ALTERNATE SCALING PARAMETER FUNCTIONS IN A HIERARCHICAL BAYES MODEL OF U.S. COUNTY POVERTY RATES

Jana Asher and Robin Fisher *
Jana Asher, U.S. Census Bureau, Room 124-2, Washington, DC 20233-8500

Key Words: Hierarchical Bayes, Small Area Estimation, Scaled Binomial Model

Abstract: The U.S. Census Bureau Small Area Income and Poverty Estimates program (SAIPE) currently uses an empirical Bayes estimation method similar to the Fay and Herriot (1979) model to produce biennial intercensal estimates of the poverty rates and counts of poor within counties. The dependent variable is formed from a three-year average of the March Current Population Survey (CPS) supplement, and the independent variables are formed from administrative data. The model includes two error terms. Problems with this 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 address these problems, a hierarchical Bayes model based on a scaled binomial kernel has been developed (see Fisher and Asher (1999)). The scaling factor corrects for both the overdispersion of the variance and the complexity of the CPS sample design. This paper will discuss the effect of different scaling factor functions on the implementation and quality of this proposed model.

1 Introduction

The U.S. government, through the Department of Education, allocates approximately $7 billion under Title I of the Elementary and Secondary Education Act to school districts with high numbers or proportions of poor children. 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 biennial basis. Through the Small Area Income and Poverty Estimates program (SAIPE), 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. The CPS sample size, however, is small for most counties, and for any given estimate year only about 1300 counties are included in the CPS sample. 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:

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

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 accommodate 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 represents CPS sample size for county i.

Three concerns arise from the implementation of this model. The first is that for approximately 200 counties each year, the CPS estimated number poor for the county is zero. Our naive strategy for these sampling zeros has been to remove these counties

*This paper reports the results of research and analysis undertaken by U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The authors wish to acknowledge the support and assistance of William R. Bell, Beverly Causey, Donald Luery, Donald Malec, and Paul Siegel.

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 use of census estimates of number poor for determining the structure of the error terms; the underlying assumption is that $\mathbf{V_u}$ is the same for both the census and CPS data, and we use census data to estimate v_u before using CPS data to determine values for the coefficients and σ^2. 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, especially for small counties.

In order to address these issues, we test 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. 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. 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. Preliminary results were presented in Fisher and Asher (1999). This paper continues that work by improving the procedure for sampling from the posterior and attempting different functions for a scaling parameter in the scaled binomial kernel.

2 Methodology

Let $y_{ij}^{\star}$ be the CPS estimate of total poor for county i in income year j; this variable is a weighted sum of the number poor found in the CPS sample. The variable available for analysis is y_i, a weighted three year CPS average of the total number of poor for county i.[1] We take k_{ij} to be the CPS sample size for county i, year j, and $k_i = \sum k_{ij}$ to be the derived sample size for y_i. Also available is n_i, an estimate of the total population for county i.[2] Because the Current Population Survey follows a complex sample design and, additionally, y_i is a weighted three-year average, attempting to model y_i directly as a binomial random variable with total population n_i would cause a severe underestimation of the true variance of y_i. Similarly, using the CPS sample directly and modeling number of poor in CPS sample for county i and year j as a binomial distribution with sample size k_{ij} would cause an underestimation of the variance of the weighted sum $y_{ij}^{\star}$ due to the complexity of the CPS sample design. Additionally, attempting to create a modeling procedure that accounts for the CPS sample design and the effect of the weighted three year averaging would be prohibitively complicated.

An intermediate solution is to model the sample design by taking a random variable z_i to represent the number poor drawn from a simple random sample (SRS) with a sample size that provides the same information as is available through the three year average of the CPS total poor for county i. Then $z_i \mid p_i \sim bin(p_i, m_i)$, where m_i represents a sample size of a SRS which contains the same information as the more complex CPS design and weighted averaging procedure, 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 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 variance associated with y_i (due to the CPS design, weighting in the summation to create the $y_{ij}^{\star}$'s and three year weighted averaging) and a SRS design. The scaling factor therefore corrects for both the overdispersion of the variance and the complexity of the sample design at once. A direct transformation of the distribution of z_i would yield the following equation:

$$g(y_i|p_i, f_i) = \frac{\Gamma(f_i n_i + 1)}{\Gamma(f_i y_i + 1)\Gamma(f_i n_i - f_i y_i + 1)} \cdot p_i^{f_i y_i}(1 - p_i)^{f_i n_i - f_i y_i} \quad (1)$$

where y_i has discrete support $(0, \frac{1}{f_i}, \frac{2}{f_i}, ..., n_i)$ given a particular value of f_i. This is a probability function for y_i only if f_i takes a value such that $f_i n_i$ is an integer. For this to occur, f_i must have the discrete

[1]Note that y_i is not the product of a log transformation, as $y_i^{\star}$ is in the current SAIPE modeling procedure. Also note that y_i is described in greater detail later in this paper.

