# A Bayesian Zero-One Inflated Beta Model for Small Area Shrinkage Estimation *

Jerzy Wieczorek† Ciara Nugent Sam Hawala‡

**Abstract**
We evaluate revisions to a Bayesian beta regression model proposed in Wieczorek and Hawala (2011), for U.S. county poverty rates. For small areas, some of which have survey estimates of poverty rates of 0 or 1, a zero-one inflated rate model extends the beta distribution to allow for these extreme estimates. The addition of a model error term allows the model to produce shrinkage estimates. We can estimate the model parameters and shrinkage estimates for the small areas via Bayesian computation techniques. Using simulated draws from a "pseudo population" based on American Community Survey (ACS) data, we compare the results to ACS-like direct estimates and to the Census Bureau's current small-area model for county poverty estimation.

**Key Words:** small area estimates, SAIPE, MCMC, beta regression, pseudo population

## 1. Introduction

The Small Area Income and Poverty Estimates (SAIPE) program at the U.S. Census Bureau uses small area estimation techniques to create model-based estimates of selected poverty and income statistics on an annual basis. The estimates are intended to be more timely than direct estimates from the decennial census or five-year American Community Survey (ACS), as well as more precise and stable than single-year ACS direct estimates for small areas.

In this paper, we are concerned with estimating the number of related poor children aged 5-17 in U.S. counties. Poverty status is determined by comparing the child's family income to thresholds that vary with family size and composition; these thresholds are described further in U.S. Census Bureau (2012). These estimates are provided to the Department of Education and used in the allocation of federal funding to local programs.

The existing county-level approach is based on a Fay-Herriot "log-level" model, i.e. a model on the natural log of the number of related poor children in each area. The model combines single-year ACS direct estimates with regression predictors from administrative data records including Internal Revenue Service (IRS) tax data and Supplemental Nutrition Assistance Program (SNAP) (formerly "food stamp") data. The data inputs and this Fay-Herriot model are described in more detail on the SAIPE website (U.S. Census Bureau, 2010 and 2011). The current SAIPE model is tractable and well-established, but it is worth considering alternative models that may have advantages over the current approach.

In particular, some counties have ACS direct estimates of zero related children in poverty. Since $\log(0)$ is undefined, these counties must be dropped from the estimation procedure, with a resulting loss of information and efficiency. During four out of the five years from 2005 to 2009, over 5% of counties had ACS direct estimates of zero related children in poverty.

---

*This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

†Contact author: jerzy.wieczorek@census.gov, Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

‡Social, Economic, and Housing Statistics Division, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Furthermore, Census Bureau staff have found other concerns with the log-level model, including biased direct variance estimates on the log scale (Huang and Bell, 2009), and have suggested modeling poverty rates rather than log poverty counts.

Wieczorek and Hawala (2011) proposed to account for both of these issues with a zero-one inflated beta (ZOIB) regression model. The beta distribution allows us to model poverty rates directly on a continuous range between 0 and 1 exclusive; and a multinomial component allows us to account for cases where the ACS rate estimate is either 0 or 1. The sampling model is equivalent to the inflated beta distributions of Ospina and Ferrari (2010). The present paper extends the 2011 work by adding random effects to the linking model. The ZOIB model details are discussed in section 2.

Hawala and Lahiri (2012) and Liu at al. (2007) compare the Fay-Herriot and beta-logistic regression models on rates in (0,1). However, they do not account for direct estimates of exactly 0 or 1, unlike the ZOIB model presented here. Bauder, Luery, and Szelepka (2012) do account for 0 and 1 estimates, with a related but distinct model that they call a "three-part model" after Pfeffermann et al. (2008), whose two-part unit-level model also combines 0s with continuous responses in the data but is different from the area-level approach pursued here.

The ZOIB hierarchical model is difficult to fit by classical/analytical methods but lends itself well to Bayesian treatment by Markov Chain Monte Carlo (MCMC) methods. Furthermore, posterior distributions allow for useful model-checking approaches. The computational details are discussed in section 3.

The models are evaluated via a simulation study, described in section 4. Finally, section 5 presents the evaluation results. We conduct a simulation study, replicating the ACS sampling process to form draws from a "pseudo population" based on real data. On each sample, we evaluate the direct, Fay-Herriot, and ZOIB estimators of the poverty rate, and compare their performance against the pseudo-population truth:

- We evaluate whether each model's assumptions hold. Do the sampling distributions and linking models assumed for each approach (Fay-Herriot or ZOIB) appear appropriate?
- Following the advice of Little (2006), we check whether the estimates from each approach (direct, Fay-Herriot, and ZOIB) are calibrated to the true values from the pseudo population. We compare biases, variances, and mean square errors (MSEs) of the point estimates, as well as biases of their estimated MSEs. We also check coverage rates for confidence or credible intervals (CIs).

## 2. Models

Section 2.1 reviews the Fay-Herriot-like model currently in use by the SAIPE program. Section 2.2 describes the zero-one inflated beta (ZOIB) model where observed poverty rates can be modeled by a modified beta regression which also allows for observed rates of 0 or 1.

### 2.1 Fay-Herriot model

The SAIPE program is built around the model proposed by Fay and Herriot (1979). Both the sampling and linking models use log counts. In the sampling model, the sample estimates of log counts of persons in poverty are normally distributed around the true log counts. The linking model connects these true log counts to a regression on log-level auxiliary data. (Normality is not required to estimate a Fay-Herriot model in general using the empirical best linear unbiased predictor (EBLUP) approach. However, for purposes of this paper, we do assume normality in order to compare a fully specified Fay-Herriot model to the ZOIB model presented below. Also, the normal assumption is used when deriving the log-bias correction factor for transforming estimates back from the log scale to the native scale.)

Let $Y_i$ denote the true poverty count in county $i$, and let $y_i$ be the design consistent survey estimate of $Y_i$. Then the sampling and linking models are:

$$\log(y_i) \sim N(\log(Y_i), \sigma_i^2)$$
$$\log(Y_i) \sim N(\log(\boldsymbol{x_i})'\boldsymbol{\beta}_{(FH)}, \sigma^2_{u(FH)})$$

Here, $\boldsymbol{x_i}$ is a vector of area-level adminstrative data (log tax poverty rate, log tax nonfiler rate, and log SNAP participation rate), $\boldsymbol{\beta}_{(FH)}$ is a vector of regression coefficients, $\sigma_i^2$ is the sampling variance in county $i$, and $\sigma^2_{u(FH)}$ is the regression model error variance, assumed to be constant across all areas. The $\sigma_i^2$ are treated as known but $\sigma^2_{u(FH)}$ and $\boldsymbol{\beta}_{(FH)}$ have to be estimated, the former using an iterative approach from the Fay-Herriot paper and the latter using weighted least squares conditional on $\hat{\sigma}^2_{u(FH)}$. (Note that these parameters are estimated using frequentist methods, not the fully Bayesian approach described for the ZOIB model below.)

