RESEARCH REPORT SERIES
(Statistics #2012-04)

An Empirical Study on Using Previous American Community Survey Data Versus Census 2000 Data in SAIPE Models for Poverty Estimates

Elizabeth T. Huang
William R. Bell

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: April 25, 2012

An Empirical Study on Using Previous American Community Survey Data Versus Census 2000 Data in SAIPE Models for Poverty Estimates

Elizabeth T. Huang and William R. Bell

Abstract

The Census Bureau's Small Area Income and Poverty Estimates Program (SAIPE) produces model-based poverty estimates at the county and state level. SAIPE uses Fay-Herriot (1979) models with dependent variables obtained from direct survey poverty estimates (currently obtained from ACS, but prior to 2005 obtained from CPS), and regression predictor variables derived from tabulations of IRS tax data, SNAP (Supplemental Nutrition Assistance Program, formerly food stamps) program data, and previous census estimates (since 2000, these have been the Census 2000 long form estimates). Although the latter have consistently been important predictors in the state and county models, as time advances and the Census 2000 poverty estimates become further removed from the production year, questions arise about their continued value in the model, and particularly about whether they might be somehow replaced in the model by ACS estimates for previous years. At the county level this would suggest consideration be given to replacing Census 2000 estimates with ACS 5-year estimates formed from data for the 5 years preceding the production year (because the only estimates published for all counties are 5-year estimates.) At the state level, the Census 2000 estimates could be replaced by single-year ACS estimates for the year immediately preceding the production year.

In using previous census poverty estimates to define regression variables, SAIPE has ignored the fact that these are survey estimates obtained from the long form and so contain sampling error. At the state level, the sampling errors of the Census 2000 long form estimates used by SAIPE are essentially negligible and can be ignored. This is less true at the county level, however, particularly for small counties. Furthermore, in considering the replacement in the model of previous census estimates with previous ACS estimates, this issue becomes more pressing, as the ACS sampling variances are higher. We illustrate this point in the report. When a predictor variable, such as Census 2000 long form data or previous ACS data, contains non-negligible sampling error, a bivariate Fay-Herriot model with that predictor as a second dependent variable, can account for that uncertainty. We take that approach in this study, using bivariate models in which "current year" ACS estimates define one of the dependent variables, and either Census 2000 estimates or previous ACS estimates define the second dependent variable. We then compare prediction error variances (posterior variances) from these models to assess which predictor variable—Census 2000 estimates or previous ACS estimates—yields the

lowest prediction error variances for the current year. We do this for models at the state and county level for which ACS single-year estimates for 2009 provide the current year estimates. We also obtain county model results with ACS single-year estimates for 2010 providing the current year estimates.

There are two general conclusions from our study. One is that the differences in prediction error variances depending on which data define the second dependent variable are generally not large. The second conclusion is that, among the three candidates, prediction error variances from using ACS multi-year estimates from previous years as the second dependent variable were in some cases lower, and were generally not higher, than those from the other two candidate models. This suggests that replacing the univariate Fay-Herriot models that use Census 2000 estimates to define regression predictor variables with bivariate Fay-Herriot models that use previous 5-year ACS estimates as the second dependent variable may yield improvements as we move further away from Census 2000, and this change is unlikely to do worse.

1. Introduction

The U. S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program produces median household income and age group poverty estimates for states and counties using linear small area models of the form proposed by Fay and Herriot (1979). Dependent variables in these models are single-year direct estimates from the American Community Survey (ACS).[1] Regression variables are obtained from previous census income and poverty estimates and from administrative record data. The latter include tax data available under an agreement with the Internal Revenue Service (IRS) and Supplemental Nutrition Assistance Program (SNAP) participant data obtained from the U.S. Department of Agriculture.[2]

The previous census estimates provide a very important predictor variable in the SAIPE models. These estimates come from the previous census long form sample which, in this decade, has been the Census 2000 long form. As time progresses, the Census 2000 long form estimates are becoming more distant in time from the income and poverty conditions measured by the current ACS estimates, something that one would expect to reduce the usefulness of the previous census estimates in the models.[3] Since the 2010 Census did not include a long form sample, it does not provide an updated replacement for the Census 2000 estimates used in the SAIPE models (as was the case when Census 2000 long form estimates replaced 1990 Census long form estimates in the SAIPE models). This naturally raises the question as to whether more current ACS estimates (particularly 5-year ACS estimates that are available for all counties) should take the place of the Census 2000 estimates in the SAIPE models.

Despite the increased timeliness of recent ACS estimates compared to the Census 2000 estimates, another factor that must be considered is their higher sampling variability compared to the Census 2000 estimates. Noise in predictor variables will reduce their usefulness, leading to higher prediction error (posterior) variances than would be obtained with noise-free data, with more noise (higher sampling variances) having larger negative effects. In addition, while the SAIPE models have used previous census estimates to define what have been treated as fixed regression variables, their sampling error actually presents a classical "error in variables" problem. At the state level, the variances of the census long form income and poverty estimates used in the SAIPE models are effectively negligible, so this problem does not arise. At the county level, however, and particularly for small counties, variances of the long form estimates are not necessarily small. Ignoring the sampling error of the long form estimates can contribute to inefficiencies in the predictions, and will also contribute to misstatements of the true

---

[1] Through 2004, the dependent variables were direct estimates from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS). Use of ACS data for this purpose started with the 2005 SAIPE estimates.

[2] The SNAP program was previously called the food stamp program.

[3] Section 4 presents some evidence that as the time interval between estimates (e.g., between current ACS and Census 2000 long form estimates) increases, the value of the prior Census estimates in the models decreases.

prediction error variances by those produced with the assumed model. Because ACS 5-year estimates have higher sampling variances than the long form estimates, replacing the census long form estimates by ACS estimates would exacerbate these problems.

One way to account for the sampling error in the census or ACS estimates when they are used as a predictor variable is via a bivariate model, with the census or ACS estimates providing the dependent variable in the second equation. Sampling variances of these estimates are taken as known (from available variance estimates). This approach, which is discussed in the context of SAIPE models by Bell (1997, 1998a), is used here. It facilitates comparisons of prediction error variances from models that use as a predictor either Census 2000 long form estimates or a more recent ACS alternative. Making such comparisons is our primary objective here.

The remainder of this report proceeds as follows. Section 2 defines a general bivariate model for SAIPE estimation, and Section 3 specifies three particular bivariate models and defines criteria for comparing their posterior variances. For 2009 (2010) estimates, all three models use ACS 2009 (2010) poverty estimates to define the dependent variable in the first equation of the bivariate model, but the models differ as to whether Census 2000 or more recent ACS data define the dependent variable in the second equation. Section 4 provides empirical results when applying the three bivariate models at the county level. Section 4.1 provides results for county models of log number of school-age children in poverty, while Section 4.2 provides results for models of county poverty rates of school-age children. Section 4.3 examines the structure of the bivariate model in more detail to show the drawbacks of simply using past ACS or previous census estimates as an added regression variable in a univariate Fay-Herriot model. Section 5 provides bivariate model results for state poverty rate models for various age groups for 2009. Finally, Section 6 summarizes the results and discusses future research.

2. General Bivariate Model

Let Y_{1i} and Y_{2i} be the “true poverty characteristics” for area (county or state) i that are being estimated by surveys 1 and 2, respectively, for areas $i = 1, \dots, n$. Let y_{1i} and y_{2i} be the direct survey estimates of Y_{1i} and Y_{2i}. In general, these “two surveys” could be the same survey conducted in two different years or two different surveys conducted in the same year or different years. Then we have

$$y_{1i} = Y_{1i} + e_{1i} \tag{1}$$

$$y_{2i} = Y_{2i} + e_{2i} \tag{2}$$

where the sampling errors e_{1i} and e_{2i} are assumed to be independently distributed as $N(0, v_{ji})$,

j = 1,2. The v_{ji} are assumed known although, in practice, they are estimated. We assume that Cov(e_{1i}, e_{2i}) = 0, though the model is easily generalized to allow e_{1i} and e_{2i} to be correlated (with their correlations also assumed known.)

The model for the true poverty characteristics is:

$$Y_{1i} = x_{1i}'\beta_1 + u_{1i} \tag{3}$$

$$Y_{2i} = x_{2i}'\beta_2 + u_{2i} \tag{4}$$

where x_{1i} and x_{2i} are vectors of regression variables, β_1 and β_2 are the corresponding vectors of regression parameters, and (u_{1i} ,u_{2i}) are independently and identically normally distributed over i with zero means. Note that the vectors x_{1i} and x_{2i} could be the same or different depending on the nature of Y_{1i} and Y_{2i}. In this research, x_{1i} and x_{2i} will generally be nominally the same variables but defined for different years in the two equations. We write

$$\mathrm{Var}(u_{1i}) = s_{11}, \quad \mathrm{Var}(u_{2i}) = s_{22}, \quad \text{and} \quad \mathrm{Corr}(u_{1i}, u_{2i}) = \rho.$$

Equations (1) – (4) specify the full bivariate model.

We use a Bayesian treatment of the model with the following noninformative prior distributions for the model parameters:

$\boldsymbol{\beta} = (\beta_1', \beta_1')'$ is distributed as multivariate $N(\mathbf{0}, cI)$, with c large (c = 1,000),

s_{11} and s_{22} are independently distributed as Uniform (0, m_1) and (0, m_2), with m_1 and m_2 large,

ρ is distributed as Uniform (−1,1).

If ρ = 0, then the bivariate model reduces to separate univariate models for each of the two years.

To implement Bayesian inference for the bivariate model, we used Gibbs sampling via the JAGS program (Plummer 2010) for county models and the WinBUGS program (Spiegelhalter, et al. 2003) for state models. (Both programs handle essentially the same model forms and do the same calculations, but the much larger number of observations for the county models led to use of JAGS, while WinBUGS was simpler to implement for the state models.) For any of the models, we used JAGS or WinBUGS to simulate a large number (5000 for county models, 10,000 for state models) of sets of the model parameters (ρ, s_{11}, s_{22}) from their posterior distribution. The posterior means and variances of Y_{1i} were then approximated by appropriately averaging results over simulations of [(ρ, s_{11}, s_{22}) | $\mathbf{y}$] to approximate the following formulas:

$$E(Y_{1i}|\boldsymbol{y}) = E_{\rho,s11,s22|\boldsymbol{y}}[E(Y_{1i}|\boldsymbol{y},\rho,s_{11},s_{22})] \tag{5}$$

$$Var(Y_{1i}|\boldsymbol{y}) = E_{\rho,s11,s22|\boldsymbol{y}}[Var(Y_{1i}|\boldsymbol{y},\rho,s_{11},s_{22})] + Var_{\rho,s11,s22|\boldsymbol{y}}[E(Y_{1i}|\boldsymbol{y},\rho,s_{11},s_{22})] \tag{6}$$

where $\mathbf{y} = \{(y_{1i}, y_{2i}), i = 1,\ldots,n\}$ is the observed data.