[2]More technically, n_i is a weighted average of the CPS poverty universes for related children age 5-17 for income years 1989, 1990, and 1991; n_i is described in greater detail later in this paper.

support $f_i : n_i \bmod f_i^{-1} = 0$. Furthermore, the support of f_i must account for the possible values of y_i in the data. This makes f_i 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 direct effect of this relaxation of the support of f_i is the resulting required relaxation of the support of y_i; y_i can now potentially take values on the continuous range $[0, n_i]$.

2.1 Approximated PDF

The result of allowing y_i to have a continuous range is that arguments of the normalizing function for the kernel in (1) must include p_i; (1) 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 use the following form for the approximated probability distribution function of y_i:

$$\begin{aligned} g^\star(y_i|p_i, f_i) &= \frac{c_i\Gamma(f_i n_i + 1)}{\Gamma(f_i y_i + 1)\Gamma(f_i n_i - f_i y_i + 1)} \\ &\quad \cdot \; p_i^{f_i y_i}(1 - p_i)^{f_i n_i - f_i y_i}, \end{aligned}$$

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 $\hat{p}$:

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

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

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

Then,

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

and

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

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 $g^\star(y_i|p_i, f_i)$ is a normalized (uniform) distribution function for y_i when $f_i \to 0$ and (consequently) $c_i \to \frac{1}{n_i}$, and it is a normalized (binomial) distribution function with discrete support for y_i when f_i is 1 if $c_i = f_i$. We can also show that as $f_i n_i \to \infty$, $g^\star(y_i|p_i, f_i)$ is approximately normalized by $c_i = f_i$. Finally, empirical evidence suggests taking $c_i = f_i + \frac{1}{n_i}$ works reasonably well for non-extreme values of p_i, as is presented in Fisher and Asher (1999).

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.

2.2 Proposed models for f_i

The parametric form proposed for f_i in Fisher and Asher (1999) is as follows:

$$f_i = n_i^{-1} \exp(-\gamma_0) k_i^{\gamma_1}$$

where $k_i = \sum k_{ij}$, and k_{ij} is the CPS sample size (# of households) for county i, year j. This form is based on a generalized variance function for $cv^2(\hat{p}_i)$ from Bell and Kramer (1998) and the suggestion to base the squared coefficient of variation of $\hat{p}_i$ on a function of $\frac{1-p_i}{p_i}$ from Zaslavsky (1997).

An alternative approach to determine the parametric form of f_i is to start with the form of the three year CPS average used in the SAIPE modeling procedure. We note:

$$\begin{aligned} n_i &= \sum_j (\frac{k_{ij}^\star}{k_i^\star}) n_{ij}^\star \\ y_i &= \sum_j (\frac{k_{ij}^\star}{k_i^\star})(\frac{y_{ij}^\star}{n_{ij}^\star}) \cdot n_i \end{aligned}$$

where n_i is the SAIPE population estimate, y_i is the three year CPS average, $k_i^\star = \sum_j k_{ij}^\star$, $k_{ij}^\star$ is the number of households in county i in year j in the CPS sample that contain related children aged 5-17, $n_{ij}^\star$ is the CPS poverty universe estimate for county i in year j, and $y_{ij}^\star$ is the CPS estimate of number of poor in county i in year j. We note that for year j, the variance for county i can be modeled as follows:

$$Var(y_{ij}^\star) = \frac{(n_{ij}^\star)^2 p_{ij}(1-p_{ij})}{k_{ij}} \text{deff}(y_{ij}^\star)$$