The regression estimates and direct estimates are composed into a shrinkage estimator:

$$\hat{w}_i = \sigma_i^2/(\sigma_i^2 + \hat{\sigma}^2_{u(FH)})$$
$$\widehat{\log(Y_i)} = \hat{w}_i \log(\boldsymbol{x_i})'\hat{\boldsymbol{\beta}}_{(FH)} + (1 - \hat{w}_i)\log(y_i)$$
$$\widehat{MSE}(\widehat{\log(Y_i)}) = \hat{w}_i \hat{\sigma}^2_{u(FH)} + \hat{w}_i^2 \widehat{Var}(\log(\boldsymbol{x_i})'\hat{\boldsymbol{\beta}}_{(FH)})$$

At this point, the estimates are converted from the log scale to the original count scale, using a log-bias correction factor. Slud and Maiti (2006) give a second-order correction for the log-bias in such scenarios, although we use a simpler first-order correction factor in the SAIPE program and consequently in this paper as well. Next the point estimates and their MSE estimates are adjusted for raking, and finally the counts are converted to rates by treating the poverty universe population estimates as known. More thorough details are given in the SAIPE documentation (U.S. Census Bureau, 2010).

In the simulations for the present paper, the raking step is omitted but the remainder of the estimation procedure closely follows the SAIPE approach. In particular, although we begin with estimates of the log counts, we convert to estimated native-scale counts and then to rates in order to allow comparison to the ZOIB model estimates. This SAIPE-like version of the Fay-Herriot model will be referred to as “the FH model” in the rest of this paper.

### 2.2 Zero-one inflated beta model

Let $Z_i$ denote the true poverty rate in county $i$, and let $z_i$ be the design consistent survey estimate of $Z_i$. For $z_i$ we use

$$z_i = \frac{\text{ACS estimated number in poverty}}{\text{ACS estimated number in poverty universe}}$$

The sampling model is a zero-one inflated beta distribution as given by Ospina and Ferrari (2010), who show that the inflated beta distributions are three- or four-parameter exponential family distributions of full rank. This sampling model can also be thought of as a multinomial trial with three possible results: a 0, a 1, or a draw from a beta distribution.

Let $p_i^{(0)}$ be the probability that county $i$ has an observed direct estimate of 0, and similarly let $p_i^{(1)}$ be the probability of observing 1. Otherwise, with probability $(1 - p_i^{(0)} - p_i^{(1)})$, the county's rate is drawn from the $\mathrm{Beta}(a_i, b_i)$ distribution.

$$z_i = \begin{cases} 0 & \text{with probability } p_i^{(0)}, \\ 1 & \text{with probability } p_i^{(1)}, \\ \sim \mathrm{Beta}(a_i, b_i) & \text{with probability } (1 - p_i^{(0)} - p_i^{(1)}). \end{cases}$$

In order to incorporate covariate information into a beta regression model, we reparameterize the beta family in terms of its mean and a parameter related to its variance: $\mu_i = a_i/(a_i + b_i)$ and $\gamma_i = a_i + b_i$.

In Ospina and Ferrari's (2010) notation, this sampling model is equivalent to

$$z_i \sim \mathrm{BEINF}(\alpha = p_i^{(0)} + p_i^{(1)}, \gamma = \frac{p_i^{(1)}}{p_i^{(0)} + p_i^{(1)}}, \mu = \mu_i, \phi = \gamma_i)$$

We assume that the ACS direct estimate is unbiased for the truth, so the target small area value $Z_i$ is the expectation of the BEINF sampling distribution:

$$Z_i = E(z_i|\mu_i, p_i^{(0)}, p_i^{(1)}) = (1 - p_i^{(0)} - p_i^{(1)})\mu_i + p_i^{(1)}$$

In our fully Bayes approach, our estimator $\hat{Z}_i$ is the posterior mean $E(Z_i|z_i)$.

We use "BEINF" to refer to the zero-inflated beta *distribution*, i.e. the sampling distribution used in this model. We use "ZOIB" to refer to the zero-one inflated beta *hierarchical model*, i.e. the small area model proposed here with BEINF as the sampling distribution and with the linking model as given below.

Our approximate estimates of $\gamma_i$ come from writing the variance of a (non-inflated) beta distribution as the variance of a rate estimate scaled by an effective sample size:

$$X \sim Beta(a_i, b_i), \quad Var(X|a_i, b_i) = \frac{\mu_i(1-\mu_i)}{\gamma_i + 1} = \frac{\mu_i(1-\mu_i)}{\mathrm{neff}_i}, \quad \gamma_i = \mathrm{neff}_i - 1$$

We approximate $\mathrm{neff}_i$ via a variance function suggested by Hawala and Lahiri (2010):

$$Var(N_i\hat{\mu}_i) \approx \mu_i(1-\mu_i)d_i \quad \text{where} \quad d_i = \sum_{h\in i}(\sum_{c\in h} w_{ihc})^2$$

where $w_{ihc}$ is the survey weight of child $c$ in household $h$ in county $i$, and $N_i$ is the total population of the poverty universe in county $i$. This suggests a heuristic estimator of $\mathrm{neff}_i$:

$$\mu_i(1-\mu_i)d_i \approx N_i^2 \frac{\mu_i(1-\mu_i)}{\mathrm{neff}_i} \quad \text{so} \quad \widehat{\mathrm{neff}}_i \approx \frac{N_i^2}{d_i}$$

Much like the Fay-Herriot model, the ZOIB model currently treats sampling variance component estimates $\hat{\gamma}_i = \widehat{\mathrm{neff}}_i - 1$ as known. The $\mu_i$ are assigned a regression linking model on their logits.

