Hierarchical Bayes Estimation of Poverty Rates *

Sam Hawala† Partha Lahiri‡

Abstract
In practice many applications of small area models use a 'Normal-Normal-Linear' assumption, i.e., a normality assumption for the design-based survey estimates and for the area-level random effects and a linear regression function relating the true parameters to available covariates. We compare the performance of rate models by slightly changing the assumptions and using internal and external checks. when area sample sizes are in the hundreds, empirical analyses using a 'Normal-t-Linear' to protect against outliers, or a seemingly reasonable 'Beta-logistic' assumption for rates, show no gain over the 'Normal-Normal-Linear' type model. However, the same type of analyses show additional benefit from including historical data through a cross-sectional and time series model. We use Monte Carlo Markov Chain (MCMC) to implement the proposed models, posterior predictive checks, as well as external checks for model comparisons.

Key Words: Small Area Estimation, SAIPE, Time-Series and Cross-Sectional model, Posterior Predictive Checks.

1. Introduction

In this paper we compare the performance of a few small area models and assess the benefit from incorporating additional information into model features. The specific focus of our data analysis is on the U.S. Census Bureau's Small Area Income and Poverty Estimates program (SAIPE) models. SAIPE uses area level models to develop state and county estimates of poor school-age children. These estimates are an important component of the administration of federal funds each year under Title I of the Elementary and Secondary Education Act.

The SAIPE production models are of the Fay Herriot type (see Fay & Herriot (1979)), which we will denote throughout the paper by *FH*. These models use a 'Normal-Normal-Linear' assumption, i.e., normal distributions for the 'direct' design-based estimates and for the area-level random effects, and a linear regression function relating the true poverty rates to covariates from administrative and other data sources. The SAIPE models have been widely reviewed and evaluated (see for example the NRC Report, Citro and Kalton eds., 2000). The NAS panel also recommended continued research to determine if we can improve upon the current production models. Wiezcorek, Nugent & Hawala (2012), use a sampling design, roughly consistent with that of the American Community Survey (ACS), to show that for more counties the direct survey estimated poverty rate follows a beta rather than a normal distribution and that a zero-one-inflated beta regression model outperforms the SAIPE *FH* model, estimated on the non-zero observations, with regards to bias, mean square error (MSE), and confidence interval coverage. Nugent & Hawala (2012) use a censored *FH* model, analogous to that in Slud et al.

*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.

†U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

‡JPSM, University of Maryland, and U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

(2011), to reveal the bias incurred when we drop the poverty counts that are zero from the estimation procedure.

In general, small area models can improve upon 'direct' i.e., survey-only estimates in several ways. Models use auxiliary information, available for all areas, as covariates to predict area characteristics such as poverty rates. Estimation of model parameters use the auxiliary data from all areas, consequently allowing areas to borrow information (strength) from each other. This is especially important for areas where survey data on the area characteristic of interest are not available. The models that allow for borrowing strength across areas are labeled cross-sectional. Models can also take into account repeated observations, such as annual surveys. For example, last year's poverty rate provides some information on the current year's poverty rate. Models that use data collected at previous time points, labeled time-series, borrow strength across time. Models that use information both from other areas and other time points are labeled time-series and cross-sectional. Work in this area include those of Pfeffermann and Burck (1990), Rao and Yu (1994), Ghosh et al. (1996), and Datta et al. (1999).

Our main application focuses on the SAIPE state model. Our data analysis uses the Current Population Survey (CPS) state data for model comparisons. The SAIPE program produces state and county level annual estimates of poverty rates and counts of poor school age children using a Bayesian and empirical best (EBLUP) estimation of a *FH* type model. For more details on the SAIPE methodology the reader should consult Bell et all (2007). From 1993 to 2005, SAIPE fitted the models to direct income and poverty estimates from the Annual Social and Economic Supplement (ASEC) of CPS, which is administered in March of each year. Starting in 2006, the response variable in the state model became the direct estimate of the poverty percentiles from the ACS. In this paper we use CPS-ASEC data to explore alternatives to the *FH* model and we compare the results. These data contain observations made at 11 time points (years), which is a longer time series, than what is currently available from the ACS.

There is a huge literature on both theory and applications of model-based small area estimation. For a comprehensive review on the subject, see Rao (2003), or Jiang and Lahiri (2006). Malec (2005) considers a multivariate multinomial/binomial model for estimating housing unit characteristics and poverty in census tracts. This housing unit level model drops the assumptions of normality and known sampling variances. However, working with unit-level data presents many challenges, not the least of which is data confidentiality. We focus on area-level models and we do not pursue unit level models at this time.

The World Bank method for poverty mapping — the so called *ELL* method —- developed by Elbers, Lanjouw, and Lanjouw (2003), uses a unit level (e.g., household level) mixed model that establishes a relationship between a welfare variable and a variety of explanatory variables that are common between the household survey and the previous census. The model is fitted using the survey data and the fitted model is used to multiply impute the welfare variable for all households in the census file. These "census like files" are then used for producing poverty maps and the associated measures of uncertainty. For a comprehensive description of the World Bank methodology and applications of the method in different countries, see Elbers et al. (2003, 2008), Neri et al. (2005), and numerous other applications in different countries, see http://go.worldbank.org/9CYUFEUQ30.

