Bayesian Hierarchical Modeling of U.S. County Poverty Rates

Robin Fisher
Jana Asher

November 1, 1999

Abstract

The U.S. Census Bureau Small Area Income and Poverty Estimates program produces biennual intercensal estimates of the povery rates and counts of poor within counties for use in determining the allocation of federal funds to local jurisdictions. Numbers of poor are currently modeled through an empirical Bayes estimation method centered on a linear regression; the dependent variable is a log transformation of the three-year average of the March Current Population Survey (CPS) estimate of the number of poor for each county, and the independent variables are log transformations of administrative data. We assume the variability of the CPS estimates is the sum of a model error term with constant variance, and a sampling error term whose variance is proportional to the inverse of a power of the CPS sample size. Maximum likelihood estimation is used to jointly determine the values of the regression coefficents and the sampling variance components. Problems with the current estimation technique include a loss of data points due to the log transformation for counties whose CPS sample of poor is zero, and the requirement of using decennial census data to estimate the model error variance term. To eliminate these problems and improve the overall quality of our estimates, we have developed a hierarchical Bayesian model which assumes the observed number of poor can be modeled through an approximated probability distribution function with a scaled binomial kernel that is dependent on the underlying poverty rate. This poverty rate, in turn, has a beta prior which relies on a set of parameters that includes the regression coefficents for the administrative data. Posterior probability distributions for regression parameters, variance parameters, and true proportion poor are generated using Markov Chain Monte Carlo techniques. We will discuss the Bayesian model and compare the results of the original and new estimation methods.

1 Introduction

The U.S. government, through the Departments of Education, Health and Human Services, Housing and Urban Development, and Labor, allocates approximately $30 billion in funds annually to programs to aid economically disadvantaged areas of the United States. Approximately $7 billion is distributed under Title I of the Elementary and Secondary Education Act by the Department of Education to school districts with high numbers and proportions of poor children. Traditionally, these funds have been allocated on the basis of small area estimates created from the most recent decennial census, which resulted in poverty figures that were not updated for a decade or more. The fact that poverty not only fluctuates within the span of a decade, but also fluctuates unevenly across different geographic areas of the country (for an example see Dalaker and Naifeh (1998)), suggests that more recent, intercensal estimates of poverty should be available. In 1994, through the "Improving America's Schools Act," the United States Congress required the U.S. Census Bureau to begin to produce estimates of poverty for counties on a biennual basis. Through the Small Area Income and Poverty Estimates program (SAIPE), and with the support of a consortium of five federal agencies, the U.S. Census Bureau has released county-level poverty estimates for income years 1993 and 1995, and will release estimates for income year 1997 in the year 2000.

In order to create county-level estimates, SAIPE uses an empirical Bayes estimation method centered on a linear regression model with a log transformation of a three-year average of the March Current Population Survey (CPS) estimates of the number of poor for each county as the dependent variable. To our knowledge, the March CPS is the best source currently available of direct estimates of poverty for counties in the United States. The CPS sample size, however, is small for most counties, and in any given year only about 1300 counties are included in the CPS sample. We use a three year average of CPS poverty counts in order to increase the number of counties for which a sample is available, and to borrow strength from the years surrounding the target year. Independent variables for the regression are formed from administrative data sources and also undergo a log transformation: these variables include the number of poor from the previous decennial census, the number of poor as aggregated from tax returns, the number of food stamp participants, the population, and the total number of tax returns for each county. The model takes the following form:

$$\mathbf{y}^* = \mathbf{X}\beta + \mathbf{u} + \mathbf{e} \tag{1.1}$$

where $\mathbf{y}^*$ is the vector of the log of the three year average of CPS poor, and $\mathbf{X}$ is the matrix of log values of variables from administrative records. Two error terms are included in the model to accomodate the sampling error associated with the CPS and the model error; $\mathbf{u}$ represents the model error and is distributed Normal$[0, \mathbf{V_u}]$, and $\mathbf{e}$ represents the sampling error and is distributed Normal$[0, \mathbf{V_e}]$. $\mathbf{V_u}$ is assumed to take the form $v_u\mathbf{I}$; $\mathbf{V_e}$ is assumed to be a diagonal matrix whose entries take the form $\frac{\sigma^2}{k_i}$, where k_i is CPS sample size for county i.