3. Alternative Model Specifications and Model Comparison Criteria

We use three bivariate models to estimate 2009 poverty (county or state level) and examine which of these models produces lower posterior variances of the poverty characteristic Y_{1i}. In all three bivariate models we use the ACS 2009 poverty estimate as y_{1i} in equation (1), and use contemporaneous regression variables such as those currently used in 2009 SAIPE production as x_{1i}, *except* we omit the Census 2000 poverty variable. Thus, equation (1) is the same in all three models. The three models differ according to the definition of the dependent variable, y_{2i}, in the second equation:

- Model I. y_{2i} ~ Cen 2000, meaning that y_{2i} is the Census 2000 long form poverty estimate (which is for income reported in 1999).
- Model II. y_{2i} ~ $\overline{\text{ACS}}_{05-08}$, where $\overline{\text{ACS}}_{05-08}$ indicates the average of the ACS estimates for 2005 through 2008. This stand-in for an ACS 5-year estimate is necessitated by the fact that, for modeling 2009 ACS estimates, only four years of production ACS data prior to 2009 are available (the first production 5-year estimates being for 2005-2009).
- Model III. y_{2i} ~ ACS_{2008}, meaning that y_{2i} is the 2008 ACS poverty estimate. We will additionally present some results for cases where y_{2i} is the county ACS estimate for some still earlier year, which will be indicated in the same way (e.g., y_{2i} ~ ACS_{2007}).

In Models I and III, the regression variables in x_{2i} are the same as the variables in x_{1i}, but they are defined for the same year as y_{2i}. As is the case with x_{1i}, x_{2i} does not include the prior census estimate. In Model II, the regression variables in x_{2i} are the averages of the corresponding variables for the years 2005-2008 (again, leaving out the census data).

The county poverty models we consider take one of two forms: y_{1i} is either the logarithm of the ACS single-year estimate of the county number of 5-17 year-old children in poverty, or the ACS single-year county estimate of the county age 5-17 poverty rate. Then, y_{2i} is, correspondingly, either the log of the number of 5-17 children in poverty, or the 5-17 poverty rate, obtained from Census 2000 or previous ACS estimates, depending on the model. At the state level, we examine only models for poverty rates, but we consider models for four age groups: 0-4, 5-17, 18-64, and 65+.

We extend the analysis of county bivariate models to let y_{1i} equal the 2010 ACS poverty estimate (log number in poverty or poverty rate), with the regression variables in x_{1i}, the second dependent variable y_{2i}, and the latter's vector of regression variables x_{2i} all shifted forward one year. Model II provides a slight exception in that, when applying it for 2010, the multi-year $\overline{\text{ACS}}_{05-08}$ data defining y_{2i} in the 2009 model is replaced by the five-year ACS_{05-09} estimates. Also, in this case the regression variables in x_{2i} are the regression variables used in the SAIPE production model for 2007 (again, leaving out census data). Note that 2007 is the mid-year of the 2005-2009 5-year estimate period. We did not extend the analysis of state poverty models to 2010 ACS data.

For each area i used in the modeling, we obtain the posterior mean and variance of Y_{1i} using equations (5) and (6). We focus on the posterior variances, computing ratios of these from the different bivariate models as follows (shown for the case of y_{1i} ~ ACS 2009 estimates):

$$R1_i = Var(Y_{1i} \,|\, \boldsymbol{y}_1, \boldsymbol{y}_2 \sim \text{ACS}_{2008}) / Var(Y_{1i} \,|\, \boldsymbol{y}_1, \boldsymbol{y}_2 \sim \text{Cen 2000})$$

$$R2_i = Var(Y_{1i} \,|\, \boldsymbol{y}_1, \boldsymbol{y}_2 \sim \overline{\text{ACS}}_{05-08}) / Var(Y_{1i} \,|\, \boldsymbol{y}_1, \boldsymbol{y}_2 \sim \text{Cen 2000})$$

$$R3_i = Var(Y_{1i} \,|\, \boldsymbol{y}_1, \boldsymbol{y}_2 \sim \overline{\text{ACS}}_{05-08}) / Var(Y_{1i} \,|\, \boldsymbol{y}_1, \boldsymbol{y}_2 \sim \text{ACS}_{2008}).$$

When y_{1i} ~ ACS 2010 county estimates, we shift the variables in the above expressions forward one year, as discussed above. The above three ratios provide measures, comparable across areas, indicating which of the three models provides the lowest MSE (posterior variance). Values of the ratios close to 1.0 indicate not much difference in the MSEs of the two models involved in the ratio. Examining the ratios for the large number of individual areas involved is somewhat impractical (especially for the large number of counties), so we instead provide summary statistics—mean, minimum, first quartile (Q1), median, third quartile (Q3), and maximum—of the ratios across areas. As the results could vary by size of area, for counties the Appendix provides results within various groups defined by county population sizes.

4. Analyses of County Bivariate Models

This section presents empirical results from county bivariate models applied to the two forms of dependent variables mentioned above. Section 4.1 considers models of survey estimates of logarithms of the number of 5-17 children in poverty. Section 4.2 considers models of survey estimates of 5-17 child poverty rates. These two sections present analogous results, with the exception that Section 4.1 presents some additional results from models where the y_{2i} are defined to be logs of ACS single-year poverty estimates for each of several previous years ranging from 2005-2008 (4 different bivariate models), and the y_{1i} are the corresponding ACS log poverty estimates for 2009. These additional results give some indication of how the value of using past ACS data in the models varies as the past estimates are pushed further back in time.

4.1 Bivariate models for county log poverty estimates of related children age 5-17

Table 1 below presents results from fitting the various bivariate models of log poverty estimates of related children age 5-17 with y_{1i} corresponding to ACS 2009 estimates. This includes Models I and II, and four versions of Model III with y_{2i} corresponding to ACS estimates for single years 2008, 2007, 2006, and 2005. The first row of Table 1 shows how many counties were included in the model fitting. Because these models involve taking logarithms of the ACS or Census 2000 estimates of the number of 5-17 year-old children in poverty, and because some of these estimates are zero (so logs cannot be taken), not all 3,141 U.S. counties could be included in the model fitting. The number of counties included thus varied each year depending on how many counties had direct survey estimates of zero from either of the two estimates (both had to be nonzero for a county to be included in the modeling). Since the models all have the same first equation, the variations in n come from variation in the zero poverty estimates related to the second equation. Census 2000 had the largest sample among those in the table, resulting in the fewest counties being dropped from the modeling due to zero estimates, and so has the largest n. The various ACS single-year estimates resulted in the next fewest counties being dropped, with fairly similar n values for each of the four years. The ACS 2005-2008 average estimates resulted in the most counties being dropped because, for this model, counties were dropped if any of their four single-year poverty estimates was zero. (The four-year average here is computed after taking logarithms of the single-year ACS poverty estimates, rather than averaging the ACS 2005-2008 poverty estimates and then taking logarithms.)

While the dropping of counties from model fitting is not a desirable feature, it is necessary here since the models involve taking logarithms, and has been the practice with the SAIPE county production model from the beginning. The percentage of counties dropped here is about 10 percent when single year ACS estimates provide y_{2i}. It is somewhat higher for the ACS 2005−2008 average estimates and somewhat lower for the Census 2000 estimates. While this may seem a large enough percentage to be of potential concern, the actual percentage of the ACS sample effectively dropped is far less than 10 percent, since the counties with zero poverty estimates are invariably small counties, and so have small samples. A county with a small sample contributes much less to the model fitting than does a county with a large sample,[4] so that

[4] The model fitting weights county estimates inversely proportional to their variances, which implies weights roughly proportional to their sample sizes.

the model fitting results should not be much affected by the dropping of the zero estimate counties. Our summary measures of the MSE ratios for the model prediction results would be affected primarily in the sense that the results somewhat under-represent small counties, but that is partly addressed in the Appendix by examining results by county size classes.

Table 1 provides model fitting results on the posterior means of the parameters ρ, s_{11}, and s_{22}. Note that the posterior mean of ρ, the correlation between the model errors in the two equations, is lowest when the second equation uses Census 2000 data. This is perhaps not surprising given that the Census 2000 estimates are the most distant in time from the ACS 2009 estimates. There is, in fact, a general indication of lower posterior means of ρ the further the estimates in the second equation are pushed back into the past, which is again not surprising. The posterior means of the model error variances, s_{11} and s_{22}, seem roughly comparable over years and across the two equations, with one exception—the posterior mean of s_{22} for the equation using the ACS 2005-2008 estimates. This is, in fact, the one case where the model error variance is not comparable to the others, as taking the four-year average of the (logged) ACS estimates y_{2i} also takes, from equation (2), the four-year average of Y_{2i}, which will hence reduce the model error variance in the second equation. Due to positive correlation between the model errors across years, the reduction in the model error variance due to this averaging would not be expected to be by a factor of ¼, but rather by a smaller amount.[5]

Table 1. Posterior means of parameters (and observation counts *n*) for bivariate models of log county poverty estimates of children age 5-17: Results for y_{1i} ~ ACS 2009 estimates.

$y_{2i} \Rightarrow$	**Cen 2000**	$\overline{\mathbf{ACS}}_{\mathbf{05-08}}$	$\mathrm{ACS}_{\mathbf{2008}}$	$\mathrm{ACS}_{\mathbf{2007}}$	$\mathrm{ACS}_{\mathbf{2006}}$	$\mathrm{ACS}_{\mathbf{2005}}$
n	2,946	2,661	2,836	2,844	2,862	2,849
ρ	0.52	0.80	0.88	0.71	0.79	0.74
s_{11}	0.0224	0.0191	0.0212	0.0210	0.0201	0.0212
s_{22}	0.0208	0.0152	0.0253	0.0295	0.0234	0.0272

[5] With a first order autoregressive process with parameter .88 (to match the correlation of the model errors in the ACS 2009 and ACS 2008 equations), the variance of the mean of four successive observations would be .86 times the variance of a single observation. Dividing the posterior mean of the model error variance for $\overline{\mathrm{ACS}}_{05-08}$ (.0152) by the average of the posterior means of the 2005-2008 ACS model error variances from the 2005-2008 ACS equations (.02635) gives an even lower value, .58. So the model error variance for $\overline{\mathrm{ACS}}_{05-08}$ actually appears somewhat lower in relation to the other model error variances than would be expected, given the high values of the correlations.

Table 2. Posterior means of parameters (and observation counts n) for bivariate models of log county poverty estimates of children age 5-17: Results for y_{1i} ~ ACS 2010 estimates.