Recently, Molina and Rao (2010) put forward an empirical Bayes method for poverty mapping using a unit level model like the one proposed by *ELL* except that it uses explicitly stated small area specific random effects generally used in small area estimation. In their paper, they claimed their method to be superior to the *ELL* method. However, a common thread in the *ELL* and Molina-Rao approaches is the use of census like files with an imputed welfare variable.

The central issue in poverty mapping, just like any other small area estimation problem, has always been the choice of a specific model-based poverty mapping technique and not whether one should or should not use models. It is possible that different poverty mapping methods may be applicable in different situations depending on the specific needs. For example, if the sample sizes in many small areas are very small or zero, and various non-sampling errors arising out of the old census frame and dissimilarity of explanatory variables across survey and the census (e.g., Tarozzi and and Deaton , 2009) are negligible, the *ELL* method could be a sensible method to apply. In the presence of such negligible non-sampling errors, one can also consider the Molina-Rao method if most of the small areas have some sample in them.

The *ELL* method is based entirely on imputed data. The Molina-Rao method, on the other hand, uses actual survey data. However, unlike the *ELL* method, the Molina-Rao method requires linking of the survey data to the previous census data at the household level, which can be a formidable task. In addition, the non-sampling errors issues raised earlier may be problematic.

We develop our paper as follows. In section 2.1 we use the year 1989 CPS-ASEC data, and a common set of regressors to compare two different cross-sectional small area models to the *FH* model. In section 2.2 we use eleven years of CPS-ASEC data to estimate a time-series and cross-sectional model. In section 3 we discuss model evaluations and comparisons. Section 4 concludes our paper by providing a summary of our findings and directions for further research.

2. Data Analysis

2.1 Cross-Sectional Models

We used Markov Chain Monte Carlo (MCMC) estimation available in R (R Development Core Team, 2012) and JAGS (Plummer, 2003) to test alternatives to modeling the CPS-ASEC state data. In what follows we present the formulation of two basic cross-sectional models, a *FH* type model with normally distributed random effects, a *FH* model but with a t_distribution for the random effects (denoted as *FH_t*) and a Beta-Logistic model (denoted as *BL*).

2.1.1 The Fay Herriot model for poverty rates

In the state level CPS-ASEC data, we have $m = 51$ areas (50 states and the District of Columbia). For $i = 1, \dots, m$, let p_i and P_i be the CPS-ASEC weighted estimate and the true rates of children in poverty in state i, respectively. Just as in the SAIPE models we assume that the sampling variances D_i are known without error. In reality the D_i's are estimated via a GVF function, see Bell et al. (2007).

Let $\boldsymbol{\beta}$ be a vector of k regression parameters including the intercept.

$$\begin{aligned} p_i|P_i &\overset{\text{ind}}{\sim} \mathcal{N}\left(P_i, D_i\right) \\ P_i &= \mathbf{x_i}'\boldsymbol{\beta} + u_i \\ u_i|\sigma_u^2 &\overset{\text{iid}}{\sim} \mathcal{N}\left(0, \sigma_u^2\right) \\ f(\boldsymbol{\beta}, \sigma_u^2) &\propto \mathcal{N}_k(0, (1/\epsilon)I_k) \end{aligned}$$

where I_k is the $k \times k$ identity matrix. In this model, we set a 'noninformative' (or weakly informative) prior $Uniform(0, \frac{1}{\epsilon})$, with $\epsilon = 10^{-6}$, on the random-effects variance parameter σ_u^2. Also for each of the $k-1$ regression parameters $\beta_j, j = 1, \ldots, k-1$, independently, we use a vague normal prior with $\epsilon = 10^{-6}$.

Through simulations, based on the original structure of the CPS-ASEC data but with known true values for the fixed effect, variance parameters, and estimation of the variance component model:

$$p_i = \beta_0 + u_i + e_i, \quad i = 1, \ldots, m$$
$$u_i \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \sigma_u^2\right), \quad e_i \overset{\text{ind}}{\sim} \mathcal{N}\left(0, D_i\right)$$

we found that other non-informative prior distributions, such as the uniform $U(0, \frac{1}{\epsilon})$, on σ_u, suggested by Gelman (2006), and the gamma $\Gamma(\epsilon, \epsilon)$ on $1/\sigma_u^{-2}$ produce incorrect inferences for the *FH* model.

In the full model the matrix $\mathbf{X}$ consists of a column of $1's$ and 4 covariates making up the other 4 columns (so here $k = 5$):

- x_1: state rate of IRS Child Tax-Poor Exemptions
- x_2: state rate of IRS non-filers
- x_3: state rate of Food Stamp participation
- x_4: "census residuals" obtained by regressing previous census poverty percentiles on the three preceding variables concurrently measured with the census.

For each area i the IRS poverty rate x_{i1} is the number of child exemptions for the households in the area, whose reported adjusted gross income is under the poverty level, divided by the total number of child exemptions for all households in the area.

