, E. and Maiti, T. (2004) MSE estimation in transformed Fay-Herriot models, with application to SAIPE. Census Bureau preprint.