$y_{2i} \Rightarrow$	**Cen 2000**	**ACS_{05-09}**	**ACS_{2009}**
n	2,975	2,976	2,860
ρ	0.50	0.61	0.66
s_{11}	0.0238	0.0218	0.0226
s_{22}	0.0212	0.0196	0.0214

Table 2 presents results analogous to those of Table 1, but for models with y_{1i} corresponding to ACS 2010 estimates. First note that the numbers of counties n used in the model fitting are similar to the numbers n in Table 1, except when y_{2i} corresponds to the multi-year ACS_{05-09} estimate we have n = 2976, which is much larger than the value of n = 2665 in Table 1 when $y_{2i} \sim \overline{\text{ACS}}_{05-08}$. This occurs because when $y_{2i} \sim \overline{\text{ACS}}_{05-08}$ we omitted from the model fitting any county for which the ACS estimate was zero in any of the individual years 2005-2008, whereas when $y_{2i} \sim \text{ACS}_{05-09}$ we omitted only those counties for which the ACS estimate was zero in all of the years 2005-2009 (because logs were then taken of the 5-year estimates, not of an average of the single-year estimates).

The posterior means of the correlation ρ between the model errors of the two equations of the model with Census 2000 data in the second equation are very similar in Tables 1 and 2: ρ = .52 when $y_{1i} \sim \text{ACS}_{2009}$, and ρ = .50 when $y_{1i} \sim \text{ACS}_{2010}$. However, for the models where past ACS estimates provide y_{2i}, ρ is smaller when $y_{1i} \sim \text{ACS}_{2010}$ (.61 for $y_{2i} \sim \text{ACS}_{05-09}$ or .66 for $y_{2i} \sim \text{ACS}_{2009}$) than when $y_{1i} \sim \text{ACS}_{2009}$ (.80 for $y_{2i} \sim \overline{\text{ACS}}_{05-08}$ and .88 for $y_{2i} \sim \text{ACS}_{2008}$). We believe this simply shows that results like these can vary over time. Posterior means of the model error variances, s_{11} and s_{22}, seem roughly comparable over years and across the two equations. The value of s_{22} when $y_{2i} \sim \text{ACS}_{05-09}$ (.0196, in the model with $y_{1i} \sim \text{ACS}_{2010}$) is not as low as when $y_{2i} \sim \overline{\text{ACS}}_{05-08}$ (.0152, in the model with $y_{1i} \sim \text{ACS}_{2009}$)[6], but the former did not involve the same averaging as did the latter.

Table 3 provides summary statistics for the ratios $R1_i$, $R2_i$, and $R3_i$ comparing the posterior variances of the three bivariate models that use ACS 2009 or ACS 2010 estimates in the first equation, and the three corresponding alternative estimates in the second equation.[7] For $y_{1i} \sim \text{ACS}_{2009}$, the results for $R1_i$ suggest little difference, in terms of the posterior variances of Y_{1i}, to using ACS_{2008} versus Census 2000 data—the ratios are fairly concentrated around 1.0.

[6] The value of s_{22} for ACS_{05-09} of 0.0196 is a little higher than 0.0172 = 0.6778750 × 0.02536 that we would obtain from a first order autoregressive process with adjacent year correlations of 0.66, and single-year model error variances for 2005-2009 as given in Tables 1 and 2.

[7] Since the highest correlation between model errors for the ACS 2009 equation and another single-year ACS equation occurs for y_{2i}~ACS_{2008}, the lowest posterior variances of Y_{1i} among these alternatives would be expected for y_{2i}~ACS_{2008},, and so Table 3 shows results only for this case.

The results for $R2_i$ and $R3_i$, however, show that the posterior variances of Y_{1i} when using the model with $\overline{\text{ACS}}_{05-08}$ data are generally lower than those when using either ACS_{2008} or Census 2000 data. While the differences are not large—mean and median variance reductions from using $\overline{\text{ACS}}_{05-08}$ data instead of the others being just a little over 10 percent—one does certainly not see any advantage from using either ACS_{2008} or Census 2000 data instead of the $\overline{\text{ACS}}_{05-08}$ data. The general conclusion here is thus that, for modeling logs of ACS 2009 county poverty estimates of children age 5-17, use of $\overline{\text{ACS}}_{05-08}$ estimates for y_{2i} in the second equation is preferred.

Table 3. Summary statistics of the variance ratios $R1_i$, $R2_i$, and $R3_i$ for bivariate models of log county poverty estimates of children age 5-17.

Results for y_{1i} ~ ACS 2009 estimates

	$R1_i = \frac{\text{Var with ACS}_{2008}}{\text{Var with Cen 2000}}$	$R2_i = \frac{\text{Var with }\overline{\text{ACS}}_{05-08}}{\text{Var with Cen 2000}}$	$R3_i = \frac{\text{Var with }\overline{\text{ACS}}_{05-08}}{\text{Var with ACS}_{2008}}$
# counties	2,830	2,655	2,661
mean	1.01	0.88	0.87
minimum	0.69	0.65	0.73
Q1	0.97	0.83	0.84
median	1.03	0.89	0.87
Q3	1.06	0.93	0.89
maximum	1.19	1.05	1.15

Results for y_{1i} ~ ACS 2010 estimates

	$R1_i = \frac{\text{Var with ACS}_{2009}}{\text{Var with Cen 2000}}$	$R2_i = \frac{\text{Var with ACS}_{05\text{-}09}}{\text{Var with Cen 2000}}$	$R3_i = \frac{\text{Var with ACS}_{05\text{-}09}}{\text{Var with ACS}_{2009}}$
# counties	2,853	2,969	2,858
mean	1.06	0.96	0.90
minimum	0.94	0.84	0.78
Q1	1.03	0.93	0.88
median	1.06	0.95	0.90
Q3	1.09	0.98	0.93
maximum	1.18	1.10	1.01

For y_{1i} ~ ACS_{2010}, the ratios $R1_i$, $R2_i$, and $R3_i$ are slightly higher, a result most likely due to the lower values of ρ shown in Table 2 for the cases where y_{2i} corresponds to past ACS data. However, we still conclude that using past ACS multi-year estimates for the dependent variable in the second equation seems preferred, or at least does not do worse than the other two alternatives.

Appendix A provides corresponding tables giving summary statistics within county population size groups for $R1_i$, $R2_i$, and $R3_i$ from bivariate models of county log poverty estimates. Table A.1 presents results for y_{1i} ~ ACS_{2009}, and Table A.2 results for y_{1i} ~ ACS_{2010}. Note that these results come from bivariate models fitted to all counties (except those with zero ACS estimates), with the prediction error variances then summarized by county size groups.

The upper section of Table A.1 shows that, for small counties, using $y_{2i} \sim \mathrm{ACS}_{2008}$ to borrow information from last year's ACS estimates produces slightly higher posterior variances of Y_{1i} for 2009 than does using $y_{2i} \sim \mathrm{Cen}\ 2000$ (values of $R1_i$ tend to exceed 1.0), while, for large counties, using $y_{2i} \sim \mathrm{ACS}_{2008}$ produces substantially lower posterior variances. This suggests that the stronger correlation of the model errors for the estimates closer together in time (the ACS_{2009} and ACS_{2008} estimates) is of benefit for larger counties, but for smaller counties the lower sampling variance of the Census 2000 estimates outweighs this advantage. The upper section of Table A.2 shows somewhat similar results for posterior variances of Y_{1i} using $y_{2i} \sim \mathrm{ACS}_{2009}$ versus $y_{2i} \sim \mathrm{Cen}\ 2000$, though the advantages to the former for large counties are less, presumably due to the reduced value of ρ from Table 2, whereas the corresponding ρ with $y_{2i} \sim \mathrm{Cen}\ 2000$ changes little from Table 1.

The middle section of Table A.1 shows that borrowing information from the four-year average ACS estimates ($\overline{\mathrm{ACS}}_{05-08}$) yields lower posterior variances of Y_{1i} for 2009 than does borrowing information from Census 2000 estimates. As with $y_{2i} \sim \mathrm{ACS}_{2009}$, the advantage to using the $\overline{\mathrm{ACS}}_{05-08}$ estimates presumably comes from their being closer together in time to the ACS_{2009} estimates than are the Census 2000 estimates. For large counties this is the determining factor, while for small counties this is partly, though not completely, offset by the lower sampling variances of the Census 2000 estimates. The middle section of Table A.2 shows somewhat similar results for borrowing information from $ACS_{05\text{-}09}$ for predicting Y_{1i} for 2010 versus borrowing information from Census 2000 although, as in the preceding paragraph, the advantages to borrowing from the ACS estimates are less due to the reduced value of ρ for this case from Table 2.

The lower section of Table A.1 shows that borrowing information from the four-year average ACS estimates ($\overline{\mathrm{ACS}}_{05-08}$) yields mostly lower posterior variances of Y_{1i} for 2009 than does borrowing information from last year's (ACS_{2008}) estimates. While the ACS_{2008} equation model errors have a higher correlation with the model errors from the ACS_{2009} equation, this advantage is offset by the lower sampling variance of the $\overline{\mathrm{ACS}}_{05-08}$ estimates. In the group of largest counties (> 1,000,000 population), the result is about a wash, with the $R3_i$ values concentrated around 1.0. As counties get smaller, however, the $\overline{\text{ACS}}_{05-08}$ estimates generally tend to do better. The lower section of Table A.2 shows similar results for posterior variances of Y_{1i} for 2010.

The general conclusion is that using last year's ACS estimate (of log of 5-17 children in poverty) in a bivariate model can do better for large counties than using the previous census estimate in the model, though the former might do worse for small counties. Using the most recent prior ACS multi-year estimate appears to do better, or at least as well, as the other two alternatives for both large and small counties.

4.2 Bivariate models for county poverty rates of related children age 5-17

In this section we provide results corresponding to those of Section 4.1, but using models for (untransformed) county poverty rates. In practice, these models have the disadvantage that they have the potential of producing negative predictions of poverty rates, or at least of producing prediction intervals that include negative values. The models are useful for our purposes here to provide a robustness check on the results of Section 4.1, by allowing us to see if changing the model form has an appreciable influence on the results.

Regression variables in our model for untransformed county poverty rates are untransformed ratio variables analogous to the regression variables used in a model for log county poverty rates explored by Bell, et al. (2007). These are the tax poor child exemption rate, the tax child filing rate, the SNAP (food stamp) participation rate, and the child (0-17) population proportion (county population of ages 0-17 divided by the county population of all ages). The last of these, which is obtained from the Census Bureau's Population Estimates Program (PEP), replaces the log county 0-17 population variable used in a log poverty rate model by Bell, et al. (2007, pp. 36-39). The child population proportion seems an appropriate variable for an untransformed poverty rate model. Again, we omit Census 2000 estimates from the regression variables so we can assess their value in a bivariate model relative to the use of the ACS alternatives.