We carry out 3 runs of the MCMC sampler for 20,000 iterations each, taking every 4th, and following a 5,000-iteration burn-in period. We use the popular Gelman and Rubin (1992) diagnostic measure to monitor convergence for the model parameters and the small area means. In our analysis, we check and see if Gelman's "potential scale reduction" factors are all 1.

Table 1 shows posterior means and standard deviations of the *FH* model parameters.

Table 1: *FH* model parameter posterior means and standard deviations

Parameter	Mean	SD
Intercept	0.006	0.02
β_1	0.53	0.10
β_2	0.23	0.15
β_3	0.70	0.22
β_4	0.72	0.37
σ_u	0.012	0.006

2.1.2 The Fay Herriot model with t_distribution

The *FH* type model has been thoroughly tested and reviewed. However, the normality assumption for the random effects is not easy to check and does not protect against possible outliers in the data. For this reason we compare the *FH* model to the 'Normal-t-Linear' (*FH_t*) model.

The model formulation when we replace the normal distribution by a t_distribution becomes:

$$\begin{aligned} p_i|P_i &\overset{\text{ind}}{\sim} \mathcal{N}\left(P_i, D_i\right) \\ P_i &= \mathbf{x_i}'\boldsymbol{\beta} + u_i \\ u_i|\sigma_u^2 &\overset{\text{iid}}{\sim} t\left(0, \sigma_u^2, \nu\right) \\ f(\boldsymbol{\beta}, \sigma_u^2) &\propto \mathcal{N}_k(0, (1/\epsilon)I_k) \end{aligned}$$

The t_distributions have heavier tails than normal distributions, researchers use them as a robust approach to handle influential outliers or, in our case, excessively small or large direct survey weighted observations. We initially let the degrees of freedom ν be a random parameter and set a categorical distribution as a prior on ν, in the range $\{2, \ldots, 50\}$. We obtained a credible interval covering the entirety of this range. Moreover, in view of the results obtained in Bell and Huang (2006), we estimated the remaining parameters by setting $\nu = 8$.

Table 2 shows posterior means and standard deviations of the *FH_t* model parameters.

Table 2: *FH_t* model parameter posterior means and standard deviations

Parameter	Mean	SD
Intercept	0.007	0.02
β_1	0.54	0.11
β_2	0.22	0.16
β_3	0.68	0.23
β_4	0.72	0.38
σ_u	0.011	0.005

2.1.3 A Beta-Logistic Regression model

Since we are modeling rates in $(0, 1)$ we explored a 'Beta-logistic' assumption, i.e., a beta distribution for the survey-weighted rates, and a logit regression relating covariates to the true poverty rate. The modeling approach using the beta distribution simultaneously accounts for the asymmetry of the distribution of poverty rates and the non-constant nature of the variance. Again let P_i denote the true poverty rate in state i and let p_i denote the estimated (through CPS-ASEC) poverty rate in state i. The density function of the beta distribution (with parameters a and b) is

$$f(p|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}$$

We use the same first three covariates as in section 2.1.1 (tax poverty rate, tax non-filing rate, food stamp participation rate). We did not use the 'Census Residuals' because of lack of convergence of the MCMC when we include this last covariate. With the logit link function, transforming the true poverty rate P, the regression parameters can be interpreted in terms of odds ratio. The inverse-logit back-transformation guarantees P to be in $(0,1)$.

In order to incorporate the information from the covariates into a beta regression model, we parameterize the beta family in terms of its mean, $P = E(p|P) = \frac{a}{a+b}$, and a parameter related to its variance, $\gamma = a + b$. Inversely, the parameters a, b can be expressed as $a = \gamma P$ and $b = \gamma(1 - P)$. Note that the variance of a beta distribution is

$$Var(p|P) = \frac{P(1-P)}{\gamma + 1} = \frac{ab}{(a+b)^2(a+b+1)}$$

The variance does depend on the mean P, and larger values of γ correspond to less heterogeneity in the data.

Assuming the sampling variances D_i are known, we can rewrite $Var(p_i|P_i)$ as a variance of rate estimates rescaled by $\gamma_i + 1$:

$$D_i = Var(p_i|P_i) = \frac{P_i(1-P_i)}{\gamma_i + 1}$$

This leads to

$$\gamma_i = \frac{P_i(1-P_i)}{D_i} - 1$$

The Bayesian beta-logistic model we consider along with the prior distributions is as follows:

$$\begin{aligned}
p_i|P_i &\overset{\text{iid}}{\sim} \text{Beta}\left(\gamma_i P_i, \gamma_i(1-P_i)\right) \\
\text{logit}\,(P_i) &= \mathbf{x}_i'\boldsymbol{\beta} + u_i \\
u_i|\sigma_u^2 &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_u^2) \\
f(\boldsymbol{\beta}, \sigma_u) &\propto \mathcal{N}_k(0, (1/\epsilon)I_k)
\end{aligned}$$

We use the same hyper-parameter values as in the previous section; $\epsilon = 10^{-6}$. However, the prior on random effect variance parameter is now $Uniform(0, \frac{1}{\epsilon})$. The posterior estimates for the parameters in this *BL* model are given in Table 3.