Estimation of the variances of the error terms is accomplished through two steps. First the model error variance term is estimated jointly with the coefficients through maximum likelihood estimation by use of a regression equation where the log transformation of the census direct estimates of poverty is substituted as the dependent variable in the place of the CPS estimates, and the census sampling error variance is estimated with a generalized variance function (see U.S. Buerau of the Census (1990) for a description of the generalized variance function). The model is then rerun with CPS poor as the dependent variable and with the model error variance term fixed to the value determined from the census model; maximum likelihood estimation is used to jointly determine the values of the regression coefficents and the sampling variance component (see Fisher (1997)).

Three concerns arise from the implementation of this modeling procedure. The first is that for approximately 200 counties each year, the CPS estimated number poor for the county is zero. These "sample zeros" occur particularly in small counties which have small sample size within the CPS. Our naive strategy for these sampling zeros has been to remove these counties from the regression, thereby losing the information available in these cases. A new modeling strategy that allows the inclusion of these counties is therefore desirable. A second issue is the reliance on the census estimates of poor in order to determine the structure of the error terms; the underlying assumption is that v_u is the same for both the census and CPS data. This is questionable both because the census and CPS use slightly different definitions of poverty, and also because this reliance becomes more problematic the further in time we move from the previous decennial census. A final issue is that the underlying assumption of a gaussian model for the error terms seems to be inappropriate for small counties.

In order to address these issues, we began to explore the use of hierarchical Bayesian modeling techniques to create a model which both fits our data better and is feasible for the production of a large number of estimates on a regular basis. We use for our analysis data for the 1990 income year measuring poverty in related children age 5-17; the CPS variable is the three-year average for income years 1989, 1990, and 1991. Our desire is to keep the model relatively simple while capturing the important features of our data. We consider the size of the sample in a given county fixed in the CPS; a binomial distribution therefore seems more appropriate than more common models for count data, such as the Poisson distribution. In Section 2 we describe our model in detail and the implementation of that model: to capture the complexity of the CPS sample, we rely on an approximated probability distribution function with a scaled binomial kernel for CPS poor given the true underlying poverty rate. Similar approaches to approximating the sampling distribution have been made by Wedderburn (1974) in his discussion on quasilikelihoods and by West (1985) in his develpoment based on deviance functions. We use mostly diffuse priors and hyperpriors; specifically, a beta prior is used for the elements of the poverty rate vector p so that the joint posterior of the hyperparameters can be easily determined. A Metropolis-Hastings algorithm is employed to sample from the joint posterior of the hyperparameters with good results; using these results a sample is drawn for the poverty rates p. In Section 3 we evaluate the results of the simulation and note that the estimates produced match our expectations from other modeling techniques. We conclude that this modeling procedure holds promise for the production of small area poverty estimates in the future, and we will continue to explore this modeling technique.

2 Methods

2.1 Model of Number Poor and Justification of Approximated Probability Distribution Function

The Current Population Survey follows a complex sample design: one primary sampling unit (PSU) is selected from each of 754 strata using probability proportional to population; within each PSU clusters of households are selected after sorting by geographic, demographic, and socioeconomic characteristics. As such, attempting to model the three year CPS average of total poor for county i, y_i, directly as a binomial random variable with total population for county i as estimated by the CPS, n_i, would cause a severe underestimation of the true variance of the sample. Similarly, using the CPS sample directly and modeling it as a binomial distribution with sample size k_i would cause an underestimation of the variance. Direct modeling of the CPS sample design, however, would be difficult. An intermediate solution is to model the design by taking a random variable z_i to represent the number poor drawn from a simple random sample (SRS) with the same information available through the 3 year average of the CPS total poor for county i. Then $z_i \mid p_i \sim bin(m_i, p_i)$, where m_i represents a sample size from a SRS which contains the same information as the more complex CPS design, and p_i represents the underlying proportion of poor. Then y_i is taken to be the scaled value $y_i = z_i/f_i$, and the population for county i is taken to be $n_i = m_i/f_i$. The scaling factor f_i is taken to be greater than zero and describes the amount of information in each estimated person relative to an observation from the underlying binomial distribution; it can be interpreted as a ratio of the variances of the CPS design and a SRS design. The scaling factor therefore corrects for both the overdispersion of the variance and the complexity of the sampling design at once. A direct transformation of the distribution of z_i would yield the following equation:

$$f(y_i|p_i, f_i) = \frac{\Gamma(f_i n_i + 1)}{\Gamma(f_i y_i + 1)\Gamma(f_i n_i - f_i y_i + 1)} p_i^{f_i y_i}(1 - p_i)^{f_i n_i - f_i y_i} \tag{2.1}$$

This is a probability function for y_i only if f_i takes a value such that $f_i n_i$ and $f_i y_i$ are integers for $y_i \in 0, ..., n_i$. If we accept this constraint, a prior distribution for f_i would have a discrete support, and would be difficult to model due to our desire to parameterize the vector of f_i with a small number of parameters. We therefore relax this constraint to allow f_i a continuous support, and anticipate $f_i < 1$.