Table 4 provides results on the posterior means of the parameters ρ, s_{11}, and s_{22} for the bivariate poverty rate models for 2009 and 2010. For both years the posterior mean of ρ is lowest when the second equation uses Census 2000 data, and considerably lower than the values given for the 2009 and 2010 log number 5-17 in poverty models in Tables 1 (0.52) and 2 (0.50). These lower values of ρ suggest there will be less benefit in the poverty rate models to borrowing information from the Census 2000 estimates, than there was in the log number in poverty models. The values of ρ are higher when the second equation uses past ACS single-year or multi-year estimates, and not so different from the values in Tables 1 and 2. The slightly higher values of ρ when y_{2i} is last year's ACS estimate rather than the previous ACS multi-year estimate, suggests that again this higher correlation will compete against the lower sampling variance of the ACS multi-year estimate to determine which provides the most benefit to predicting the current year's Y_{1i}. The posterior means of s_{22} are distinctly different for the three alternative models for each year, more so than was the case in Tables 1 and 2, although, again, the values corresponding to the ACS multi-year estimates stand out as the lowest.

Table 4. Posterior means of parameters (and observation counts *n*) for bivariate models of county poverty rates of children age 5-17

Results for $\boldsymbol{y_{1i}}$ ~ ACS 2009 estimates

$y_{2i} \Rightarrow$	**Cen 2000**	$\overline{\mathbf{ACS}}_{\mathbf{05-08}}$	$\mathrm{ACS}_{\mathbf{2008}}$
n	2,952	2,665	2,840
ρ	0.30	0.72	0.80
s_{11}	0.0023	0.0022	0.0024
s_{22}	0.0013	0.0005	0.0024

Results for $\boldsymbol{y_{1i}}$ ~ ACS 2010 estimates

$y_{2i} \Rightarrow$	**Cen 2000**	$\mathrm{ACS}_{\mathbf{05-09}}$	$\mathrm{ACS}_{\mathbf{2009}}$
n	2,973	2,976	2,865
ρ	0.33	0.62	0.77
s_{11}	0.0034	0.0035	0.0034
s_{22}	0.0013	0.0007	0.0024

Table 5 presents summary statistics for the ratios $R1_i$, $R2_i$, and $R3_i$ comparing the posterior variances of the bivariate county poverty rate models that use ACS 2009 or ACS 2010 estimates in the first equation and one of the three corresponding alternative estimates in the second equation. First, the results for $R2_i$ appear quite comparable to those in Table 3, suggesting roughly the same amount of potential improvement from using previous ACS multi-year estimates versus using Census 2000 estimates for y_{2i}. The results for $R1_i$, on the other hand, are more favorable to use of last year's ACS estimates versus Census 2000 estimates for y_{2i} than was the case in Table 3 (in which the mean and median values of $R1_i$ are close to, and even slightly larger than, 1). The results for $R3_i$ are rather neutral between the choice of last year's ACS estimates versus previous ACS multi-year estimates for y_{2i} — the mean and median values of $R3_i$ are close to 1 in 2009, and slightly larger than 1 in 2010.

Table 5. Summary statistics of the variance ratios $R1_i$, $R2_i$, and $R3_i$ for bivariate models of county poverty rates of children age 5-17.

Results for y_{1i} ~ ACS 2009 estimates

	$R1_i = \frac{\text{Var with ACS}_{2008}}{\text{Var with Cen 2000}}$	$R2_i = \frac{\text{Var with }\overline{\text{ACS}}_{05-08}}{\text{Var with Cen 2000}}$	$R3_i = \frac{\text{Var with }\overline{\text{ACS}}_{05-08}}{\text{Var with ACS}_{2008}}$
# counties	2,834	2,659	2,664
mean	0.90	0.91	1.03
minimum	0.45	0.73	0.85
Q1	0.83	0.88	0.95
median	0.91	0.91	1.01
Q3	0.98	0.94	1.05
maximum	1.10	1.00	2.03

Results for y_{1i} ~ ACS 2010 estimates

	$R1_i = \frac{\text{Var with ACS}_{2009}}{\text{Var with Cen 2000}}$	$R2_i = \frac{\text{Var with ACS}_{05\text{-}09}}{\text{Var with Cen 2000}}$	$R3_i = \frac{\text{Var with ACS}_{05\text{-}09}}{\text{Var with ACS}_{2009}}$
# counties	2,858	2,969	2,864
mean	0.92	0.97	1.06
minimum	0.54	0.80	0.86
Q1	0.88	0.94	1.01
median	0.93	0.97	1.04
Q3	0.97	1.00	1.08
maximum	1.06	1.08	1.79

The main impression from the combined results of Tables 3 and 5 is that using previous ACS multi-year estimates for y_{2i} instead of Census 2000 values has some chance of producing better results, and there is no suggestion this would do worse. Results for using last year's single-year ACS estimates are more equivocal—they perform comparably overall to the ACS multi-year estimates for the poverty rate models (Table 5), but don't do quite as well as the ACS multi-year estimates for the log number in poverty models (Table 3). With this limited amount of data, the logical choice would be to replace the Census 2000 estimates in a bivariate model with the previous ACS multi-year estimates. (Note this does not suggest simply using previous ACS multi-year estimates as regression variables in a univariate model for current single-year ACS estimates—see Section 4.3.) This choice is also appealing in that it will continually keep the y_{2i} variable reasonably current, whereas continuing to use Census 2000 estimates for y_{2i} would result in its becoming increasingly distant from the current year, with an expected deterioration in performance.

Appendix B provides tables giving summary statistics within county population size groups for $R1_i$, $R2_i$, and $R3_i$ from bivariate models of the county 5-17 poverty rate estimates. Tables B.1 and B.2 present results for y_{1i} ~ ACS_{2009} and for y_{1i} ~ ACS_{2010}, respectively. In comparison to the results in Tables A.1 and A.2 for the log number in poverty models, the results of Tables B.1 and B.2 show somewhat less indication that potential advantages from using ACS estimates in place of Census 2000 estimates are larger for large counties and smaller for small counties. The results in the first two sections of Tables B.1 and B.2 do show this tendency for county population size groups up to 65,000 – 250,000, but less so beyond this point. (However, even Tables A.1 and A.2 show slight upticks in $R1_i$ and $R2_i$ for the group of largest counties.)

Another result of note in Table 5 is that two values are extreme: the minimum $R1_i$ of 0.45 and the maximum $R3_i$ of 2.03. The first of these may seem to suggest that for one county (and probably others as well that are not individually discernible from the summary statistics of Table 5), use of the single year ACS 2008 estimates for y_{2i} produced a dramatically lower posterior variance compared to use of Census 2000 estimates. The second may seem to suggest a similar result for use of the single year ACS 2008 estimates in comparison to use of $\overline{\mathrm{ACS}}_{05-08}$ estimates. These particular results seem surprising given the overall results. What is the explanation?

The explanation lies in errors in sampling variance estimates based on small samples. One assumption of the standard Fay-Herriot models, univariate and bivariate, is that sampling variances, v_{ji}, of the survey estimates are known when, in reality, they are estimated. When samples are small, both the direct survey point estimates and the direct sampling variance estimates will be subject to large errors (large sampling errors in y_{1i} being the motivation for doing modeling in the first place). For counties with small samples, some of the sampling variance estimates will substantially underestimate the true variances, and some will substantially overestimate the true variances. Note that, for small counties, if the true variance (which will be large when the sample is small) were known and were used, the direct estimate for the county would get little weight in the posterior mean, and the posterior variance would be similar to that which would be obtained if there were no direct estimate. Thus, for counties with small samples, overestimating the sampling variance will have little effect.

When the sampling variance is substantially underestimated for a small county, however, the calculations will give too much weight to the direct point estimates in forming the posterior mean, and the posterior variance calculation will place too much confidence in the direct point estimates, so the posterior variance will be substantially too low. Bell (2008) illustrates this phenomenon for univariate Fay-Herriot models, but the same principle applies to the bivariate versions as well. The problem for the calculation of posterior variances is essentially that the term $Var(Y_{1i}|y_{1i}, y_{2i}, \rho, s_{11}, s_{22})$ in equation (6) (taking a slight liberty with the notation to make the point clear), will be understated (for all nonzero ρ) whenever the sampling variance of either y_{1i} or y_{2i} is substantially underestimated. If the sampling variance of y_{1i} is underestimated, this will affect the posterior variances calculated for all three of our bivariate models in similar ways, and thus have only minor effects on the variance ratios $R1_i$, $R2_i$, and $R3_i$. If the sampling variance of last year's single-year ACS estimate is substantially underestimated, however, this likely will result in substantial understatement of the posterior variance only for the model where y_{2i} corresponds to last year's single-year ACS estimate. It is unlikely that the sampling variance of the corresponding ACS multi-year estimate will be similarly underestimated, since this estimate involves additional years of data and a 4-5 times larger sample. Thus, the posterior variances for the models using ACS multi-year estimates or Census 2000 estimates are much less likely to be substantially understated. The net effect on the variance ratios of substantial underestimation of the sampling variance of last year's single-year ACS estimate for a given county will thus be a value for $R1_i$ that is substantially too low, and a value for $R3_i$ that is substantially too high.

The table in Appendix C provides evidence of this phenomenon. In this table, for the bivariate county 5-17 poverty rate model with y_{1i} corresponding to ACS 2009 estimates, we show, for each county size group shown, the five lowest values of $R1_i$ along with the corresponding values of $R3_i$, the ACS 2008 and 2009 county poverty rate estimates, and the corresponding sampling variance estimates. Note that, especially for the first three groups corresponding to the smallest county population sizes, the ACS 2008 direct poverty rate estimates and sampling variances are all quite low. In fact, the first row shows a vanishingly small ACS 2008 poverty rate estimate of 0.0067, and a correspondingly low sampling variance estimate of 0.000103 (corresponding to a standard error of about .01, implying extreme confidence that the true poverty rate is indeed very small). Corresponding entries in other rows of the table are also quite small. It appears these are all likely to be substantial underestimates due to a large amount of random estimation error arising from small ACS (single-year) samples. This leads to the extreme values shown for $R1_i$ and $R3_i$, as discussed in the previous paragraph.