Table 3: *BL* model parameter posterior means and standard deviations, parameters here are on logit scale

Parameter	Mean	SD
Intercept	-2.82	0.18
β_1	4.36	0.94
β_2	1.87	1.46
β_3	3.44	1.90
σ_u	0.218	0.129

2.2 Time-Series and Cross-Sectional Model

The analysis of data from either the CPS or the ACS annual surveys cannot ignore the autocorrelation over time. It is especially relevant when modeling ACS county data, where an area may have no poor people in sample one year, leading, unrealistically, to an ACS direct survey estimate of zero for that year. It is reasonable to borrow information from data collected in previous years for the same area to estimate the true rate for the year when data are unavailable. To get some idea of the poverty rate over time, we plotted the state level IRS poverty rate for DC, California (CA), NH, and CT, see Figure 1. We also we plotted the state level CPS poverty rate for 11 years for four areas - the National rate, for New Hampshire (NH), the District of Columbia (DC), and Connecticut (CT), see Figure 2. These figures show that for large areas CA and US the past is a strong predictor for CPS and IRS poverty rates. For smaller areas the CPS rates can vary considerably in time.

Figure 1: IRS poverty rates over time

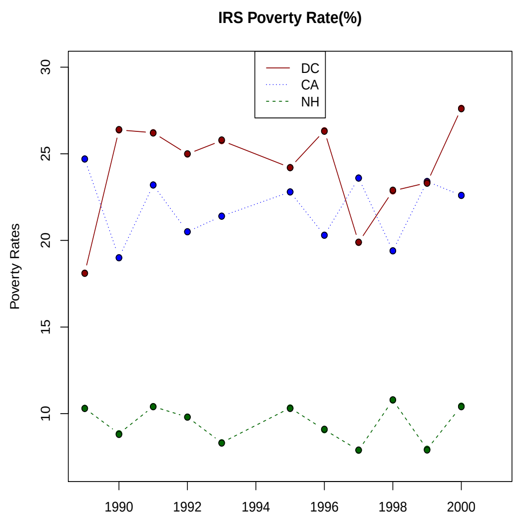
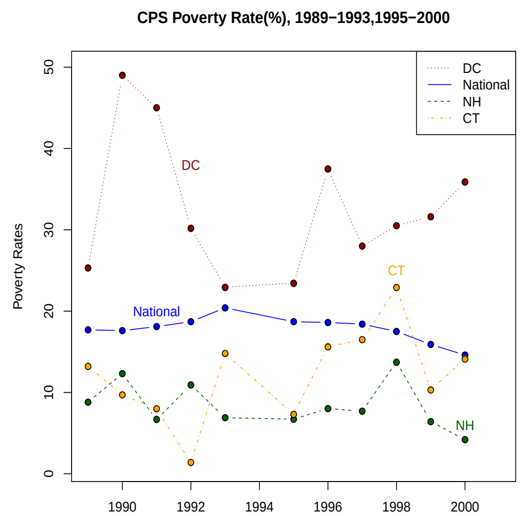


Figure 2: CPS poverty rates over time



To borrow information across time we incorporate a time series component by considering an autoregressive model of order 1 (AR(1)) on the true rate P, similar to Rao and Yu (1994) but with regression parameters $\boldsymbol{\beta_t}$ varying with time t. In this paper our time-series and cross-sectional (TSCS) model uses CPS-ASEC data on income from $T = 11$ years (1989-1993, 1995-2000) on each of the 50 states and the District of Columbia, $m = 51$ areas. 1994 data were not available because SAIPE production, of model based estimates, started in 1993 but skipped 1994. We will assume that skipping 1994 has no effect on our model estimation.

Let $p_{i,t}$ be the direct survey estimator, of the true percent in poverty $P_{i,t}$, for the i-th small area at time point $t, (i = 1, \ldots, m; \quad t = 1, \ldots, T)$. $p_{i,t}$ is assumed to be unbiased for $P_{i,t}$. As in the *FH* model, of section 2.1.1, we assume that the sampling variances D_{it} are known.

2.2.1 AR(1) Time-series and cross-sectional, Fay Herriot structure

The next set of statements and equations describe the formulation of an AR(1) time-series and cross-sectional model in the Fay Herriot structure. Let ρ be the first order autoregressive parameter.

$$
\begin{array}{rcl}
p_{i,t}|P_{i,t} & \overset{\text{ind}}{\sim} & \mathcal{N}\left(P_{i,t}, \quad D_{i,t}\right) \\
P_{i,t} & = & \mathbf{x}'_{i,t}\boldsymbol{\beta_t} + u_i + \eta_{i,t} \\
\eta_{i,t} & = & \rho\eta_{i,t-1} + \epsilon_{i,t}, \qquad |\rho| < 1 \\
u_i|\sigma_u^2 & \overset{\text{iid}}{\sim} & \mathcal{N}(0, \sigma_u^2) \\
\epsilon_{i,t} & \overset{\text{iid}}{\sim} & \mathcal{N}(0, \sigma_\eta^2) \\
f(\boldsymbol{\beta}, \sigma_u^2, \rho) & \propto & \mathcal{N}_k(0, (1/\epsilon)I_k)
\end{array}
$$