In the main linking model, the logit of the mean of the beta distribution follows a linear model with a common model-error variance $\sigma^2_{u(ZOIB)}$ across all areas. (The addition of the random effect is new relative to the model in Wieczorek and Hawala (2011).) $\boldsymbol{\beta_\mu}$ is the vector of regression coefficients for estimating $\mathrm{logit}(\mu_i)$.

$$\mathrm{logit}(\mu_i) \sim N(\boldsymbol{x_i}'\boldsymbol{\beta_\mu}, \sigma^2_{u(ZOIB)})$$

There are also linking models for the logits of $p_i^{(0)}$ and $p_i^{(1)}$, using $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$; however, they do not yet incorporate random effects and are strictly synthetic:

$$\text{logit}(p_i^{(0)}) = \boldsymbol{x_i}'\boldsymbol{\beta_0} \quad \text{and} \quad \text{logit}(p_i^{(1)}) = \boldsymbol{x_i}'\boldsymbol{\beta_1}$$

In future work we intend to evaluate other linking models for these probabilities, including a comparison with the Small Area Health Insurance Estimates (SAHIE) program model described in Bauder, Luery, and Szelepka (2012).

We use an intercept term and the same three regressors as for the FH model (tax poverty rate, tax non-filing rate, and SNAP participation rate) for $\boldsymbol{\beta_\mu}$. We add a fourth regressor, the natural log of ACS sample size in poverty universe, for each of $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$. We expect that the coefficients of sample size will be negative, so that large samples have lower probabilities of 0 and 1.

Let $\mathbf{x}_i = (1, x_{i1}, x_{i2}, x_{i3})$ be the vector containing 1 and the three main regressors for each of $m$ small areas, $i = 1, \ldots, m$. Let $\mathbf{x}_{i+}$ denote this vector augmented with the last regressor. The combined zero-one inflated Bayesian beta regression model we consider is

$$\begin{aligned}
z_i|\mu_i, p_i^{(0)}, p_i^{(1)} &\sim \text{BEINF}(p_i^{(0)} + p_i^{(1)}, \frac{p_i^{(1)}}{p_i^{(0)} + p_i^{(1)}}, \mu_i, \gamma_i) \\
\text{logit}(\mu_i)|\boldsymbol{\beta_\mu}, \sigma^2_{u(ZOIB)} &\sim N(\mathbf{x}_i'\boldsymbol{\beta_\mu}, \sigma^2_{u(ZOIB)}) \\
\text{logit}(p_i^{(0)})|\boldsymbol{\beta_0} &= \mathbf{x}_{i+}'\boldsymbol{\beta_0} \\
\text{logit}(p_i^{(1)})|\boldsymbol{\beta_1} &= \mathbf{x}_{i+}'\boldsymbol{\beta_1} \\
\boldsymbol{\beta_\mu}, \boldsymbol{\beta_0}, \boldsymbol{\beta_1} &\sim N(\mathbf{0}, 1000^2\mathbf{I}) \\
\sigma^2_{u(ZOIB)} &\sim \text{Unif}(0, 300)
\end{aligned}$$

where $\mathbf{I}$ denotes the identity matrix (of size $4 + 5 + 5 = 14$ in this case). We place diffuse normal priors on each regression coefficient and a diffuse uniform prior on the model error variance, and we treat $\gamma_i$ as known. (An alternate approach to try in the future will be to model the conditional logit of $Z_i$ rather than of $\mu_i$.)

## 3. Computational Approach

In our MCMC procedure we draw samples $j = 1, ..., J$ from the posteriors of each parameter. Letting tildes denote these sampled posterior draws, we have

$$\begin{aligned}
\tilde{Z}_{ij} &= \tilde{p}_{ij}^{(01)}\tilde{\mu}_{ij} + \tilde{p}_{ij}^{(1)} \\
\hat{Z}_i &= J^{-1}\sum_{j=1}^{J}\tilde{Z}_{ij}
\end{aligned} \tag{1}$$

For each dataset, we use the R software (R Development Core Team, 2012) to call the JAGS software (Plummer, 2003) to generate a MCMC sample from the joint posterior distribution of all the $Z_i$.

We run several chains at once (starting with several overdispersed initial values). This way, we can evaluate the convergence for each parameter by checking whether its multiple chains are converging to the same distribution, using the potential scale reduction factor $\hat{R}$ (Gelman et al., 2004, p. 297).

## 4. Simulation study

This section summarizes the setup for a simulation study designed to evaluate the poverty estimates from our ZOIB model and from the FH model. Subsection 4.1 explains how the simulation study datasets were generated. Subsection 4.2 summarizes the experimental design used for the following evaluations. The results are presented in Section 5.

### 4.1 Overview of pseudo population simulation study

In describing models for Census data, Little (2006) suggests that "the model assessment component would be helped by building research pseudo-populations of records from earlier censuses that form the basis for simulation assessments of different model procedures." These remarks are in the context of a discussion about "Calibrated Bayes," i.e., assessing Bayesian models by frequentist standards. We follow this advice by creating a pseudo population from which we simulate draws of replicate subsample datasets, which can be used both for assessing model assumptions and for comparing the estimates from competing models.

We used the 2010 single-year ACS data for the 100 largest counties as a pseudo population. In other words, the sampled housing units and persons in county $i$ were treated as a complete pseudo population for that county, and the poverty count and rate observed in that ACS sample (among related children ages 5-17) were treated as the true count and rate for that pseudo population. (In what follows, "true" will be used to denote these known pseudo population counts and rates, not the unknown values in the real-world population.)

The pseudo population sizes of the 100 largest counties ranged from approximately 2,000 to 40,000 housing units. The actual population sizes of many small U.S. counties fall into this range, so this seemed to be a reasonable pseudo population for the purpose of testing small area models.

In each of the 100 counties, we drew $R = 100$ independent samples or replicates. (Note that the term "replicates" here is unrelated to the "replicate weights" used by the ACS-like weighting process within any one sample.) We followed a simplification of the ACS methodology for sampling, weighting the results, setting replicate weights, and calculating the poverty estimates and their estimated sampling variances. The ACS process is documented by U.S. Census Bureau (2009), particularly in chapters 4, 11, and 12. Full details of this pseudo population sampling process will be available in a forthcoming technical report by Wieczorek (2012). Across these $100 * 100$ county-by-replicate combinations, the resulting replicate sample sizes (number of sampled children in the poverty universe, not of all persons) ranged from 4 to 404, with a median of 36 children sampled in a single replicate for a single county.