We can explain further in general terms how this arises. For small counties a single year of ACS data typically provides a small sample. For estimating the poverty rate of 5-17 year-olds, the sample effectively shrinks because we use data from only those households containing at least one 5-17 year old child. This might be only 10-15 households. With such a small sample, there may be no households in poverty (leading to an estimated poverty rate of zero), or there might be one household in poverty. In the latter case, it could occur that this one household has only one 5-17 child. When this happens, the direct estimated poverty rate is the sampling weight for this one poor 5-17 child divided by the sum of the sampling weights for all the 5-17 children in the sample. Sampling weights vary across persons depending on differential sampling rates across geographic areas, on differential effects of population controls, and on whether the household's data came from a mail or telephone response versus a CAPI (computer assisted person interview) response. CAPI cases get higher weights because they are subsampled.[8] If the one poor 5-17 child came from a mail or telephone response, it would have a much lower weight than would the 5-17 child responses from CAPI cases. If the one 5-17 child also happens to have a sampling weight lower than the weights of other 5-17 child mail or telephone responses, this can result in a very low estimated 5-17 poverty rate. It turns out that the corresponding variance estimate will then also be very low, for reasons we won't bother going into here.

Without disclosing specifics of the data, this is essentially the explanation for the very low estimated poverty rates and variances in the table of Appendix C. This led to low posterior variances for these counties from the model with y_{2i} defined by these ACS 2008 estimates, resulting in extremely low values of $R1_i$ and, generally, extremely high values of $R3_i$. (There are a few exceptions to the last point, especially in the largest county size group, where sample sizes are not so small.) These results are not truly indicating superior performance for these counties of the model with y_{2i} defined by the ACS 2008 estimates, they are merely reflecting the effects on sampling variance estimates, and the resultant effects on the posterior variance ratios, of random estimation error from small samples.

4.3 Further analysis of the bivariate model

We now reconsider the general bivariate model, which we restate for convenience here in a way that combines equations (1) and (3), and also combines equations (2) and (4):

$$y_{1i} = Y_{1i} + e_{1i} = x'_{1i}\beta_1 + u_{1i} + e_{1i} \tag{7}$$

$$y_{2i} = Y_{2i} + e_{2i} = x'_{2i}\beta_2 + u_{2i} + e_{2i.} \tag{8}$$

We can project (regress) Y_{1i} on y_{2i}, using results on conditional expectations in a normal distribution, to write

$$Y_{1i} = x'_{1i}\beta_1 + \gamma_i(y_{2i} - x'_{2i}\beta_2) + w_i \tag{9}$$

where the regression coefficient on the second equation regression residuals, $y_{2i} - x'_{2i}\beta_2$, is

$$\gamma_i = \frac{s_{12}}{s_{22}+v_{2i}} \tag{10}$$

[8] For more information on the ACS sample design and estimation, see the ACS web site at http://www.census.gov/acs/www/methodology/methodology_main/.

and the residual w_i has variance

$$Var(w_i) = Var(Y_{1i}|y_{2i}) = s_{11} - \frac{s_{12}^2}{(s_{22}+v_{2i})}. \tag{11}$$

Note that both γ_i and $Var(w_i)$ vary over counties as v_{2i}, the sampling variance of y_{2i}, varies. Note also that as v_{2i} gets large, γ_i approaches zero. Thus, for counties with very small samples and correspondingly large sampling variances, the variable $y_{2i} - x'_{2i}\beta_2$ coming from the second equation will get very little weight in the predictions of Y_{1i}. This makes intuitive sense—we should give more weight to past data that is quite reliable (low sampling variance) than to past data that is very unreliable (high sampling variance).

Combining equations (7) and (9) gives

$$y_{1i} = Y_{1i} + e_{1i} = [x'_{1i}\beta_1 + \gamma_i(y_{2i} - x'_{2i}\beta_2) + w_i] + e_{1i}, \tag{12}$$

which looks like a univariate Fay-Herriot model with the one added regressor, $y_{2i} - x'_{2i}\beta_2$, but with a regression coefficient and model error variance, γ_i and $Var(w_i)$, that vary over counties. If this variation over counties is small, which will be the case if the v_{2i} are all small (as occurs with the use of Census 2000 estimates in the SAIPE state models), then treating (12) as a univariate Fay-Herriot model—forcing a constant γ and $Var(w_i)$—will provide a good approximation to the bivariate model. If the variation in v_{2i} is large (relative to s_{22}), however, then doing this will give too much weight to $y_{2i} - x'_{2i}\beta_2$ from counties with small samples, and too little weight to this data from counties with large samples.

Figures 1 and 2 illustrate these results for the case where y_{1i} corresponds to ACS 2010 county poverty estimates. Figure 1 shows results for the log number in poverty model, and Figure 2 shows results for the poverty rate model. In both figures the left column of plots shows the values of γ_i for the three models (with y_{2i} corresponding to ACS 2009 estimates, or ACS 2005-2009 estimates, or Census 2000 estimates, respectively), and the right column of plots shows the corresponding values of $Var(w_i)$. The values of γ_i and $Var(w_i)$ were obtained by plugging posterior means of the parameters s_{11} , s_{12}, and s_{22} into (10) and (11).

Figures 1 and 2 show considerable variation over counties in the values of γ_i and $Var(w_i)$, with more variation for the models using ACS data for y_{2i} than for the models using Census 2000 data. This is because of the lower sampling variances of the Census 2000 estimates. This shows why using a bivariate county model is more important when borrowing information from past ACS estimates than when borrowing information from previous census estimates.

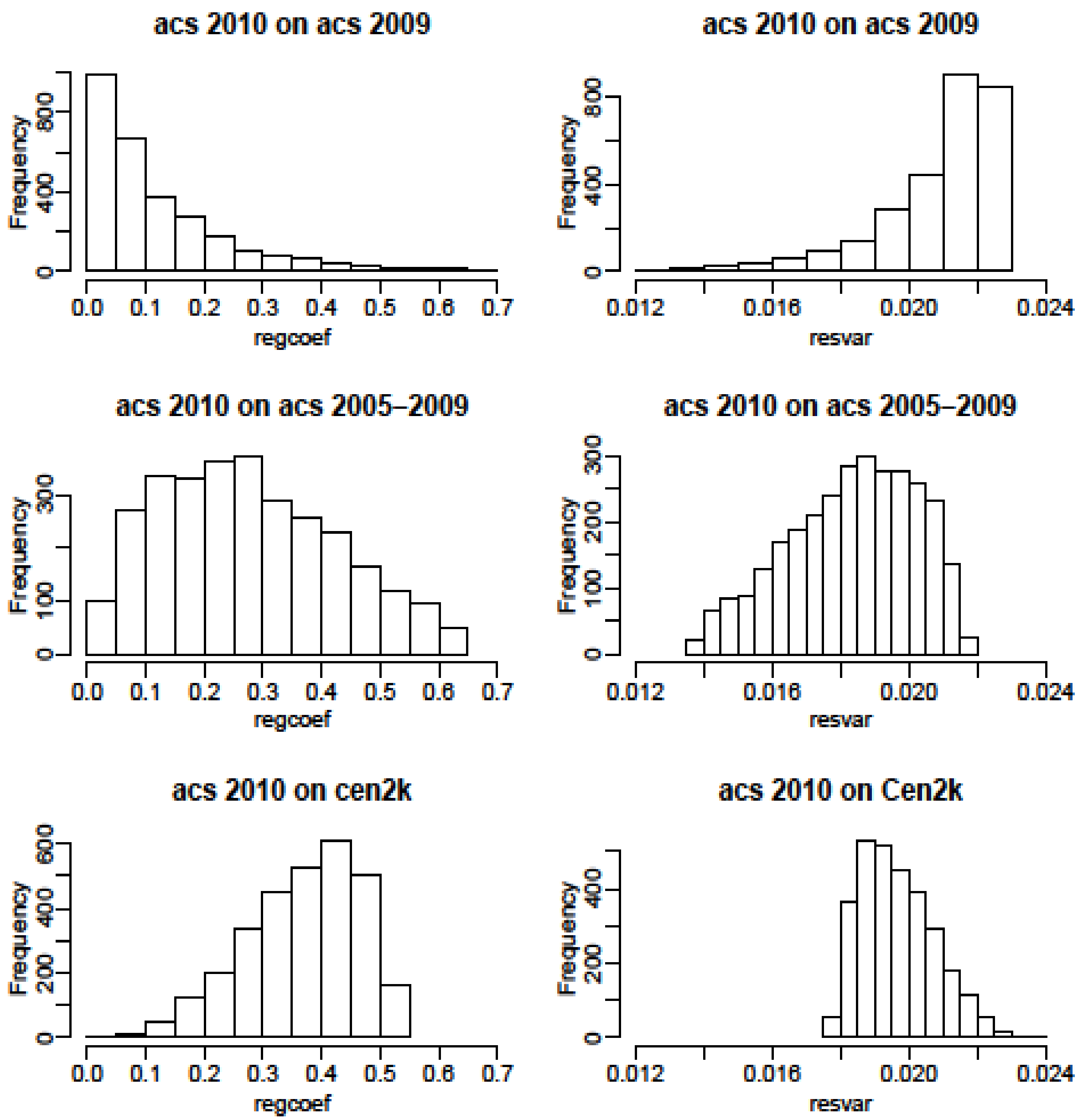

Figure 1. Implied regression coefficients and residual variances from county bivariate models of log number of children in poverty. The first equation corresponds to the ACS 2010 estimates, the second to either the ACS 2009 estimates, the average of the ACS 2005-2008 estimates, or the Census 2000 estimates. The graphs show the implied regression coefficients on the second equation regression residuals when used as a predictor variable in the first equation, and the corresponding residual variances, which become the variances of the county level random effects in the implied univariate model for the ACS 2010 estimates. See equations (7)-(12).