We assume that the component $\eta_{i,t}$ of the model errors follow an autoregressive process of order 1, so that, for any $t = 1, \ldots, T$, and any $i = 1, \ldots, m$, the correlations of these errors at lag s is

$$Corr(\eta_{i,t}, \eta_{i,t-s}) = \rho^s, \quad s = 0, \ldots, T-1$$

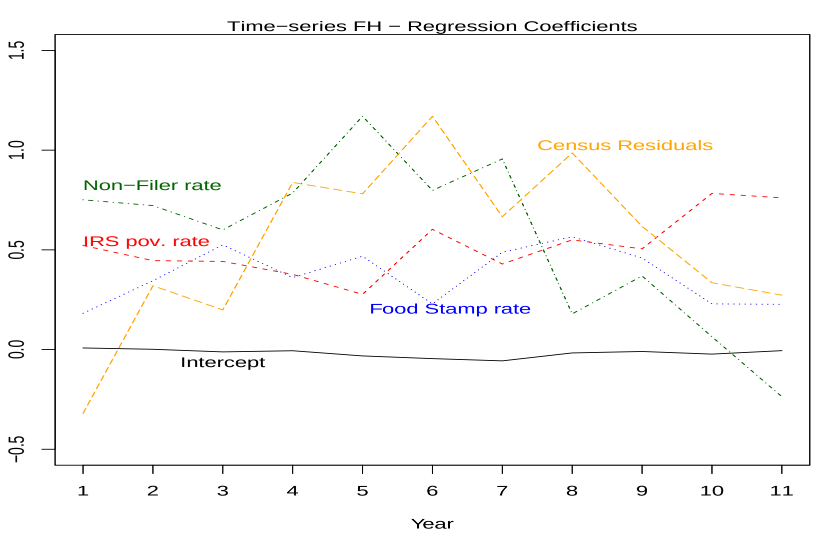
For the prior distributions we again use $\epsilon = 10^{-6}$. In Table 4, we provide the posterior estimates of the autocorrelation coefficient ρ and the (square roots) of the variance components: σ_u and σ_η.

Table 4: *TSCS* autocorrelation and variance components posterior means and standard deviations

Parameter	Mean	SD
ρ	0.60	0.185
σ_u	0.017	0.009
σ_η	0.012	0.006

Figure-**3** shows the estimates of the *TSCS* models' regression parameters $\boldsymbol{\beta_t}$ and their values for the 11 years.

Figure 3: TSCS model regression parameters for 1989-1993; 1995-2000



3. Model Checks and Comparisons

3.1 Internal Checks

We explored several diagnostic tools to assess the adequacy of the models we considered. We used the same MCMC method used for estimation, to calculate model checks, each of which looks for a particular flaw in the model but cannot confirm the correctness of the model as a whole. We can check the adequacy of the model fit, and appropriateness of model assumptions using the predictive distribution. This distribution enables *"criticism"* of the model in light of the current data. Our main interest is in comparing models. For this purpose, predictive distributions are directly comparable while posterior distributions are not. So we don't focus only on the posterior means and variances, which are parametrized by the model.

Hierarchical models in general require several assumptions at each level of the hierarchy. It is more likely to obtain misleading inferences if any of the multiple parts of the model is poorly specified. The presence of multiple sources of variation, coming from each level of the model, make residual analyses, for example, more complicated because of possible confounding in the sense that the residuals are errors emanating from multiple sources.

Researchers usually take two broad approaches: model checking and sensitivity analysis. In the first they try to inspect the results to determine if any of the assumptions are plainly violated and if the model cannot reproduce some features of the data. In the second, modifications are made to the assumptions in question and changes in posterior quantities are evaluated for practical significance.

We first try to determine if the models provide adequate fit to the data. We use posterior predictive checks as in Gelfand, Dey and Chang (1992), Rubin (1984), Meng (1994) and Gelman, Meng and Stern (1996). Again, these checks may be necessary, but they are not sufficient. Inconsistencies between the proposed model and *'the true model'* may not even be detectable.

If the model fits to $\boldsymbol{p}_{obs}$ are adequate, replicated values $\boldsymbol{p}_{new}$ generated from the model would be similar to $\boldsymbol{p}_{obs}$. We generate replicates from the posterior predictive distribution:

$$f(\boldsymbol{p}|\boldsymbol{p}_{obs}) \quad = \quad \int f(\boldsymbol{p}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{p}_{obs})d\boldsymbol{\theta}$$

where $f(\boldsymbol{p}|\boldsymbol{\theta})$ is the likelihood function and $\pi(\boldsymbol{\theta}|\boldsymbol{p}_{obs})$ is the joint posterior distribution of all the model parameters $\boldsymbol{\theta}$ given the data $\boldsymbol{p}_{obs}$. We use the posterior predictive distribution to calculate the following divergence measure proposed in Laud and Ibrahim (1999) :

$$D_{LI} = d(\boldsymbol{p}_{new}, \boldsymbol{p}_{obs}) = E\left(\parallel \boldsymbol{p}_{new} - \boldsymbol{p}_{\boldsymbol{obs}} \parallel^2 |\boldsymbol{p}_{\boldsymbol{obs}}\right)$$

approximated by

$$\hat{D}_{LI} = \frac{1}{mK}\sum_{k=1}^{K} \parallel \hat{\boldsymbol{p}}_{new} - \boldsymbol{p}_{obs} \parallel^2$$

In comparing models, the one resulting in the smallest D_{LI} outperforms the others. We give values of $\hat{D}_{LI}$ for each model in Table 5.