Now, assuming the first two moments for p_{ij} are consistent for every year j, and assuming the same underlying distribution of the design effect for every year, we can determine the variance for y_i:

$$\begin{aligned} Var(y_i) &= Var(\sum_j (\frac{k_{ij}^\star}{k_i^\star})(\frac{y_{ij}^\star}{n_{ij}^\star}) \cdot n_i) \\ &= (\frac{n_i^2}{(k_i^\star)^2}) Var(\sum_j (\frac{k_{ij}^\star y_{ij}^\star}{n_{ij}^\star})) \\ &= (\frac{n_i^2}{(k_i^\star)^2}) u_i p_i (1-p_i) \text{deff}(y_i^\star) \end{aligned}$$

where u_i is $(\sum_j (\frac{(k_{ij}^\star)^2}{k_{ij}}) + .9\frac{k_{i1}^\star k_{i2}^\star}{\sqrt{k_{i1}k_{i2}}} + .9\frac{k_{i2}^\star k_{i3}^\star}{\sqrt{k_{i2}k_{i3}}})$, and $\text{deff}(y_i^\star)$ is the design effect for county i taken to be common across the three years. To find u_i, we assume the correlation for adjacent year CPS estimates is .45 and for non-adjacent year CPS estimates is 0, as given in U.S. Bureau of the Census (1995). Equating this to the form for $Var(\hat{p}_i)$ from our model yields

$$\frac{p_i(1-p_i)}{f_i n_i} = \frac{1}{(k_i^\star)^2} u_i p_i (1-p_i) \text{deff}(y_i^\star)$$

This yields

$$f_i = \frac{(k_i^\star)^2}{n_i u_i} \frac{1}{\text{deff}}$$

There are several options for modeling the design effect in f_i. One is to assume a constant form across counties, e.g. deff $= e^{\gamma_0}$. Another is to assume that the design effect varies by sample size, e.g. deff $= (k_i^\star)^{\gamma_1} e^{\gamma_0}$.

For this paper, we have chosen to focus on the four potential parameterizations of f_i listed below:

Model	f_i
1	$\frac{(k_i)^{\gamma_1}}{n_i} \exp(-\gamma_0)$
2	$\frac{(k_i^\star)^{\gamma_1}}{n_i} \exp(-\gamma_0)$
3	$\frac{(k_i^\star)^2}{n_i u_i} \exp(-\gamma_0)$
4	$\frac{(k_i^\star)^2}{n_i u_i} (k_i^\star)^{-\gamma_1} \exp(-\gamma_0)$

Model 1 is the original as stated in Fisher and Asher (1999). In our procedure, all four models use the 1259 counties for which $k_i^\star$ is greater than zero.

2.3 Priors and Hyperpriors

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

$$p_i \sim Beta(\exp(X_i\beta)n_0, (1-\exp(X_i\beta)n_0),$$

Under this parameterization the expected value of $p_i|\beta, n_0$ is $\exp(X_i\beta)$. X_i is formed through a principal components analysis based on the correlation matrix of predictors taken to be log values of measures of poverty. Use of principal components analysis reduces the dependence of the β's, which allows for simpler implementation of a MCMC algorithm. $\exp(X_i\beta)$ represents the inverse-link function. The hyperpriors used for each of the four f_i's are as follows:

For all four models:

$$n_0 \sim \text{Gamma(1/1,000,000, 1/100,000,000)}$$
$$\beta \sim \text{Uniform}(R^6)$$

This parameterization of the prior for n_0 yields a mean of 100 and a variance of $1X10^{10}$.

For Models 1 and 2:

$$\gamma_0 \sim \text{Normal(-5,10)}$$
$$\gamma_1 \sim \text{Gamma(5/3, 5/6)}$$

The prior for γ_1 has a mean of .5 and a variance of .3. For Models 3 and 4:

$$\gamma_0, \gamma_1 \sim \text{Normal(0,5)}$$

The priors for all parameters were picked to be diffuse, and are diffuse compared to the resulting posterior distributions. The prior on γ_1 for Models 1 and 2 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. Even so, this prior is diffuse compared to the posterior distributions for both models.

2.4 Implementation

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

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

where $g_i = \exp(X_i \beta) n_0$.

We form the posterior for p_i in the standard way:

$$p(p_i|y_i, \beta, n_0, \gamma) = Beta(f_i y_i + g_i, f_i n_i - f_i y_i - g_i + n_0) \quad (2)$$

The posterior distributions of the hyperparameters are simulated using a multi-step Metropolis algorithm. Within each iteration, a gaussian candidate generating function is formed for each parameter separately, whose mean is the previous value for the parameter. Then, identically for each iteration, a jump for each parameter is determined as follows. Let θ be a vector of the hyperparameters. Let $\theta_{-\mathbf{k}}$ represent the vector of hyperparameters with θ_k removed. Let $\theta_k^\star$ be a draw of parameter k at iteration t. Then

$$r = \frac{p^\star(\theta_k^\star, \theta_{-\mathbf{k}}^{\mathbf{t-1}}|\vec{y})}{p^\star(\theta_k^{t-1}, \theta_{-\mathbf{k}}^{\mathbf{t-1}}|\vec{y})}$$