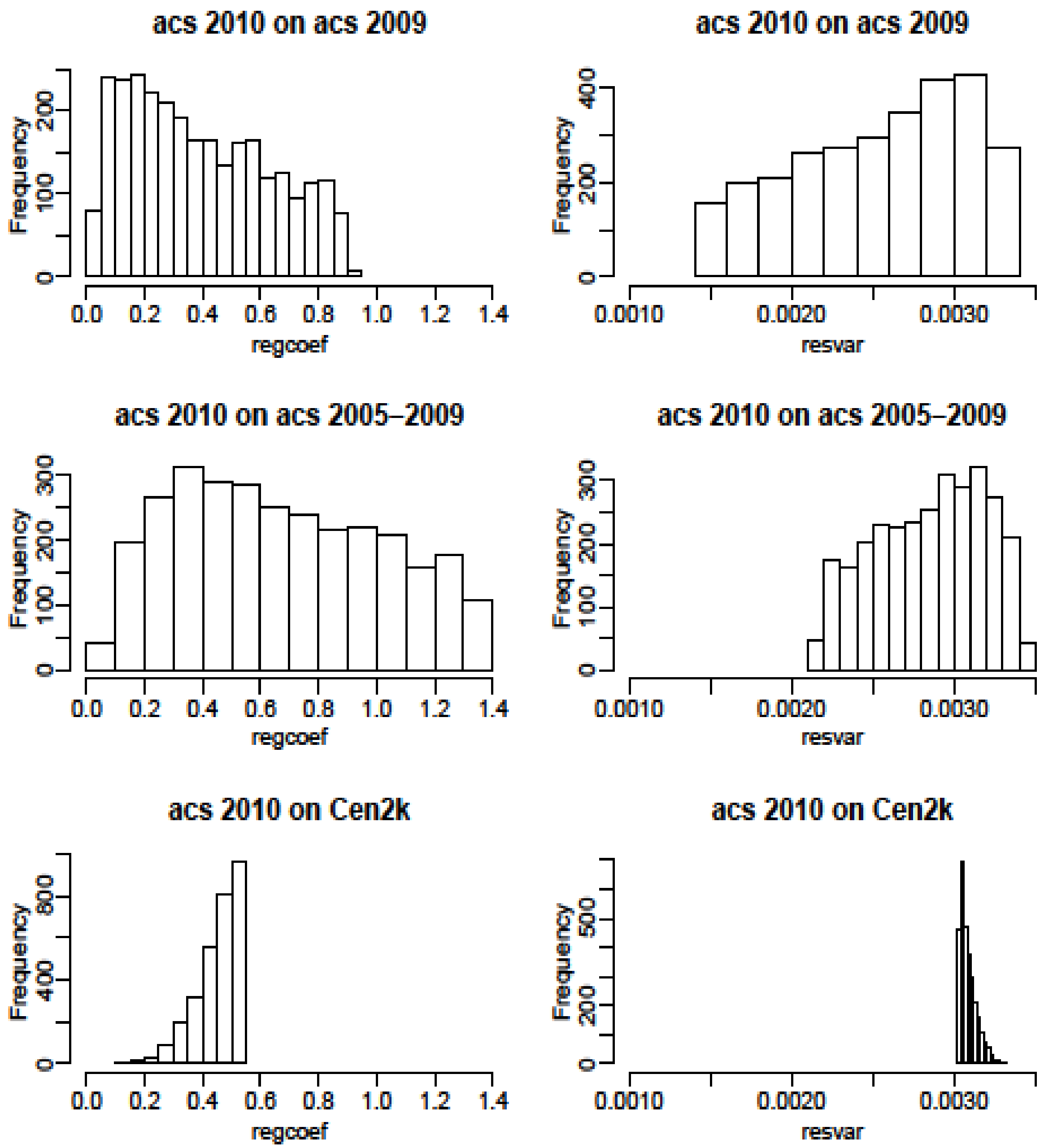

Figure 2. Implied regression coefficients and residual variances from county bivariate models of 5-17 child poverty rates. The plots are analogous to those of Figure 1.

There are other differences in the results across the three different forms of the bivariate models. For the log number in poverty model (Figure 1), the consistently largest values of γ_i occur for the model with y_{2i} corresponding to the Census 2000 estimates, and the consistently smallest values of γ_i occur for the model with y_{2i} corresponding to the ACS 2009 estimates. For the poverty rate model (Figure 2), this is not the case as many counties have values of γ_i for the model with y_{2i} corresponding to the ACS 2009 estimates that are larger than the values of γ_i for any county from the model with y_{2i} corresponding to the Census 2000 estimates. Recall that the value of γ_i depends on s_{12} (in addition to v_{2i}), which in turn depends on ρ, s_{11} (roughly the same for all three models), and s_{22}. The low values of ρ (0.30 and 0.33) for the poverty rate models using Census 2000 estimates lead to somewhat lower values of γ_i.

The general message from Figures 1 and 2 is that simply using past ACS data to define a regression variable for a univariate Fay-Herriot model would misstate the relationship between the target population quantities Y_{1i} and the past ACS data. In small counties, the strength of the relationship would be overstated, while in large counties it would be understated. Model error variances would also be misstated, leading directly to misstatements of posterior (prediction error) variances. These problems exist to an extent in the current SAIPE univariate Fay-Herriot county models, which use Census 2000 data, but to a lesser degree than would be the case with ACS data, due to the lower sampling variances of the Census 2000 estimates.

5. Bivariate Models for 2009 State Poverty Ratios by Age Groups

In this section we provide results on models for ACS 2009 state poverty ratios by 4 age groups (age 0-4, 5-17, 18-64, and 65+). The SAIPE state poverty ratio model for age 5-17 related children and its regression variables are given in Bell, et al. (2007, pp. 66-69). These models use regression variables analogous to those of the county poverty rate model discussed in Section 4.2. In addition to the intercept term, these are ratios related to poverty, including the tax-poor child exemption rate, the tax nonfiler ratio, the state SNAP participation rate, and "census residuals" obtained by regression of the previous census (Census 2000) state 5-17 poverty ratio estimates on the other regression variables defined in the census income year (1999 for Census 2000). Since sampling errors of the Census 2000 state poverty ratios are negligible, use of the census residuals as a regression variable is nearly equivalent to use of the bivariate model[9]—note the development in Section 4.3 if $v_{2i} = 0$. For age 65+, the Supplemental Security Income (SSI) state participation rate is used instead of the SNAP variable, and the census poverty rate for age 65+ is used instead of the "census residual".

[9] There is a difference between the two models in how the model parameters would be estimated, but the prediction results given the estimated parameters are identical.

Table 6 presents results on the posterior means of the model parameters in the state poverty ratio bivariate models. The posterior mean of the correlation between the model errors in the two equations of the bivariate state poverty ratio model is lowest, 0.65, when y_{2i} ~ Census 2000, and highest, 0.86, when $y_{2i} \sim \overline{\text{ACS}}_{05-08}$. The correlation when $y_{2i} \sim \text{ACS}_{2008}$ is in between at 0.77. These results suggest that using $y_{2i} \sim \overline{\text{ACS}}_{05-08}$ may be best, given the high correlation and the fact that sampling variances of the $\overline{\text{ACS}}_{05-08}$ state poverty ratios are small, if somewhat larger than those of the Census 2000 estimates. The posterior mean of s_{22} is smallest when $y_{2i} \sim \overline{\text{ACS}}_{05-08}$, which again may have something to do with the 4-year averaging.

Table 6. Posterior means of parameters (and standard errors of ρ) for bivariate models of 2009 state poverty rates of children age 5-17.

y_{2i} ⇒	**Cen 2000**	$\overline{\textbf{ACS}}_{\mathbf{05-08}}$	$\textbf{ACS}_{\mathbf{2008}}$
ρ	0.65 (0.12)	0.86 (0.07)	0.77 (0.10)
s_{11}	1.618	1.593	1.516
s_{22}	1.750	1.542	1.666

The posterior means of the parameters of the state poverty ratio models for the other age groups are given in Table D.1 of Appendix D. For all age groups, the posterior means of ρ are higher when y_{2i} corresponds to past ACS estimates than when it corresponds to Census 2000 estimates, due, presumably, to the smaller time difference. Note the extremely high value of ρ for age 65+. This reflects the greater stability over time of the true age 65+ poverty ratios.

Table 7 provides summary statistics for the ratios $R1_i$, $R2_i$, and $R3_i$ comparing the posterior variances of the three bivariate models for the 5-17 state poverty ratios. The results for $R1_i$ suggest not much difference, in terms of the posterior variances of Y_{1i}, to using ACS_{2008} versus Census 2000 data—the mean ratios are fairly concentrated around 0.98. It appears that the higher value of ρ when using the ACS_{2008} estimates is offset by their having non-negligible sampling variances at least for small states. The values for $R2_i$ and $R3_i$ show that, as expected, the model with $y_{2i} \sim \overline{\text{ACS}}_{05-08}$ performs best. Table D.2 in Appendix D provides corresponding tables giving summary statistics for $R1_i$, $R2_i$, and $R3_i$ for the models for the other three age groups of 0-4, 18-64, and 65+. These also show better performance for the models with $y_{2i} \sim \overline{\text{ACS}}_{05-08}$.

Table 7. Summary statistics of the variance ratios $R1_i$, $R2_i$, and $R3_i$ for bivariate models of 2009 state poverty rates of children age 5-17.

	$R1_i = \frac{\text{Var with ACS}_{2008}}{\text{Var with Cen 2000}}$	$R2_i = \frac{\text{Var with }\overline{\text{ACS}}_{05-08}}{\text{Var with Cen 2000}}$	$R3_i = \frac{\text{Var with }\overline{\text{ACS}}_{05-08}}{\text{Var with ACS}_{2008}}$
mean	0.98	0.82	0.84
minimum	0.90	0.70	0.72
Q1	0.95	0.77	0.79
median	0.96	0.82	0.84
Q3	0.99	0.86	0.90
maximum	1.08	0.96	0.99

Table 8 provides summary statistics of $\text{Var}(Y_{1i})$ for the direct ACS 2009 estimates, the SAIPE production model and the three bivariate models of the state 5-17 poverty ratio estimates. The table shows that all the models make substantial improvements in average state variances over the direct ACS estimates. For the largest states, however, such as California, modeling has little effect since for these states the ACS sampling variances are quite small. The variance improvement comes in the small states such as the District of Columbia (DC) and Wyoming (see note to Table 8).

Table 8. Summary statistics of $\text{Var}(Y_{1i})$ for direct ACS 2009 estimates, SAIPE production, and bivariate models of 2009 state poverty ratio estimates of children age 5-17.

	ACS 2009 direct est.	**SAIPE production**	**Variance with Census 2000**	**Variance with ACS_{2008}**	**Variance with $\overline{\text{ACS}}_{05-08}$**
mean	0.821	0.375	0.379	0.377	0.304
min (CA)	0.052	0.050	0.050	0.049	0.048
Q1	0.226	0.180	0.180	0.171	0.150
median	0.419	0.303	0.302	0.284	0.240
Q3	1.052	0.515	0.530	0.507	0.396
max1*	2.353	0.731	0.778	0.598	0.748
max2* (DC)	9.271	1.913	2.033	2.199	1.757

*Note: Starting in 2009, the SAIPE estimate of age 5-17 poverty for the District of Columbia (DC) is obtained from the county model for this age group, rather than from the state model. (Information about this change is available at http://www.census.gov/did/www/saipe/methods/09change.html.) For the table above, results were obtained from a state model that included DC, and DC had the maximum variance in all five columns. Given the production move of DC to the county model, the table also shows the entries with the maximum variances among the 50 states. Delaware had the maximum among the 50 state direct variance estimates, while the maximum in the other four columns was for Wyoming.