Table 5: Laud & Ibrahim Measure

Model	D_{LI}	$\sqrt{D_{LI}}$
FH	0.00166	.0407
FH_t	0.00200	.0447
BL	0.00280	.0529
TSCS	0.00165	.0406

We calculated the following Bayesian posterior predictive quantities:

- $p_D = Pr\left(D(\boldsymbol{p_{new}}, \boldsymbol{\theta}) > D(\boldsymbol{p}, \boldsymbol{\theta})\right)$, where

$$D(\boldsymbol{p}, \boldsymbol{\theta}) = \sum_{i=1}^{m} \frac{(p_i - E(p_i|\boldsymbol{\theta}))^2}{Var(p_i|\boldsymbol{\theta})}$$

  This is a summary measure from Gelman et al. (2004) - an omnibus measure of goodness-of-fit. p_D is the probability of observing a sum of squared "residuals" much different than the one obtained from the data. Unlike *p_values* in the classical statistics framework which are sought to either reject or not reject a hypothesis, small or large p_D's will cast doubt on a model but not outright reject it. A small or large p_D (less than .01 or bigger than .99) suggests a discrepancy between the data and the model.

  We also calculated *p_values* for each of the m order statistics. For example,

- $Q_{(1)p}$ = p-value for agreement of the smallest data values in the observed and artificial (replicated) datasets. $Q_{(1)p} = Pr\left(p_{(1),obs} \geq p_{(1),new} | \boldsymbol{p_{obs}}\right)$

- $Q_{(m)p}$ = p-value for agreement of the largest data values in the observed and artificial (replicated) datasets. $Q_{(m)p} = Pr\left(p_{(m),new} \geq p_{(m),obs} | \boldsymbol{p_{obs}}\right)$

- In general $Q_{(i)p}$ = p-value for agreement of the i^{th} order statistic $p_{(i)}$ in the observed and artificial (replicated) datasets.

$$Q_{(i)p} = Pr\left(p_{(i),obs} \geq p_{(i),new} | \boldsymbol{p_{obs}}\right)$$

  For a model comparison we summarize the *p_values* for the order statistics for the CPS-ASEC 2000 data in Figures 4 and 5.

Figure 4: FH and BL models - CPS-ASEC 2000 data

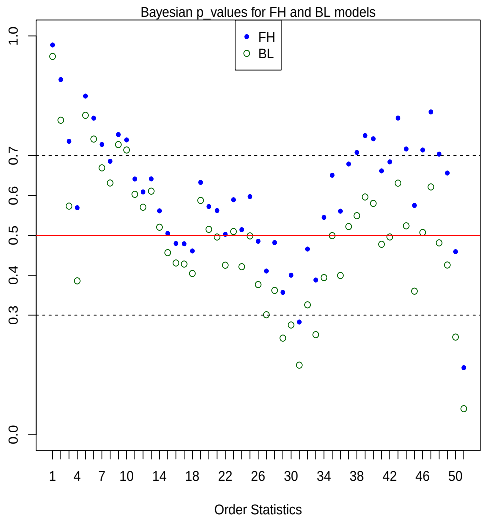
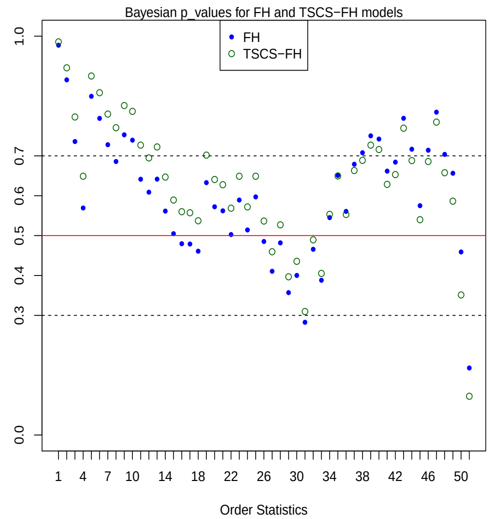


Figure 5: FH and TSCS models - CPS-ASEC 2000 data



Finally we calculated for each i the - individual - posterior predictive p-values (*ppp*-values) defined as:

$$\hat{p}\_value_i = Pr\left(p_{i,new} \leq p_{i,obs} | \boldsymbol{p}_{obs}\right); \quad i = 1, \ldots, m.$$

We made Q-Q plots of the m values of $\hat{p}\_value_i$ comparing their distribution to the uniform distribution as a reference. These plots did not reveal anything for model comparisons. They are not included in this paper but available from the authors.