We used SAS software (SAS Institute, 2008) to gather all the data, draw samples, construct weights, and store the preliminary ACS-like direct estimates of child poverty counts, log counts, rates, and their sampling variances, as well as the effective sample size estimates $\widehat{\mathrm{neff}}_i$ as given in section 2.2. At the same time we stored the true pseudo population poverty counts and rates for each county.

The resulting dataset was processed further in R and JAGS to get the Empirical Bayes estimates from the FH model and the posterior estimates from the ZOIB model. Finally, for each of the 100 replicates or samples from each county, the FH and ZOIB model estimates across all 100 counties were compared to each other and to the ACS-like direct estimates, as described in the following sections.

The same set of predictor variables were used for each sample, based on the 2010 SAIPE auxiliary data for the selected 100 counties.

We will use the term "MC estimate" to denote the Monte Carlo (MC) estimates of bias, variance, MSE, and CI coverage that we get by aggregating the $R = 100$ replicate estimates for each estimator and each county.

Let MSE (uppercase) denote our best estimate of the true mean squared error, from seeing the variation in point estimates across all 100 replicate samples; and let mse (lowercase) denote each replicate's "internal" estimate of its point estimate's mean squared error, from seeing only one replicate at a time. Looking at MSE tells us how good the point estimates are; looking at the bias of mse tells us how good the mse estimates are.

Letting r index the replicates and i index the counties, our metrics of interest are:

$$\begin{aligned}
\mathrm{Bias}_{\mathrm{MC}}(\hat{Z}_i) &= \frac{1}{R}\sum_r (\hat{Z}_{ir} - Z_i) = \bar{Z}_i - Z_i \\
\mathrm{Var}_{\mathrm{MC}}(\hat{Z}_i) &= \frac{1}{R}\sum_r (\hat{Z}_{ir} - \bar{Z}_i)^2 \\
\mathrm{MSE}_{\mathrm{MC}}(\hat{Z}_i) &= \frac{1}{R}\sum_r (\hat{Z}_{ir} - Z_i)^2 \qquad (2) \\
\mathrm{Bias}_{\mathrm{MC}}(\mathrm{mse}(\hat{Z}_i)) &= \frac{1}{R}\sum_r (\mathrm{mse}(\hat{Z}_{ir}) - \mathrm{MSE}_{\mathrm{MC}}(\hat{Z}_i)) \\
\mathrm{Coverage}_{\mathrm{MC}}(\hat{Z}_i) &= \frac{1}{R}\sum_r (I\{(\widehat{CI}_{ir}^{low} < Z_i < \widehat{CI}_{ir}^{high})\})
\end{aligned}$$

where each $\hat{Z}_{ir}$, $\mathrm{mse}(\hat{Z}_{ir})$, and the CI endpoints are computed as appropriate depending on the estimator (Direct, FH, or ZOIB). Again, the FH estimates are transformed from log counts to rates so that all three sets of estimates are directly comparable.

- The Direct $\mathrm{mse}(\hat{Z}_{ir})$ estimates are the sampling variances estimated using replicate weights, just as in the ACS. (Since the point estimates are assumed to be unbiased, the sampling variance and mse estimates are the same.) Under the FH model, the $\mathrm{mse}(\hat{Z}_{ir})$ estimates are estimated as given in section 2.1. The ZOIB estimates of $\mathrm{mse}(\hat{Z}_{ir})$ are the posterior variances of the MCMC draws for each $\hat{Z}_{ir}$.

- Both the Direct and FH CIs are 90% confidence interval estimates calculated as $\hat{Z}_{ir} \pm 1.645\sqrt{\mathrm{mse}(\hat{Z}_{ir})}$, while the ZOIB CIs are Bayesian equal-tail 90% credible intervals formed by taking the $5^{th}$ and $95^{th}$ percentiles of the MCMC draws for each $\hat{Z}_{ir}$.

Two other issues arose with the particular set of replicate samples drawn during this analysis:

- For the Direct estimates, $\mathrm{mse}(\hat{Z}_{ir})$ is treated as undefined when there were no poor children in that sample for that county. (This occurred at least once in 77 of the 100 counties. Only a quarter of the counties had more than 15 replicates with no poor children; the county with the maximum had 56 no-poor-child replicates.) Therefore, for each county's Direct estimates, $\mathrm{Bias}_{\mathrm{MC}}(\mathrm{mse}(\hat{Z}_i))$ is estimated only from the replicates with poor in the sample; the remaining replicates are dropped and the denominator is reduced appropriately from $R$ to the number of replicates actually used.

- None of the counties happened to have a replicate sample in which all sampled children were in poverty. Therefore, the data was unsuitable for the full ZOIB model. Instead, a restricted version was used, by setting $p_i^{(1)} \equiv 0$ for all $i$. This could be termed a zero-inflated beta (ZIB) model, but we continue to use the term "ZOIB" in the evaluations below since they do not rely on this particular aspect of the model.

### 4.2 Experimental Design

In summary: The pseudo population has 100 areas (counties). We create 100 replicate subsample datasets, each containing a subsample from each of the hundred counties. Taking each of the 100 replicate datasets one at a time, we estimate each county's poverty rate and its mse using all three estimators: Direct, FH, and ZOIB. Thus we have 100 replicates $*$ 100 counties $*$ 2 estimates $*$ 3 estimators, as well as 100 true poverty rates (one per county) calculated directly from the pseudo population. Then:

- In Section 5.1 we look at the distribution of Direct estimates across replicates, within each county, to check the sampling models; and we look at the distribution of true values across counties (no replicates involved) to check the linking models.
- In Section 5.2 we aggregate the information across replicates to compute the 5 metrics (from Equation 2) for each county and each estimator (Direct, FH, ZOIB), leading to 100 counties $*$ 3 estimators $*$ 5 metrics.

## 5. Results

This section presents the results of the simulation study set up in Section 4. Subsection 5.1 describes our checks on the assumptions for the FH and ZOIB models. Section 5.2 evaluates the calibration of all three estimators: Direct, FH, and ZOIB.

### 5.1 Internal checks

We used several checks for internal consistency to evaluate whether each model's assumptions hold.