Table D.3 of Appendix D provides summary statistics of $\mathrm{Var}(Y_{1i})$ analogous to those of Table 8 for the models for the other three age groups (0-4,18-64, and 65+). It also shows that the bivariate model with $y_{2i} \sim \overline{\mathrm{ACS}}_{05-08}$ makes the largest variance improvements.

Note from Tables 8 and D.3 that the results from the SAIPE production model for age groups of 0-4, 5-17, and 18-64 are very similar to those from the respective bivariate models with y_{2i} corresponding to Census 2000 estimates. This is because for these three age groups the SAIPE production models use census residuals as a regression variable, thus making the models nearly equivalent to the corresponding bivariate models. For age 65+, the results for the SAIPE production model and the bivariate model with Census 2000 estimates are not as close, because for this age group the actual census poverty rate is used as a regression variable in the SAIPE state model instead of census residuals.

6. Summary

We have examined three bivariate models to assess the value of borrowing information from Census 2000, previous ACS single-year, or previous ACS multi-year estimates when using the models to predict true current year poverty. Direct ACS single-year estimates for the current year provided data for the other dependent variable in the bivariate model. We applied the bivariate models to county estimates of the log number of age 5-17 children in poverty, as well as to county estimates of age 5-17 child poverty rates, for 2009 and 2010. We also applied the bivariate models to state estimates of 5-17 child poverty rates for 2009. We then compared posterior variances of the current year true poverty measures from the alternative models. In every case the results showed that posterior variances from the bivariate model using previous ACS multi-year estimates were generally lower than, or at least not larger than, the posterior variances from the other two models. We thus conclude that use of previous ACS multi-year estimates in SAIPE models in place of Census 2000 estimates may improve results, particularly as time advances and the Census 2000 estimates thus become more and more distant from the current year. While the differences in posterior variances are not large, at least there is no indication from the results that this change would do worse. We also conclude that use of previous ACS multi-year estimates may yield better results than using the prior ACS single-year estimates. While the greater timeliness of the prior ACS single-year estimates can produce a higher correlation between the poverty characteristics estimated by the current year and previous estimates, the results suggest that this advantage tends to be offset by the higher sampling error of the single-year ACS estimates.

Our analyses, and hence our conclusions, were limited by the use of only two years (2009 and 2010) of ACS single-year estimates to provide the first dependent variable in our bivariate models. These were the only two years that we could reasonably use at this time in a model whose second dependent variable was provided by ACS multi-year estimates for a previous set of years. When the first equation was for 2009, we in fact stretched the definition of "multi-year" to include the four-year average of ACS estimates from 2005-2008. Despite this limitation, we think our basic conclusion that use of ACS multi-year estimates may yield better performance, and should at least not produce worse performance than the two alternatives, is solid given two considerations. First, we would not expect great differences in performance between using previous ACS multi-year versus single-year estimates, just some small to moderate differences driven by the trade-off between the slightly better timeliness of the single-year estimates versus the lower sampling error of the multi-year estimates. Second, we would expect the performance of models with Census 2000 estimates to deteriorate as they become more distant from the current year.

Our analyses leave open the question of possible benefits of using more than one of our alternative dependent variables in combination with current year ACS estimates in a trivariate model. We have started to examine this question and may report some results at a later date. Given the strong relations among the variables, however, and given our results, we doubt that a trivariate model will yield substantial gains, though such a model would be more complicated. Another possibility, a little more complicated, is to combine several years of ACS single-year estimates in a multivariate model with a time series structure. Note that use of previous ACS multi-year estimates gives equal weight to the data from each of the previous years that comprise the multi-year estimate. A time series perspective would suggest effectively downweighting the data somehow the further it is from the current year—see Bell (1998b). While it seems doubtful to us that this additional complexity would yield substantial improvements, this remains for now an open question.

References

Bell, William R. (1997), "A County CPS Model with 'Census Residuals'," internal note, Statistical Research Division, U.S. Census Bureau, July, 23, 1997.

Bell, William R. (1998a), "Models for County and State Poverty Estimates," Appendix A in *Small Area Estimates of School Age Children in Poverty, Interim Report 2: Evaluation of Revised 1993 County Estimates for Title I Allocations*, eds. Constance F. Citro and Graham Kalton, Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics, Washington, D.C.: National Academy Press, 95-108.

Bell, William R. (1998b) "Borrowing Information Over Time in Small Area Estimation: Thoughts with Reference to the American Community Survey," unpublished paper presented at the American Community Survey workshop, September 13, 1998, Michael Cohen (workshop study director), Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council, Washington, DC.

Bell, William R. (2008), "Examining Sensitivity of Small Area Inferences to Uncertainty About Sampling Error Variances," *Proceedings of the American Statistical Association, Survey Research Methods Section*, [CD-ROM], Alexandria, VA: American Statistical Association, 327-334.

Bell, William R., Basel, Wesley, Cruse, Craig, Dalzell, Lucinda, Maples, Jerry, O'Hara, Brett, and Powers, David (2007), "Use of ACS Data to Produce SAIPE Model-Based Estimates of Poverty for Counties," available at http://www.census.gov/did/www/saipe/publications/methods.html.

Fay, Robert E. and Herriot, Roger A. (1979), "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *Journal of the American Statistical Association*, **74**, 269-277.

Ghosh, Malay, Nangia, Narinder, and Kim, Dal Ho. (1996), "Estimation of Median Income of Four-Person Families: A Bayesian Time Series Approach," *Journal of the American Statistical Association*, 91, 1423-1431.

Huang, Elizabeth T. and Bell, William R. (2004), "An Empirical Study on Using ACS Supplementary Survey Data in SAIPE State Poverty Models". U.S. Census Bureau. http://www.census.gov/hhes/www/saipe/asapaper/HuangBell2004asa.pdf

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003), WinBUGS Users Manual, Version 1.4, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR,UK.

Plummer, Martyn (2010), "JAGS - Just Another Gibbs Sampler, JAGS 2.1.0.," May 12, 2010, available at https://sourceforge.net/projects/mcmc-jags.

APPENDIX A. Summary statistics by county population size groups of the variance ratios $R1_i$, $R2_i$, and $R3_i$ for bivariate models of log county poverty estimates of children age 5-17.

Table A.1: Results for y_{1i} ~ ACS_{2009}

Summary statistics for $R1_i$ = Var with ACS_{2008}/Var with Cen 2000

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,830	465	571	1,006	537	125	85	41
mean	1.01	1.05	1.06	1.03	0.96	0.88	0.83	0.84
min.	0.69	0.86	0.84	0.69	0.70	0.74	0.75	0.74
Q1	0.97	1.02	1.03	1.00	0.91	0.83	0.80	0.80
median	1.03	1.05	1.06	1.04	0.97	0.89	0.83	0.84
Q3	1.06	1.09	1.09	1.07	1.02	0.92	0.85	0.86
max.	1.19	1.18	1.19	1.19	1.07	1.07	1.04	0.94

Summary statistics for $R2_i$ = Var with $\overline{ACS}_{05\text{-}08}$/Var with Cen 2000

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,655	342	532	993	537	125	85	41
mean	0.88	0.94	0.93	0.89	0.81	0.77	0.76	0.83
min.	0.65	0.82	0.81	0.67	0.65	0.67	0.70	0.74
Q1	0.83	0.92	0.91	0.86	0.77	0.74	0.74	0.77
median	0.89	0.94	0.93	0.89	0.82	0.76	0.76	0.82
Q3	0.93	0.97	0.95	0.92	0.85	0.79	0.78	0.87
max.	1.05	1.05	1.05	1.02	0.97	0.89	0.87	0.96

Summary statistics for $R3_i$ = Var with $\overline{ACS}_{05\text{-}08}$/Var with ACS_{2008}

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,661	342	534	993	537	128	85	41
mean	0.87	0.89	0.88	0.86	0.85	0.88	0.92	0.99
min.	0.73	0.82	0.79	0.73	0.73	0.75	0.80	0.84
Q1	0.84	0.88	0.86	0.84	0.82	0.84	0.88	0.95
median	0.87	0.89	0.88	0.86	0.85	0.87	0.92	1.00
Q3	0.89	0.90	0.89	0.89	0.88	0.91	0.96	1.03
max.	1.15	1.01	1.15	1.04	1.06	1.01	1.05	1.05

Table A.2: Results for y_{1i} ~ ACS_{2010}

Summary statistics for $R1_i$ = Var with ACS_{2009}/Var with Cen 2000

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,853	461	573	1,019	543	130	88	39
mean	1.06	1.07	1.08	1.07	1.04	0.99	0.97	0.97
min.	0.94	0.95	0.98	0.95	0.95	0.94	0.94	0.94
Q1	1.03	1.04	1.05	1.05	1.01	0.97	0.96	0.95
median	1.06	1.07	1.08	1.07	1.04	0.99	0.97	0.96
Q3	1.09	1.10	1.11	1.10	1.07	1.01	0.98	0.97
max.	1.18	1.18	1.18	1.18	1.17	1.10	1.06	0.99

Summary statistics for $R2_i$ = Var with $ACS_{05\text{-}09}$/Var with Cen 2000

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,969	547	594	1028	543	130	88	39
mean	0.96	0.98	0.97	0.96	0.93	0.92	0.93	0.96
min.	0.84	0.87	0.87	0.86	0.84	0.86	0.88	0.91
Q1	0.93	0.96	0.94	0.93	0.91	0.91	0.92	0.94
median	0.95	0.98	0.97	0.95	0.93	0.92	0.93	0.96
Q3	0.98	1.00	0.99	0.98	0.94	0.93	0.94	0.97
max.	1.10	1.10	1.09	1.06	1.01	0.98	0.97	0.99

Summary statistics for $R3_i$ = Var with $ACS_{05\text{-}09}$/Var with ACS_{2009}

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,858	460	575	1019	543	133	89	39
mean	0.90	0.91	0.90	0.89	0.89	0.93	0.96	0.99
min.	0.78	0.79	0.78	0.78	0.79	0.84	0.87	0.92
Q1	0.88	0.89	0.87	0.87	0.87	0.91	0.94	0.98
median	0.90	0.92	0.90	0.89	0.89	0.93	0.97	1.00
Q3	0.93	0.94	0.92	0.91	0.91	0.95	0.98	1.00
max.	1.01	0.97	0.98	1.00	0.98	1.00	1.01	1.01

APPENDIX B. Summary statistics by county population size groups of the variance ratios $R1_i$, $R2_i$, and $R3_i$ for bivariate models of county poverty estimates of children age 5-17.