3.2 External Checks

All of our internal checks did not reveal much in terms of relative model performance. We turned to comparisons between model estimates of poverty rates based on 1989 CPS and census 1990 poverty rates. The national poverty rates from 1989 CPS and from census 1990 were $y_{nat}^{cps89} = 18\%$ and $y_{nat}^{cen90} = 17.5\%$ respectively. We ratio-adjusted the model estimates of the state rates $\hat{Y}_i$ to obtain

$$\hat{Y}_i^R = \frac{y_{nat}^{cen90}}{y_{nat}^{cps89}} \hat{Y}_i = .9721 \hat{Y}_i$$

then we calculated and plotted the differences

$$diff_i = \hat{Y}_i^R - \hat{Y}_i^{cen}$$

We would like to mention that, upon review of this paper, William Bell brought to our attention a different way of adjusting the rates for comparisons. We also thought of modifications to his approach. We hope to work on these issues and publish the results in a subsequent paper.

We also calculated the 'absolute relative differences'

$$ardiff_i = \left| \frac{\hat{Y}_i^R}{\hat{Y}_i^{cen}} - 1 \right|$$

We calculated the differences $diff_i$ in two instances. One where the estimates $\hat{Y}_i$ are from models without covariates to get $diff_i^{no\_cov}$ and another with all four covariates to get $diff_i^{cov}$. We summarize these straight differences in the next two figures: Figure 6 and 7.

Figure 6

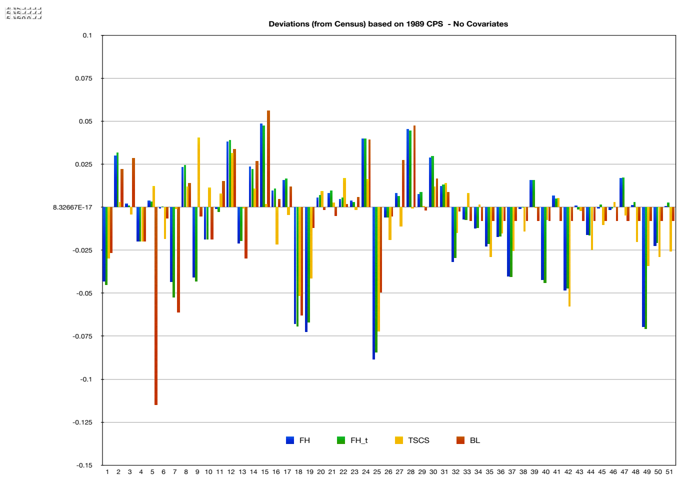
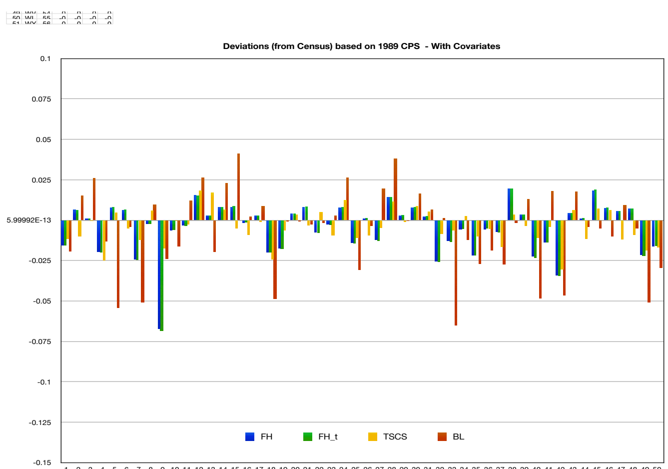


Figure 7



We did the same for the absolute relative differences, i.e. with and without covariates. These are summarized in the next two tables: Figures 8 and 9.

Figure 8

Summary of Absolute Relative Deviations Models without covariates				
	FH	FH_t	TSCS	BL
Min.	0.38%	0.06%	0.27%	0.86%
1st Qu.	3.59%	3.42%	2.64%	5.35%
Median	10.87%	11.01%	8.96%	11.41%
Mean	14.04%	14.34%	9.55%	16.74%
3rd Qu.	21.91%	21.55%	13.83%	22.43%
Max.	42.46%	51.16%	30.16%	65.89%

Figure 9

Summary of Absolute Relative Deviations Models with covariates				
	FH	FH_t	TSCS	BL
Min.	0.36%	0.39%	0.11%	0.18%
1st Qu.	2.72%	2.75%	2.75%	4.02%
Median	5.65%	5.73%	4.94%	8.80%
Mean	6.91%	6.99%	5.75%	12.21%
3rd Qu.	8.55%	8.64%	8.31%	18.60%
Max.	27.08%	27.61%	16.27%	49.48%

4. Conclusions

The Bayesian posterior predictive checks did not clearly reveal differences in performance between the models we considered. Our work followed several of the authors mentioned already. It mainly consists of calculating Bayesian p_values, which Hjort et al. (2006) described as measures of degree of surprise from data, given the prior and the model.

One of the measures for model comparisons that indicated a small difference between the models' performance is (D_{LI}) the one due to Laud and Ibrahim. Using D_{LI}, neither the Fay-Herriot model with random effects distributed as t, nor the Beta logistic model introduce any improvement. In fact the usual Fay-Herriot model does better in both cases.

The Bayesian p_values for the order statistics show that there are problems with all four models in fitting the lower and upper tails of the distribution of $\boldsymbol{p}$.