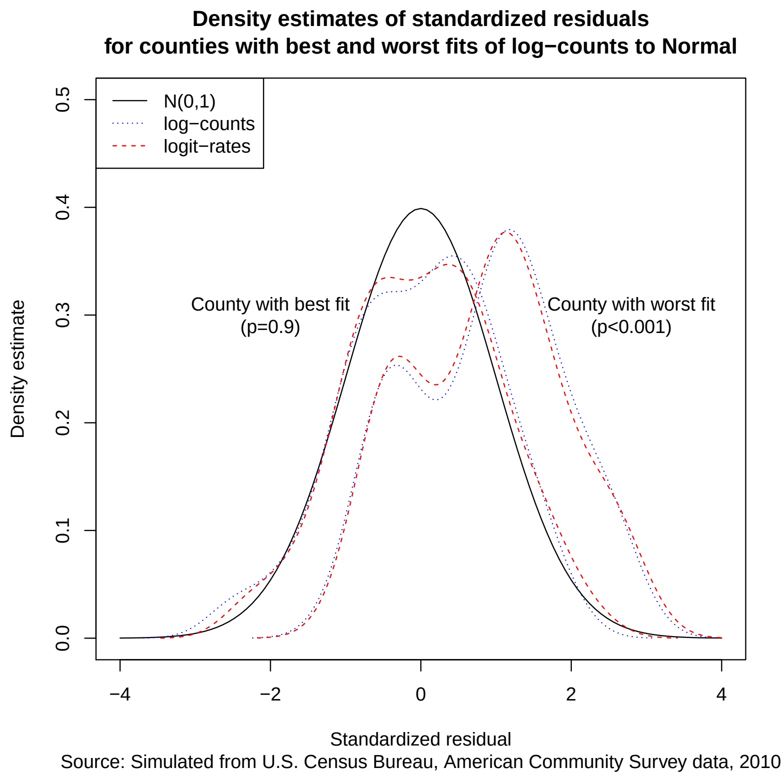
Firstly, are the sampling models appropriate, at least for the non-zero estimates? For the FH model, are the sample log count estimates approximately Normal around the true log poverty counts? For the ZOIB model, are the sample poverty rate estimates approximately Beta-distributed around the true poverty rates?

Secondly, are the linking models appropriate? Are the true log counts (for the FH model) or the logits of the true rates (for the ZOIB model) approximately Normal with constant variance, around a regression line with the given predictor variables? (Note that under the complicated structure of the ZOIB model, the true poverty rates are not really the same as the mean of the Beta distribution. However, treating them as if they were is a reasonable start for diagnostic model-checking purposes.)

The FH sampling-model check is straightforward. For each county, take all 100 samples; subtract off the true pseudo population log count; divide by the MC standard error across the 100 samples; and see if these residuals look approximately $N(0,1)$. The ZOIB sampling-model check is a bit more complicated. For data from a Beta distribution, neither the residuals nor the standardized residuals are necessarily expected to look approximately Normal. However, Espinheira et al. (2008) suggest that Beta residuals should be more approximately Normal on the logit scale than on the rate scale; the logit scale is also more appropriate for detecting outliers. Figure 1 illustrates this approach by comparing the sample densities for standardized log-count and logit-rate residuals to the standard Normal density, for two example counties. (These are the best- and worst-fitting county, according to p-values from Kolmogorov-Smirnov tests on the standardized log-count residuals, as described below.) Both assumed sampling distributions are clearly appropriate for one county but not for the other.

However, plotting every county's distribution of residuals would be difficult to review succinctly. Instead, we summarize the p-values from performing a Kolmogorov-Smirnov

**Figure 1**: Two sample densities compared to N(0,1)




(KS) test of normality on each county's across-samples set of residuals. For the FH model, we test whether log-counts are approximately Normal. For the ZOIB model, we test whether rates are approximately Beta, and also (following Espinheira et al. 2008) whether logit-rates are approximately Normal. (We use the KS test, rather than Lilliefors test, in order to allow comparisons to the Beta distribution as well as the Normal.)

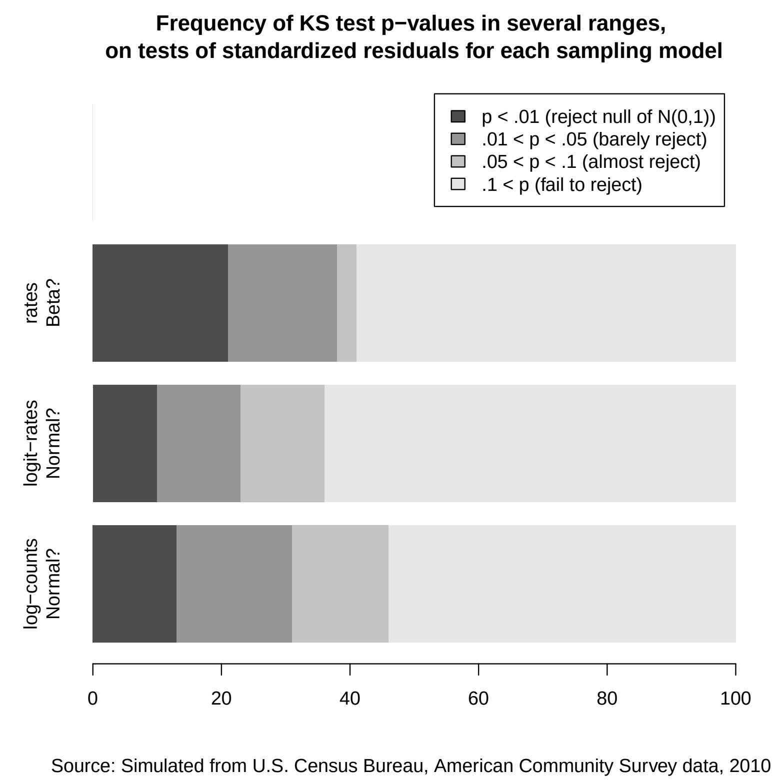
For the Normal comparisons, we take each county's samples' log-count and logit-rate estimates; subtract off the true values; divide by the MC estimates of the standard errors of these residuals; and perform KS tests on these standardized residuals. (For each county, the samples with a poverty estimate of 0 are removed before running the KS tests, since the resulting log-counts and logit-rates would be undefined.)

For the Beta comparisons, we take each county's samples' rate estimates; find the MC estimate of their variance; transform this and the true rate into parameters of the target Beta distribution; and perform KS tests on the rates using these Beta parameters.

This leads to one hundred KS tests (one per county) each on the log-counts, logit-rates, and rates. Using a critical level of $\alpha = 0.05$, 31 of those hundred counties' tests reject the null of $N(0,1)$ for the standardized log-count residuals; 23 reject the null of $N(0,1)$ for the standarized logit-rate residuals; and 38 reject the null of an appropriate Beta distribution for the rates. In other words, most of the counties fail to reject the proposed sampling distribution under each approach.