The effect of this relaxation of the support of f_i is the resulting required relaxation of the support of y_i; we now allow y_i to be continuous on $[0, n_i]$. The result is that the kernel in (2.1) can only be normalized by a function of p_i; it is no longer a legitimate probability function. We therefore treat this function as an approximated probability distribution function: our goal is to both find a normalizing constant that is not a function of p_i in order to allow the approximated probability distribution function to behave as if the moments of y_i were from the underlying binomial kernel, and also to approximate the distribution of y_i well over the range of p_i and f_i. We therefore propose the following form for the approximated probability distribution function of y_i:

$$f^{\star}(y_i|p_i, f_i) = \frac{c_i \Gamma(f_i n_i + 1)}{\Gamma(f_i y_i + 1)\Gamma(f_i n_i - f_i y_i + 1)} p_i^{f_i y_i} (1 - p_i)^{f_i n_i - f_i y_i}, \tag{2.2}$$

where $c_i = f_i + \frac{1}{n_i}$. Our justification is as follows. We first note that the behavior of the approximated probability distribution function with respect to p_i is independent of our choice of scaling factor c_i. For both the scaled binomial distribution and this approximated probability distribution function, $\hat{p}_i = y_i/n_i$. We know that for the scaled binomial distribution, $E(\hat{p}_i|p_i) = p_i$ for any value of $f_i n_i$, and $Var(\hat{p}_i|p_i) = p_i(1 - p_i)/(f_i n_i)$. Asymptotically, we can take this expected value to be true of the approximated probability distribution function as well; we note the consistency of $\hat{p}$:

$$\frac{E(\hat{p}_i|p_i)}{p_i} \stackrel{f_i n_i \to \infty}{\longrightarrow} 1 \tag{2.3}$$

We can also determine the following for the approximated probability distribution function:

$$I(p_i) = -\frac{d^2 log(f^{\star}(y_i|p_i, f_i))}{dp_i^2} = \frac{f_i y_i (1 - p_i)^2 + (f_i n_i - f_i y_i) p_i^2}{p_i^2 (1 - p_i)^2} \tag{2.4}$$

Then,

$$Var(\hat{p}_i) \stackrel{\hat{p}_i \to p_i}{\longrightarrow} \frac{1}{I(p_i)} \tag{2.5}$$

and

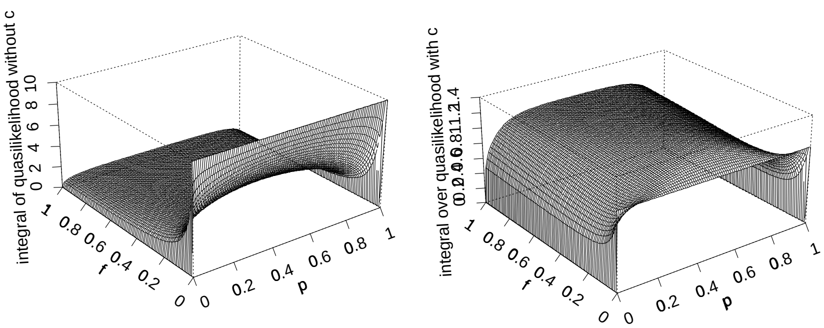
$$\frac{1}{I(p_i)} \stackrel{\hat{p}_i \to p_i}{\longrightarrow} \frac{p_i(1 - p_i)}{f_i n_i} \tag{2.6}$$

with $\hat{p}_i \to p_i$ for large $m_i = f_i n_i$.

The form of c_i is chosen to normalize the approximated probability distribution function as effectively as possible over the joint range of p_i and f_i. We note that $f^{\star}(y_i|p_i, f_i)$ is a normalized distribution function for y_i when f_i is 0 if $c_i = \frac{1}{n_i}$, and it is a normalized distribution function for y_i when f_i is 1 if $c_i = f_i$. Empirical evidence suggests taking $c_i = f_i + \frac{1}{n_i}$ works reasonably well for non-extreme values of p_i. An example follows below:

the surfaces in the following figures represent the value of the integration of $f^*(y_i|p_i, f_i)$ over the range of p_i, taking $f_i \in [0, n_i]$ where n_i is set to 10 for convenience. Note the surface is relatively flat for $c_i = f_i + \frac{1}{n_i}$; for $p_i = .16$, for example, the surface ranges between .9411 and 1.025. The surface when $c_i = 1$ behaves especially poorly as f_i approaches 0.