This is also true for other values as we move away from the median. Overall, the model predictions tend to be smaller than the observed data at the left tail of the distribution and the predictions are larger at the right tail of the distribution. For the *FH* model, about 50% of the p_values are above .6 and only about 12% are below .4. The highest ones are for the first order statistic $Q_{(1)p} \approx .98$. The lowest ones are for the last order statistics $Q_{(m)p} \approx .14$.

As we expected, and as the external checks show, there is a 5% gain in efficiency, on average, through borrowing strength from previous years of CPS-ASEC data. An absolute relative difference is analogous to a coefficient of variation. However, this gain is not as big (1% on average) when we use all the covariates in the model.

REFERENCES

Bell et al. (2007) "Use of ACS Data to Produce SAIPE Model-Based Estimates of Poverty for Counties", see SAIPE website http://www.census.gov/did/www/saipe/publications/files/report.pdf

Bell, W.R. (2008), "Examining Sensitivity of Small Area Inferences to Uncertainty about Sampling Error Variances," *Proceedings of the Section on Survey Research Methods, Alexandria,* VA: American Statistical Association, pp 327-334.

Bell et al. (2006) " Using the t-distribution in Small Area Estimation: An Application to SAIPE State Poverty Models " *Proceedings of the Section on Survey Research Methods, Alexandria,* VA: American Statistical Association, pp 3142-3149.

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

Datta, G.S., Lahiri, P., Maiti, T., and Lu, K.L. (1999), "Hierarchical Bayes Estimation of Unemployment Rates for the States of the U.S.", *Journal of the American Statistical Association,* Vol. 94, No. 448, pp. 1074-1082.

Elbers, C., Lanjouw, J., and Lanjouw, P. (2003), "Micro-Level Estimation of Poverty and Inequality", *Econometrica,* Vol. 71, 355-364.

Elbers, Chris, Peter Lanjouw, and Philippe George Leite, "Brazil within Brazil: Testing the Poverty Map Methodology in Minas Gerais," *World Bank policy research working paper* no. 4513 (2008).

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, 269-277.

Gelfand, A.E., Dey, D.K., and Chang, H. (1992). "Model determination using predictive distributions with implementation via sampling-based methods (with discussion)". *Bayesian Statistics 4*, J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds., Oxford: Oxford University Press, pp. 147167.

Gelman, A. (2006). "Prior distributions for variance parameters in hierarchical models." *Bayesian Analysis*, 1, 515-534.

Gelman, A., Meng, X.-L., and Stern, H.S. (1996). "Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica, 6*, 733-807.

Gelman, A., Rubin, D.B. (1992). "Inference from Interactive Simulation Using Multiple Sequences". *Statistica Science, 7*, 457-472.

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

Hjort, N.L., Dahl, F.A., & Steinbakk G.H. (2006) "Post-Processing Posterior Predictive p Values," *Journal of the American Statistical Association,* Vol, 101, 1157-1174.

Jiang, J. & Lahiri, P. (2006). "Estimation of Finite Population Domain Means - A Model-Assisted Empirical Best Prediction Approach," *Journal of the American Statistical Association,* Vol, 101, 301-311.

Malec, D. (2005), "Small Area Estimation from the American Community Survey Using a Hierarchical Logistic Model of Persons and Housing Units", *Journal of Official Statistics,* Vol. 21, No. 3, 411-432.

Meng, X.-L. (1994). "Posterior predictive $p-$values." *Ann. Statist.*, 22, 1142-1160.

Molina, I. and Rao, J.N.K. (2010), "Small area estimation of poverty indicators", *Canadian Journal of Statistics,* 38, 369-385.

Natarajan, R. and Kass, R.E. (2000). "Reference Bayesian Methods for Generalized Linear Mixed Models." *Journal of the American Statistical Association,* 95, 227-237.

Neri, L., Ballini, F. and Betti, G. (2005). "Poverty and inequality in transition countries". *Statistics in Transition,* 7, 135-157.

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

Pfeffermann, D., and Burck, L. (1990), "Robust Small-Area Estimation Combining Time Series and Cross-Sectional Data," *Survey Methodology*, 16, 217-237.

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.

Rao, J. N. K. (2003). "Small Area Estimation". Wiley, Hoboken, NJ.

Rao, J.N.K., and Yu, M. (1994), "Small-Area Estimation by Combining Time Series and Cross-Sectional Data," *Canadian Journal of Statistics*, 22, 511-528.

Rubin, D.B. (1984). "Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.,* 12, 1151-1172.

Slud, E. V. and Maiti T. (2011), "Small-area estimation based on survey data from a left-censored Fay-Herriot model", *Journal of Statistical Planning and Inference*, 141, 3520-3535.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde (2002), "Bayesian measures of model complexity and fit (with discussion)". *Journal of the Royal Statistical Society* Series B 64, pp. 583-639.

Tarozzi, A. and Deaton, A. (2009). "Using census and survey data to estimate poverty and inequality for small areas". *Review of Economics and Statistics,* 91, 773-792.

Wieczorek, J., Nugent, C., and Hawala, S., (2012). "A Bayesian Zero-One Inflated Beta Model for Small Area Shrinkage Estimation ". *Proceedings of the Section on Survey Research Methods, Alexandria,* VA: American Statistical Association.