Since many of the p-values are near the critical level, Figure 2 illustrates how many counties' p-values fall into ranges of clearly rejecting the null ($p < .01$), clearly failing to reject ($p > .1$), or in between. The differences in numbers failing to reject are not dramatic, so there is no conclusive evidence whether the ZOIB or FH sampling models are more appropriate.

**Figure 2**: KS test p-values




The FH and ZOIB linking model checks are simpler. We take the true log counts and the logits of the true rates for each county, regress each variable against the three main $X$ predictors, and look for heteroscedasticity in the residuals. None of the true counts are 0 in this pseudo population, so we do not have to omit any data in these linking-model checks.

See Figure 3 A and C for the plots of truth-data residuals vs fitted values, and Figure 3 B and D for QQ plots of these residuals.

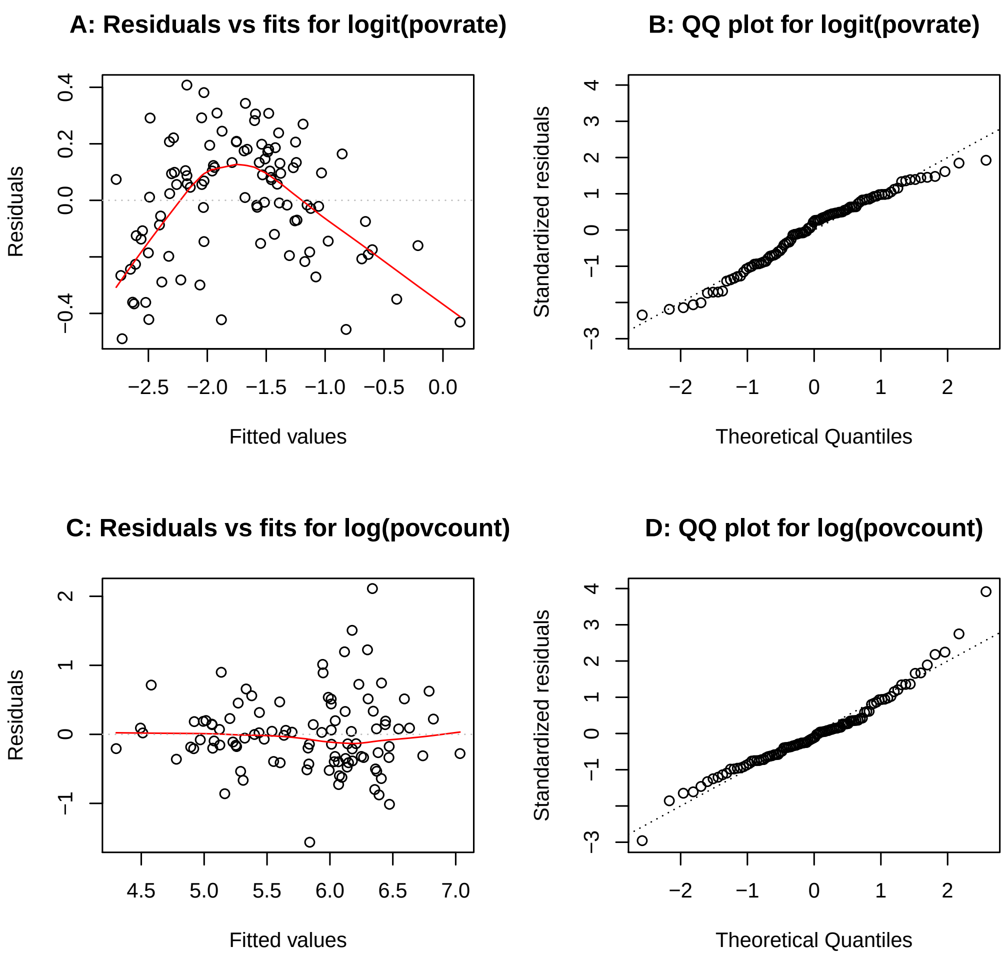
The ZOIB model residuals show a nonlinear relationship with the fitted values, so perhaps adding quadratic or interaction terms could improve the predictions. The FH model residuals show no association between their mean and the fitted values; but their variance does appear to increase with the fitted values, so perhaps the assumption of constant model error is not satisfied. The QQ plots for both show some major deviations from normality in the largest residuals.

## 5.2 External checks

We summarize (as a distribution across the 100 counties) the estimates of $\mathrm{Bias}_{\mathrm{MC}}(\hat{Z}_i)$ and $\mathrm{Bias}_{\mathrm{MC}}(\mathrm{mse}(\hat{Z}_i))$ under each approach: Direct, FH, and ZOIB estimates. We also consider the Variance, MSE, and CI coverage of the point estimates.