Effect of normalizing function on intergral of the Quasilikelihood.



The result of taking $c_i = f_i + \frac{1}{n_i}$ is an approximated probability distribution function that behaves appropriately with respect to the moments of y_i and is approximately normalized for most values of f_i and p_i. More rigorous theoretical justification of the approximated probability distribution function is in progress.

2.2 Parametric Form of f_i

Since the value of f_i is dependent on the design of the CPS, which is somewhat constant over counties, we wish a constant parametric form for the set of f_i's. Given $\hat{p}_i$ consistent and unbiased, by the results in (2.3),

$$E(\hat{p}_i|p_i) \rightarrow p_i \tag{2.7}$$

and by (2.6)

$$V(\hat{p}_i|p_i) \rightarrow \frac{p_i(1-p_i)}{f_i n_i} \tag{2.8}$$

then

$$cv^2(\hat{p}_i) = \frac{V(\hat{p}_i|p_i)}{E(\hat{p}_i|p_i)^2} \approx \frac{1-p_i}{p_i f_i n_i}. \tag{2.9}$$

Solving for f_i,

$$f_i \approx \frac{1-p_i}{p_i n_i cv^2(\hat{p}_i)}. \tag{2.10}$$

We choose to model the relvariance of $\hat{p}_i$ based on an equation suggested by Bell and Kramer (1998), in which they derive a generalized variance function for $cv^2(\hat{p}_i)$ using balanced half sample replicates. Specifically, they explore modeling the relvariance of y_i by a function of both the CPS sample size (k_i) and the poverty rate (p_i). Further motivation of this form is given by Zaslavsky (1997), in which he notes that an appropriate function of the relvariance to consider is $\frac{1-p_i}{p_i}$. We therefore propose the following form of the relvariance:

$$cv^2(\hat{p}_i) = \exp(-\gamma_0)k_i^{-\gamma_1}(\frac{1-p_i}{p_i}) \tag{2.11}$$

This leads to

$$f_i \approx n_i^{-1}\exp(\gamma_0)k_i^{\gamma_1} \tag{2.12}$$

where k_i is the CPS sample size and n_i the population size. Note from (2.10) that f_i is just the relvariance of the SRS binomial over the relvariance of y_i for the CPS sample design, which follows the heuristic explanation of this factor given in Section 2.1.

2.3 Priors and Hyperpriors

We take the mixing distribution for p, which is the 'true' poverty ratio in county i:

$$p_i \sim Beta(x_i, n_0 - x_i), \tag{2.13}$$

where $x_i = g(X_i\beta)n_0$. Under this parameterization the expected value of $p_i|\beta, n_0$ is $g(X_i\beta)$. X_i is formed through a principle components analysis of predictors taken to be log values of measures of poverty. Use of principle components analysis reduces the dependence of the β's, which allows for simpler implementation of a Metropolis-Hastings algorithm. The inverse-link function used is

$$g(\mathbf{X}\beta) = \exp(\mathbf{X}\beta). \tag{2.14}$$

The hyperpriors used are as follows:

$$\begin{aligned} n_0 &\sim \text{Gamma}(1/1{,}000{,}000,\ 1/100{,}000{,}000) \\ \beta &\sim \text{Uniform}(R^6) \\ \gamma_0 &\sim \text{Normal}(-5,10) \\ \gamma_1 &\sim \text{Normal}(.5,\ .3) \end{aligned}$$

With the exception of the prior for n_0, these priors are picked to be diffuse when compared to the resulting posterior distributions. The prior on γ_1 is picked based on previous information; in the current SAIPE modeling procedure the sampling variance is believed to be modeled well by the square root of CPS sample size. Examination of the rates at which the tails of the posterior distributions converge to zero indicates the marginal posteriors are proper.