Table B.1: Results for y_{1i} ~ ACS_{2009}

Summary statistics for $R1_i$ = Var with ACS_{2008}/Var with Cen 2000

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,834	466	573	1,007	537	125	85	41
mean	0.90	0.95	0.93	0.89	0.84	0.86	0.89	0.93
min.	0.45	0.46	0.45	0.53	0.63	0.74	0.82	0.85
Q1	0.83	0.91	0.89	0.83	0.80	0.83	0.87	0.91
median	0.91	1.01	0.97	0.91	0.85	0.86	0.89	0.93
Q3	0.98	1.04	1.02	0.96	0.88	0.88	0.91	0.95
max.	1.10	1.10	1.08	1.07	0.99	0.99	0.98	0.98

Summary statistics for $R2_i$ = Var with $\overline{ACS}_{05\text{-}08}$/Var with Cen 2000

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,659	344	533	994	537	125	85	41
mean	0.91	0.95	0.93	0.90	0.87	0.89	0.92	0.95
min.	0.73	0.77	0.74	0.72	0.74	0.82	0.85	0.88
Q1	0.88	0.94	0.92	0.87	0.85	0.87	0.90	0.94
median	0.91	0.96	0.94	0.90	0.87	0.89	0.92	0.95
Q3	0.94	0.97	0.96	0.93	0.89	0.91	0.94	0.97
max.	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99

Summary statistics for $R3_i$ = Var with $\overline{ACS}_{05\text{-}08}$/Var with ACS_{2008}

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,664	343	535	994	537	128	86	41
mean	1.03	1.03	1.02	1.02	1.04	1.04	1.03	1.02
min.	0.85	0.87	0.86	0.85	0.92	1.00	1.00	1.01
Q1	0.95	0.92	0.94	0.95	1.00	1.02	1.03	1.02
median	1.01	0.95	0.97	0.99	1.03	1.04	1.03	1.02
Q3	1.05	1.04	1.03	1.05	1.07	1.05	1.04	1.03
max.	2.03	2.03	1.98	1.66	1.34	1.11	1.06	1.05

Table B.2: Results for y_{1i} ~ ACS_{2010}

Summary statistics for $R1_i$ = Var with ACS_{2009}/Var with Cen 2000

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,858	464	575	1,019	543	130	88	39
mean	0.92	0.96	0.94	0.91	0.88	0.90	0.93	0.95
min.	0.54	0.54	0.54	0.59	0.69	0.81	0.87	0.91
Q1	0.88	0.94	0.92	0.87	0.85	0.88	0.91	0.94
median	0.93	0.99	0.97	0.92	0.88	0.90	0.93	0.95
Q3	0.97	0.94	1.00	0.96	0.91	0.92	0.95	0.97
max.	1.06	1.06	1.05	1.04	0.99	0.97	0.97	0.99

Summary statistics for $R2_i$ = Var with $\mathrm{ACS}_{05\text{-}09}$/Var with Cen 2000

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,969	546	595	1028	543	130	88	39
mean	0.97	1.00	0.98	0.96	0.94	0.95	0.97	0.98
min.	0.80	0.80	0.83	0.82	0.87	0.91	0.94	0.96
Q1	0.94	0.98	0.97	0.94	0.92	0.94	0.96	0.98
median	0.97	1.00	0.99	0.96	0.94	0.95	0.97	0.98
Q3	1.00	1.03	1.01	0.98	0.95	0.96	0.98	0.99
max.	1.08	1.08	1.06	1.05	0.99	0.99	0.99	1.00

Summary statistics for $R3_i$ = Var with $\mathrm{ACS}_{05\text{-}09}$/Var with ACS_{2009}

	All	<10k	10-20k	20-65k	65-250k	250k-500k	500k-1000k	>1000k
# counties	2,864	464	577	1019	543	133	89	39
mean	1.06	1.05	1.06	1.06	1.07	1.06	1.04	1.03
min.	0.86	0.86	0.92	0.92	0.98	1.02	1.02	1.01
Q1	1.01	1.00	1.00	1.01	1.04	1.05	1.03	1.02
median	1.04	1.02	1.03	1.04	1.06	1.06	1.05	1.03
Q3	1.08	1.06	1.07	1.09	1.09	1.08	1.05	1.04
max.	1.79	1.78	1.79	1.58	1.31	1.14	1.08	1.06

APPENDIX C. The five lowest $R1_i$ and the corresponding $R3_i$ in 4 small county population size groups from 2009 county poverty rate bivariate models

2009 pop	$R1_i$	$R3_i$	**ACS 2008 Pov rate**	**ACS 2008 Pov rate Var**	**ACS 2009 Pov rate**	**ACS 2009 Pov rate Var**
	0.46	**2.03**	0.00668	0.000103	0.18	0.0164
	0.47	2.00	0.00325	0.000024	0.16	0.0226
< 10K	0.50	1.67	0.01650	0.000322	0.30	0.0307
	0.52	1.66	0.01347	0.000180	0.09	0.0078
	0.52	1.55	0.01354	0.000189	0.12	0.0074
	0.45	1.66	0.00556	0.000031	0.23	0.0144
	0.45	**1.98**	0.00277	0.000011	0.21	0.0115
10 – 20K	0.48	1.95	0.00905	0.000096	0.27	0.0147
	0.49	1.89	0.00626	0.000044	0.11	0.0067
	0.51	1.65	0.01138	0.000158	0.17	0.0072
	0.53	1.65	0.00962	0.000068	0.21	0.0051
	0.56	**1.66**	0.02597	0.000371	0.27	0.0111
20 – 65K	0.56	1.59	0.01451	0.000262	0.18	0.0062
	0.57	1.30	0.03065	0.000395	0.11	0.0073
	0.57	1.44	0.00628	0.000048	0.09	0.0027
	0.63	**1.34**	0.05474	0.000407	0.22	0.0039
	0.63	1.16	0.09470	0.000408	0.17	0.0030
65 – 250K	0.66	1.23	0.05891	0.000277	0.20	0.0022
	0.66	1.27	0.04232	0.000476	0.14	0.0028
	0.66	1.24	0.08035	0.000269	0.20	0.0019

APPENDIX D. Results for bivariate models of 2009 state poverty rates for four age groups

Table D.1 Posterior means of parameters (and standard errors of ρ)

Age 0-4

$y_{2i} \Rightarrow$	**Cen 2000**	$\overline{\textbf{ACS}}_{\mathbf{05-08}}$	$\text{ACS}_{\mathbf{2008}}$
ρ	0.68 (0.11)	0.90 (0.06)	0.92 (0.06)
s_{11}	4.62	3.83	3.83
s_{22}	2.48	2.57	2.59

Age 5-17

$y_{2i} \Rightarrow$	**Cen 2000**	$\overline{\textbf{ACS}}_{\mathbf{05-08}}$	$\text{ACS}_{\mathbf{2008}}$
ρ	0.65 (0.12)	0.86 (0.07)	0.77 (0.10)
s_{11}	1.62	1.59	1.52
s_{22}	1.75	1.54	1.67

Age 18-64

$y_{2i} \Rightarrow$	**Cen 2000**	$\overline{\textbf{ACS}}_{\mathbf{05-08}}$	$\text{ACS}_{\mathbf{2008}}$
ρ	0.67 (0.10)	0.93 (0.06)	0.87 (0.06)
s_{11}	1.16	1.13	1.05
s_{22}	0.62	0.65	0.65

Age 65+

$y_{2i} \Rightarrow$	**Cen 2000**	$\overline{\textbf{ACS}}_{\mathbf{05-08}}$	$\text{ACS}_{\mathbf{2008}}$
ρ	0.88 (0.05)	0.99 (0.01)	0.98 (0.02)
s_{11}	1.56	1.59	1.62
s_{22}	2.01	2.16	2.33

Table D.2 Summary statistics of the variance ratios $R1_i$, $R2_i$, and $R3_i$ for three age groups (results for age 5-17 are given in Table 7).

Age 0-4

	$R1_i = \frac{\text{Var with } ACS_{2008}}{\text{Var with Cen 2000}}$	$R2_i = \frac{\text{Var with } \overline{ACS}_{05-08}}{\text{Var with Cen 2000}}$	$R3_i = \frac{\text{Var with } \overline{ACS}_{05-08}}{\text{Var with } ACS_{2008}}$
mean	0.89	0.76	0.86
minimum	0.78	0.65	0.69
Q1	0.86	0.71	0.78
median	0.89	0.71	0.84
Q3	0.91	0.81	0.92
maximum	1.04	0.94	1.03

Age 18-64

	$R1_i$	$R2_i$	$R3_i$
mean	0.92	0.79	0.86
minimum	0.84	0.62	0.69
Q1	0.90	0.72	0.80
median	0.92	0.79	0.87
Q3	0.95	0.87	0.93
maximum	0.98	0.97	0.98

Age 65+

	$R1_i$	$R2_i$	$R3_i$
mean	0.82	0.55	0.68
minimum	0.67	0.43	0.46
Q1	0.74	0.50	0.59
median	0.78	0.53	0.67
Q3	0.85	0.59	0.76
maximum	1.13	0.94	0.98

Table D.3 Summary statistics of Var(Y_{1i}) for the direct ACS 2009 estimates, the SAIPE production model, and bivariate models of the estimates for three age groups (results for age 5-17 are given in Table 8).

Age 0-4

	ACS 2009 direct est.	SAIPE production	Variance with Census 2000	Variance with ACS_{2008}	Variance with $\overline{ACS}_{05-08}$
mean	1.612	0.896	0.894	0.808	0.660
min (CA)	0.095	0.092	0.092	0.087	0.087
Q1	0.458	0.397	0.396	0.342	0.319
median	0.989	0.752	0.742	0.646	0.543
Q3	2.139	1.215	1.217	1.070	0.892
max (DC)	7.570	3.578	3.677	3.424	2.878

Age 18-64

	ACS 2009 direct est.	SAIPE production	Variance with Census 2000	Variance with ACS_{2008}	Variance with $\overline{ACS}_{05-08}$
mean	0.128	0.0997	0.09996	0.090	0.073
min (CA)	0.010	0.0099	0.0099	0.0097	0.0096
Q1	0.039	0.037	0.0366	0.034	0.0318
median	0.081	0.072	0.072	0.066	0.056
Q3	0.162	0.134	0.134	0.125	0.100
max (DC)	0.724	0.490	0.499	0.440	0.333

Age 65+

	ACS 2009 direct est.	SAIPE production	Variance with Census 2000	Variance with ACS_{2008}	Variance with $\overline{ACS}_{05-08}$
mean	0.261	0.151	0.141	0.122	0.076
min (CA)	0.021	0.020	0.020	0.020	0.019
Q1	0.084	0.071	0.068	0.050	0.040
median	0.150	0.115	0.109	0.088	0.055
Q3	0.376	0.232	0.203	0.160	0.095
max (DC)	1.932	0.516	0.580	0.654	0.391