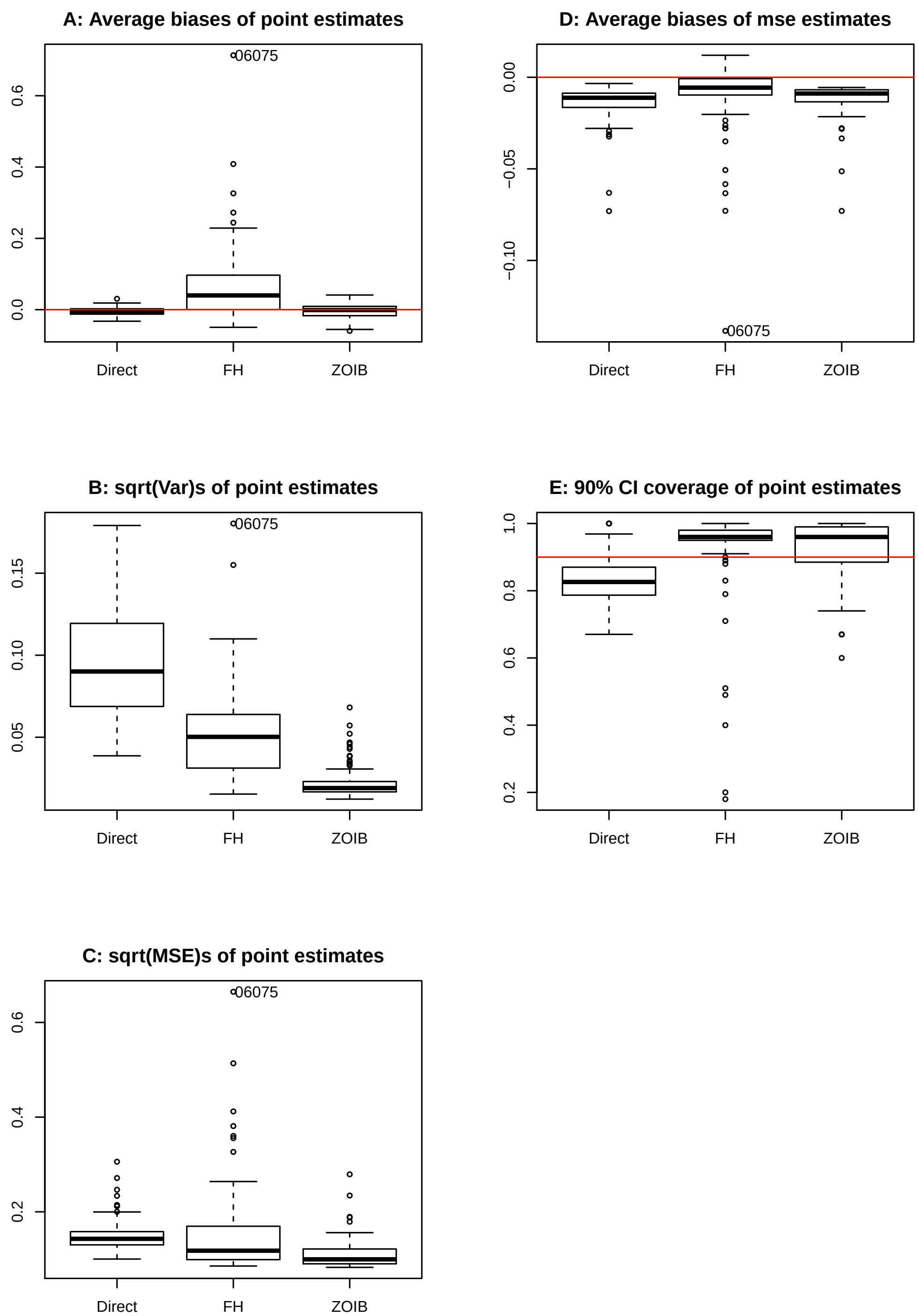
First, Figure 4A presents boxplots of $\mathrm{Bias}_{\mathrm{MC}}(\hat{Z}_i)$ for each county, grouped by estimator. The average bias of the Direct estimates is negligible for most counties, as expected. (This is also a good sign that the survey weighting process is being carried out correctly for our sampling scheme.) The ZOIB estimates also tend to have a small average bias, although some counties have a slightly larger bias than seen in any Direct estimate. The FH estimates, however, tend to be biased upwards, overestimating the true poverty rates. Their

**Figure 3**: Residuals vs fitted values, with LOESS curves (A,C)
and QQ plots of residuals (B,D)
for ZOIB (A,B) and FH (C,D) models




Source: Simulated from U.S. Census Bureau, American Community Survey data, 2010

**Figure 4**: County-to-county variability in each of several metrics,
on each of three estimators




Source: Simulated from U.S. Census Bureau, American Community Survey data, 2010 bias is not very large for most counties, but there are a few extreme outliers, particularly county 06075. (This county had the most negative residual in the linking model, leading to overestimates in the regression; and it had a high number of replicate samples with no poor children in sample, which occurred 55 times in this county. Thus it is not surprising that the poor regression fit, combined with frequent lack of shrinkage toward the direct estimate, leads to the most extreme Bias, Variance, and MSE, as well as the worst mse underestimates.)

The lack of extreme average biases among ZOIB estimates suggests that the ZOIB estimates tend to account for the information from 0s much better than the FH model (which just drops those counties from the model-fitting process entirely). Nugent and Hawala (2012) are developing a left-censored FH model that may help the FH approach to better account for these 0s in the data.

Besides looking at the average bias of the point estimates, we also consider $\mathrm{Var}_{\mathrm{MC}}(\hat{Z}_i)$ and $\mathrm{MSE}_{\mathrm{MC}}(\hat{Z}_i)$, and how these vary across counties. These are shown in Figure 4B and C (with a square-root transformation on each Variance or MSE so that the high outliers do not compress the rest of the boxplots too much). We see that the Direct point estimates tend to have the highest Variances; the FH point estimates tend to be less variable, except for a few outlying counties; and the ZOIB point estimates tend to be least variable of all. The clear reduction in Variance from Direct to FH in Figure 4B is as expected; however, Figure 4C shows that the high FH average biases (seen in Figure 4A) contribute to many counties with high MSEs of their FH point estimates. This appears to cancel out the FH model's benefit of reduced Variance compared to Direct estimates. The ZOIB model point estimates, however, tend to have lower MSEs than either Direct or FH.

Next, we move on from point estimates and look at $\mathrm{Bias}_{\mathrm{MC}}(\mathrm{mse}(\hat{Z}_i))$ under each model. These are shown in Figure 4D. Both the Direct and ZOIB estimates of mse have a small negative average bias in every county. For the FH estimates of mse, most counties have a small negative average bias, but a few do have a positive average bias, and there is one extreme negative outlier. In other words, all three approaches tend to lead to underestimates of mse in most counties, but only the FH model has a county with a severe underestimate of mse. (However, as noted above, the Direct estimates here are based on dropping the undefined $\mathrm{mse}(\hat{Z}_i)$ estimates for replicates where there were no poor in sample. Future work will address this by using an alternative estimator of $\mathrm{mse}(\hat{Z}_i)$ for such replicates.)

Finally, we also check the coverage rates $\mathrm{Coverage}_{\mathrm{MC}}(\hat{Z}_i)$ for 90% confidence or credible intervals (CIs): how often do each model's CIs cover the true value of the poverty rate? These are displayed in Figure 4E. In most counties, the Direct CIs have lower coverage than nominal: the $3^{rd}$ quartile of coverages is lower than the target of 0.90. On the other hand, the FH and ZOIB CIs for most counties have higher than nominal coverage. However, the FH counties with the lowest CI coverages (around 0.2 to 0.5) are much worse than the lowest Direct or ZOIB coverages (around 0.6). The conservative over-coverage of the ZOIB CIs is probably safer than the under-coverage of the Direct CIs, and the outliers are not as bad under ZOIB as under FH.