The prior on n_0 is chosen recognizing some characteristics of $p(\vec{y}|\beta, n_0, \gamma)$ taken as a function of n_0. Holding other parameters constant, this distribution increases quickly over small values of n_0, and then becomes quite flat after a certain level. One interpretation for this is that the data indicate that n_0 is more likely not to have a small value and that, above some level, the model is not very sensitive to n_0. The part of the real line where the distribution is increasing as a function of n_0 is well within the region where the prior for n_0 has relatively large mass. Allowing the prior for n_0 to be diffuse results in simulations that do not converge or converge very slowly. The posterior may not be proper for improper priors of n_0.

2.4 Implementation

The form of the approximated probability distribution function and prior for p_i lead to a approximate posterior for the hyperparameters of the following form:

$$p^{\star}(\beta, n_0, \gamma|\vec{y}) = p(n_0, \beta, \gamma)\cdot$$
$$\prod_i \frac{c_i\Gamma(f_i n_i + 1)\Gamma(n_0)\Gamma(f_i y_i + x_i)\Gamma(f_i n_i + n_0 - f_i y_i - x_i)}{\Gamma(f_i y_i + 1)\Gamma(f_i n_i - f_i y_i + 1)\Gamma(f_i n_i + n_0)\Gamma(x_i)\Gamma(n_0 - x_i)}$$

where $x_i = g(X_i\beta)n_0$.

We form the posterior for p_i in the standard way:

$$p(p_i|y_i, \beta, n_0, \gamma) = Beta(f_i y_i + x_i, f_i n_i - f_i y_i - x_i + n_0) \tag{2.15}$$

The posterior distributions of the hyperparameters are simulated using a Metropolis-Hastings algorithm, for which the candidate values are drawn as follows:

1. $n_0^\star$ from a Gamma candidate-generating distribution with coefficient of variance set to .1 and mean to $n_{0,t-1}$. We used this form due to the fact that n_0 can vary greatly in size; determining the ratio of the mean and the standard deviation that produces good mixing is easier than pinpointing the variance.

2. $\beta^\star$ from the candidate-generating distribution $N[\beta_{i-1}, V_\beta]$. The covariance matrix is formed with the cholesky root of the approximate covariance matrix of the posterior marginal for β and is centered on the value of β_{t-1}. The covariance matrix was estimated with short runs of the sampler and some human supervision.

3. γ_0^* from the candidate-generating distribution $N[\gamma_{0,t-1}, .03^2]$.

4. γ_1^* from the candidate-generating distribution $Beta[\gamma_{1,t-1}, .00005^2]$.

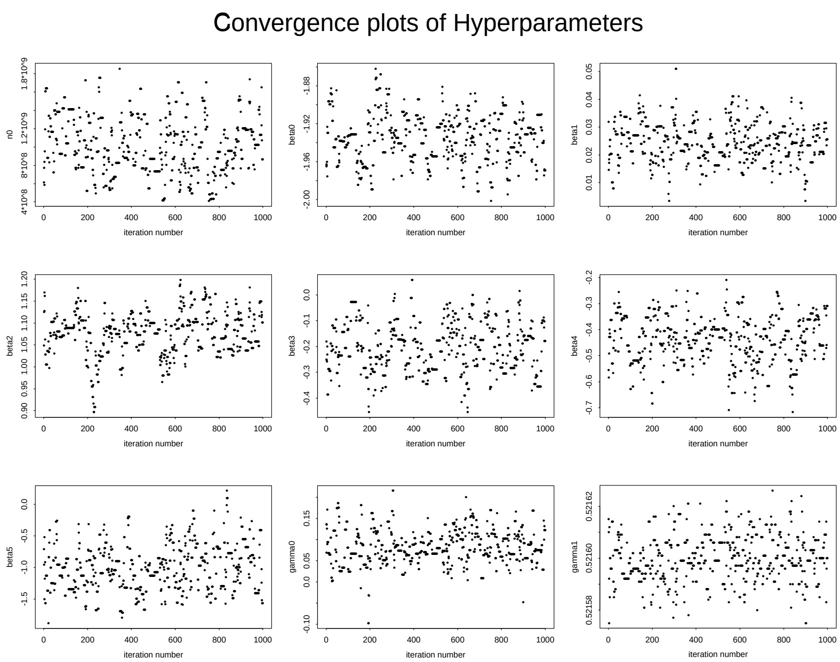
The tightness of the variances for the γ_0 and γ_1 candidate-generating functions are required by the tightness of their posteriors. Once a sample of hyperparameters is drawn, generating a corresponding sample for p_i is done through a simple random draw from the Beta distribution in (2.15).

3 Results

3.1 Posterior Distributions

To calculate the posterior distributions of the hyperparameters, we draw 100,000 samples and keep every 10th sample of the last 10000. Convergence is achieved well before this point, after about 2000 iterations. Alternate runs with varying starting points confirm the convergence behavior in all but one case: if the starting value for γ_1 is too extreme a value, the convergence of all the parameters occurs in a different location, with n_0 decaying to its minimum. Possible explanations include a second local maximum in the joint posterior; further exploration of this anomaly will occur.

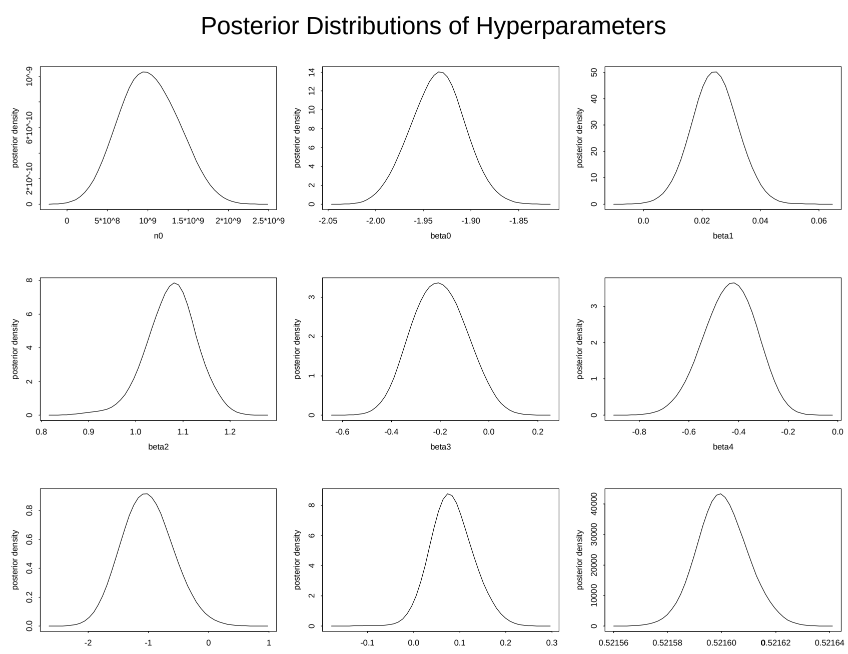
Following are the convergence plots for a typical run:



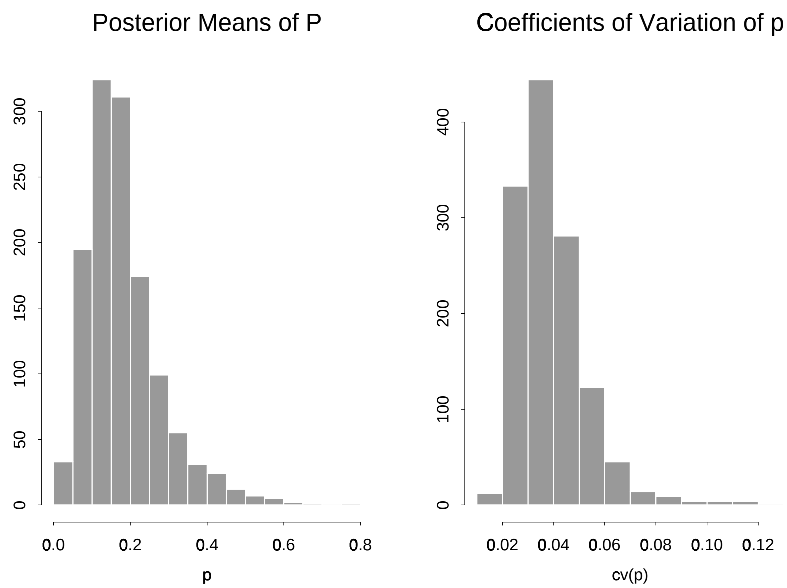
The posterior distributions of the hyperparameters behave well; a summary of their distributions for this run follows:

Parameter	1st Qu.	Median	3rd Qu.	Stand. Dev.
n_0	779,500,000	982,200,000	1,217,000,000	304,311,599
β_0	-1.953	-1.933	-1.919	0.02433934
β_1	0.02014	0.02431	0.02842	0.00663341
β_2	1.044	1.081	1.103	0.04693845
β_3	-0.2818	-0.2024	-0.1378	0.09265293
β_4	-0.495	-0.4301	-0.3619	0.08703785
β_5	-1.295	-1.034	-0.8017	0.3488488
γ_0	0.06009	0.07447	0.1076	0.03919703
γ_1	0.5216	0.5216	0.5216	.000008102818