The lower-than-nominal Direct coverage makes sense: the point estimates are approximately unbiased but the mse estimates tend to be low, so the CIs tend to be too narrow, leading to lower coverage than expected. The higher-than-nominal coverage for FH is also reasonable: these CIs are constructed using mse estimates, not variances, so the CIs will be wider than necessary. Finally, the ZOIB CIs are equal-tail Bayesian credible intervals constructed directly from the MCMC draws rather than by help of the point and mse estimates, so the biases of the point and mse estimates cannot tell us directly why the ZOIB CI coverage tends to be too high.

## 6. Conclusion

The results of Subsections 5.1 and 5.2 suggest that, for this pseudo population, the ZOIB sampling and linking models appear to be no less appropriate than their FH counterparts. The ZOIB model also outperforms the FH model with regards to bias, MSE, and CI coverage.

While the Direct point and mse estimates show the least bias, the ZOIB results are barely worse; and ZOIB point estimates have the lowest MSEs and best CI coverage of the three estimators. Thus, ZOIB appears to outperform the FH model, at least on these samples and this pseudo population. (Future work will continue these tests on a more sophisticated pseudo population and with more replicates.) ZOIB's ability to handle 0 estimates is also a clear benefit over Direct estimation for such small areas.

However, it is important to note that the FH model performs far better on the real SAIPE data than in this paper's illustrations. Here, 0s are more common and there are no large reliable counties, thus preventing the FH model from borrowing strength as much as it does on the full ACS dataset. The results here should not be taken to show that the FH model is unreliable, only that the ZOIB model may be an improvement in certain ways.

## 7. Future Research

Subsection 5.1 suggests some directions for improvement to each of the linking models, and perhaps the sampling models as well. Fixing the issues highlighted by those diagnostics may affect the results of subsection 5.2 as well. There is also a need to improve the diagnostics themselves, particularly for checking whether the ZOIB $p_i^{(0)}$ and $p_i^{(1)}$ are modeled well.

The present work ignores raking or benchmarking, which is part of current SAIPE practice. Further work should assess the effect of benchmarking to state estimates as well as the value of using alternative regressors.

Also, further work could incorporate spatial information (neighboring counties, etc.). Spatial models are able to predict random effects even for counties with no sample, which is not possible under the current non-spatial models.

## REFERENCES

Bauder, M., Luery, D., and Szelepka, S. (2012), "Small Area Estimation of Health Insurance Coverage in 2008 and 2009," U.S. Census Bureau technical report, http://www.census.gov/did/www/sahie/methods/20082009/files/sahie_20082009_technical_methodology.pdf

Espinheira, P.L., Ferrari, S.L.P., and Cribari-Neto, F. (2008), "On beta regression residuals," *Journal of Applied Statistics*, vol.35, no.4, pp.407-419.

Fay, R.E. and Herriot, R.A. (1979), "Estimates of income for small places: an application of James-Stein procedures to census data," *Journal of the American Statistical Association*, vol.74, pp.269-277.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004), *Bayesian Data Analysis*, Boca Raton: Chapman & Hall.

Hawala, S., and Lahiri, P. (2010), "Variance Modeling in the U.S. Small Area Income and Poverty Estimates Program for the American Community Survey," *Proceedings of the American Statistical Association, Section on Bayesian Statistical Science, Section on Survey Research Methods*, Alexandria, VA: American Statistical Association.

Hawala, S., and Lahiri, P. (2012), "A Hierarchical Bayes Estimation of Poverty Rates," *Proceedings of the American Statistical Association, Section on Survey Research Methods*, Alexandria, VA: American Statistical Association. Forthcoming.

Huang, E.T. and Bell, W.R. (2009), "A Simulation Study of the Distribution of Fays Successive Difference Replication Variance Estimator," *Proceedings of the American Statistical Association, Survey Research Methods Section*, Alexandria, VA: American Statistical Association.

Little, R.J. (2006), "Calibrated Bayes: A Bayes/Frequentist Roadmap," *The American Statistician*, vol.60, nr.3, pp.213-223.

Liu, B., Lahiri, P., and Kalton, G. (2007), "Hierarchical Bayes Modeling of Survey-Weighted Small Area Proportions," *Proceedings of the American Statistical Association, Section on Survey Research Methods*, Alexandria, VA: American Statistical Association.

Nugent, C., and Hawala, S. (2012), "Research and Development for Methods of Estimating Poverty for School-Age Children," *Proceedings of the American Statistical Association, Section on Survey Research Methods*, Alexandria, VA: American Statistical Association. Forthcoming.

Ospina, R., and Ferrari, S.L.P. (2010), "Inflated beta distributions," *Statistical Papers*, 51, 1, pp.111-126.

Pfeffermann, D., Terryn, B., and Moura, F.A.S. (2008), "Small area estimation under a two-part random effects model with application to estimation of literacy in developing countries," *Survey Methodology*, 34, (2), pp.235-249.

Plummer, Martyn (2003), "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling," Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 2022, Vienna, Austria.

R Development Core Team (2012), "R: A Language and Environment for Statistical Computing," http://www.R-project.org, R Foundation for Statistical Computing, Vienna, Austria.

SAS Institute (2008), Version 9.2 of the SAS System for Unix. Cary, NC, USA: SAS Institute Inc.

Slud, E., and Maiti, T. (2006), "Mean-squared error estimation in transformed Fay-Herriot models," *Journal of the Royal Statistical Society B*, 68, Part 2, pp.239257.

U.S. Census Bureau (2009), "Design and Methodology: American Community Survey," Washington, DC: U.S. Government Printing Office.

U.S. Census Bureau (2010), "2006 - 2009 County-Level Estimation Details," last modified December 8, 2010, http://www.census.gov/did/www/saipe/methods/statecounty/20062009county.html

U.S. Census Bureau (2011), "Information about Data Inputs," last modified November 29, 2011, http://www.census.gov/did/www/saipe/data/model/info/index.html

U.S. Census Bureau (2012), "How the Census Bureau Measures Poverty," last modified September 12, 2012, http://www.census.gov/hhes/www/poverty/about/overview/measure.html

Wieczorek, J. (2012), "ACS-like subsampling from a pseudo population." Forthcoming technical report, http://www.census.gov/did/www/saipe/publications/techpubs.html

Wieczorek, J. and Hawala, S. (2011), "A Bayesian Zero-One Inflated Beta Model for Estimating Poverty in U.S. Counties," *Proceedings of the American Statistical Association, Section on Survey Research Methods*, Alexandria, VA: American Statistical Association.