The size of the n_0's in the posterior distribution can be interpreted as follows: large n_0 means that the prediction equation $X_i\beta$ carries more "weight" in the formulation of the posterior values of p_i than the data point y_i. As n_0 approaches infinity, the actual value of n_0 makes less and less difference as to its importance in the posterior distribution of p_i. Note that γ_1's posterior distribution is sufficiently tight that its variability is only noticible past four significant digits. Interpretation of the β's is left to future research. Plots of the estimated marginal posterior density functions for a typical run follow.



The behavior of the posterior distributions of the p_i's is encouraging; we note the distribution of the posterior means in the first plot below, and the tight variances on the distributions of the p_i's in the second plot below. A detailed analysis of the characteristics of the posterior distributions of the p_i, including the change in the shape and moments of the posteriors given underlying county size and CPS sample size, is saved for future work.

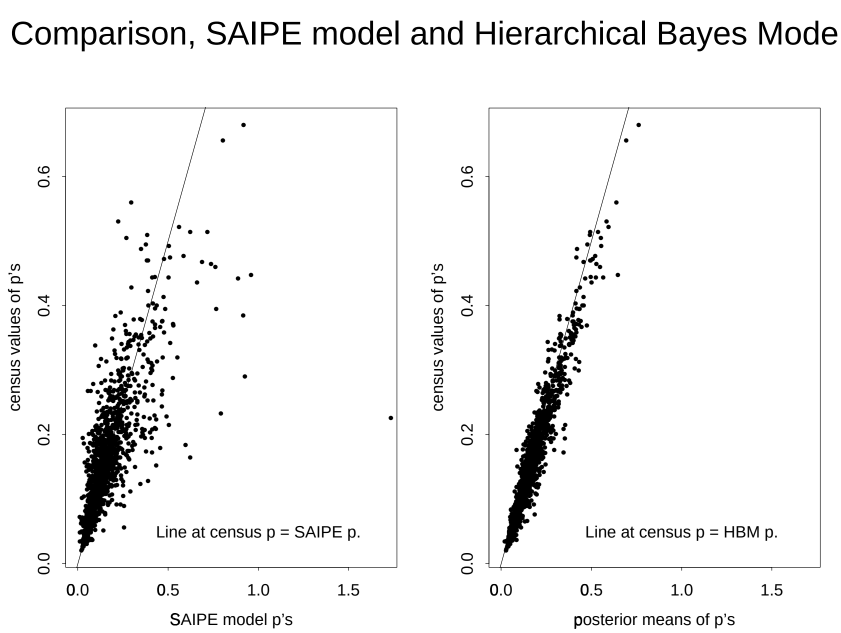


3.2 Evaluation of the Model

Because we have an established modeling procedure, and we can compare our results to the 1990 census, we have several ways of testing the quality of our modeling procedure. We must first discuss some of the differences between the census and CPS measurements of poverty. The CPS measurement of population differs from the census due to a difference in measurement of the "poverty universe", or the people that are included when a number poor within a particular county is modeled. The CPS poverty universe for related children age 5-17 is bigger than the census population count for this group; the CPS poverty universe includes unrelated subfamily members and residents of college dormitories, while the census population for this group does not. The current SAIPE model is for total poor as measured by the CPS; to create proportions of poor for counties from this model, we have traditionally divided by the census population for that county due to our belief that there is less variability in the census population measurement than the CPS population measurement. For the hierarchical Bayes modeling procedure, using the CPS poverty as the population count for county i better matched our goal of modeling the poverty rates (p_i's). For the purpose of comparing the hierarchical Bayes modeling results to the current SAIPE estimates and the census, however, we have also run the model using the census populations for the n_i's. In any comparisons of CPS model results with the census, we must assume some variation due to the fact that they measure slightly different populations.

We first measure the absolute relative difference between the posterior means of the p_i's and the poverty rates given by the 1990 census to test the quality of our model. Using a typical run, the census population as n_i yields an absolute relative difference of .1451, which is favorably comparable to the results from the current SAIPE model (.1625). Using the CPS poverty universe for n_i yields an absolute relative difference of .1293; testing with different simulation runs yields results between .1267 and .1299. We also check the mean value of the posterior means of the p_i's; using the census population for n_i yields a mean of .1814, which is comprable to our expectation given the current SAIPE model. Using the CPS poverty universe gives us .1640, which is not surprising given the CPS poverty universe is larger than the census universe. Testing with different simulation runs yields results between .1634 and .1641.

The following figure shows a comparison of both the current SAIPE model's predicted poor (using the CPS poverty universe in the denominator) and the posterior means of the p_i's (using the CPS poverty universe for n_i) to the 1990 census p_i's for a typical simulation run. We note that the posterior means follow the census p_i's more closely than the p_i's from the current SAIPE model. We also note that due to the fact that the current SAIPE model predicts y_i's independently of n_i, it produces several extremely unlikely values for proportion poor, including one value greater than one.



Supporting evidence towards the validity of the hierarchical Bayes model can also be found in the behavior of the posterior distribution of γ_0; this posterior distribution has a very tight variance around .5216. This result is supported by research done by Fisher and Asher (1999) suggesting that the square root of CPS sample size is a good function for the CPS sampling variance for the three-year average of number poor related people age 5-17.

Finally, we check the model by comparing the data to the posterior predictive distribution; using the methodology outlined in Gelman et al. (1995). Taking a test quantity T as the χ^2 discrepancy, and using a simulation-based measure of the Bayes p-value $Pr(T(y^{rep}, p) \geq T(y, p)|y)$, we have achieved values ranging between .48 and .57. These tail-probabilities are very far from extreme, and so we are positively encouraged by the fit of the model.

4 Discussion

We are satisfied with the preliminary results of our research into Bayesian hierarchical modeling of U.S. county poor, and will continue to explore this modeling procedure as an alternative to the traditional empirical Bayes approach. Future work will include testing this modeling procedure on data from intercensal years. We also will continue work on the theoretical justification of the approximated probability distribution function. Finally, we are intrigued by concept of formulating a function of f_i to more realistically model the scaled binomial distribution.

Acknowledgments

Robin Fisher and Jana Asher are mathematical statisticians in the Housing and Household Economic Statistics Division (HHES) of the U.S. Census Bureau. This paper reports the results of research and analysis undertaken by U.S. Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Census Bureau. It is released to inform interested parties of current research and to encourage discussion. The authors wish to acknowlege the support and assistance of William R. Bell, Pat Cardiff, Donald Malec, and Paul Siegel.

References

Bell, W. and Kramer, M. (1998). Generalized Variance Functions for Sampling Error Variances of Direct CPS County Poverty Estimates. Internal Memo, SAIPE, U.S. Census Bureau.

Citro, C. and Kalton, G. (1999). Small-Area Estimates of School-Age Children in Poverty, Interim Report 3: Evaluation of 1995 County and School District Estimates for Title I Allocations. Washington, DC: National Academy Press.

Dalaker, J. and Naifeh, M. (1998). Poverty in the United States: 1997. *U.S. Census Bureau, Current Population Reports, Series P60-201,* Washington, DC: U.S. Government Printing Office, ix-xii.

Fisher, R. (1997). Methods Used for Small Area Poverty and Income Estimation. *1997 Proceedings of the Section on Government Statistics and Section on Social Statistics.* Washington, DC: American Statistical Association, 177-182.

Fisher, R. and Asher, J. (1999). Alternate CPS Sampling Variance Structures for Constrained and Unconstrained County Models. Internal Technical Report Series #1, SAIPE, U.S. Census Bureau.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995). Bayesian Data Analysis. New York: Chapman and Hall.

Siegel, P. (1995). Developing Postcensal Income and Poverty Estimates for all US Counties. *1995 Proceedings of the Section on Government Statistics.* Washington, DC: American Statistical Association, 166-171.

U.S. Bureau of the Census (1993). 1990 Census of Population, Series CP(2). *Social and Economic Characteristics, Missouri, Section 1 of 2.* Washington, DC: U.S. Government Printing Office, C2-C5, C10-C11.

Wedderburn, R.W.M. (1974). Quasilikelihood functions, generalized linear models, and the Gauss-Newton Method. *Biometrika*, **61**, 439-448.

West, M. (1985). Generalized Linear Models: Scale Parameters, Outlier Accomodation and Prior Distributions. *Bayesian Statistics 2.* North Holland: Eslevier Science Publishers, 531-558.

Zaslavsky, A. (1997). A Small-Area View of Poverty from the Panel. *1997 Proceedings of the Section on Government Statistics and Section on Social Statistics.* Washington, DC: American Statistical Association, 189-191.