The jumping rule becomes: $\theta_k^t = \theta_k^\star$ with probability min(r, 1), otherwise $\theta_k^t = \theta_k^{t-1}$. To test convergence, we create chains of length 100,000, remove the first 5000 iterations, take every tenth value from the remaining 95,000 iterations, and test the posterior draws of the hyperparameters using an Augmented Dickey-Fuller Test. Jump rates are between 20 and 25 percent for most parameters. For several of our models, convergence is also tested by starting several chains at different parameter values.

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). Model fit is tested by a comparison to census values for proportion poor for 1990, and by a comparison of the posterior predictive distribution to the y_i's.

3 Results

The chart below shows results for the four models. The column labeled "M" represents the model number. Posterior distributions for the β's are not listed due to confidentiality constraints. The β's for the four models are quite similar, however: Models 3 and 4 are identical in the first two significant digits, and Models 2, 3, and 4 are closer to each other than Model 1. We would expect the different parameterizations of f_i to have minimal effect on the β's, so this result is not surprising.

Posterior Moments for n_0, γ_0, γ_1

	n_0		γ_0		γ_1	
M	μ	σ	μ	σ	μ	σ
1	395.9	287.5	.8617	.2024	.7543	.0501
2	486.5	333.1	-.1296	.1388	.7356	.0452
3	53.81	10.27	1.17	.0763	NA	NA
4	56.73	16.19	1.16	.1685	.0038	.0686

In the two models for which the parameters inform us about CPS design effect, we find that using the posterior mean for γ_0 gives us a deff of 3.18 for Model 3 and a deff close to 3.22 for Model 4. Note that the parameter γ_1 in Model 4 represents a multiplicative factor of 1 to 1.03 in the final CPS design effect estimate, suggesting that a function for the design effect based on number of households with related children aged 5-17 in CPS sample is not effective.

In the posterior distribution of the p_i's, n_0 represents the weight on the equation of predictive variables (on $\exp(X\beta)$), and $f_i n_i$ represents the weight on the $\hat{p}_i$'s. The values for n_0 in comparison to $f_i n_i$ indicate whether a model relies on the regression equation more heavily than the CPS estimates.

To test for model fit, we calculate the mean of the posterior means for the p_i's, the mean relative difference from the census poverty proportions for 1990, and the absolute mean relative difference from the census poverty proportions for 1990. These three statistics allow an external validation of the models; we use the census poverty proportions as an imperfect "gold standard." We compare these results to results from the current SAIPE model described in the introduction, using the same 1259 counties as used for the modeling procedures described in this paper. S1 represents estimates before raking to state totals, and S2 represents estimates after this raking.

We also perform a goodness of fit test using the statistic $T = \sum(y_i - E(y_i))^2/Var(y_i)$ for both the original CPS y_i's and for draws from the posterior predictive distribution for the y_i's. We determine these statistics for every iteration, and then empirically determine $Pr(T(y^{rep}, \theta) \geq T(y, \theta))$ as described in Gelman et al. (1995).

Fit Statistics

M	Posterior Mean, $\bar{p}_i$	Mean Rel. Diff.	Abs. Mean Rel. Diff.	Fit Stat.
1	.1651	.0120	.1260	.6245
2	.1664	.0275	.1265	.5925
3	.1653	.0198	.1785	.6103
4	.1652	.0195	.1765	.6237
S1	N/A	-.0027	.1346	N/A
S2	N/A	.0396	.1396	N/A

All four models produce posterior means of $\bar{p}_i$ that are identical to two significant digits and similar mean relative differences. Absolute mean relative differences for Models 3 and 4, however, are considerably larger than those for Models 1 and 2, and for the current SAIPE modeling procedure before and after raking. While the results for Models 3 and 4 may seem discouraging, the CPS universe is different than the census universe. As a result, poverty proportions measure a different population for each, and although we use the census poverty estimates as a check for model fit, we do not believe that CPS and census poverty estimates should be identical. A remaining question is what "distance" represents the true difference between CPS and census proportion poor.

Finally, a note: we know now that the results listed in Fisher and Asher (1999) for n_0, γ_0, and γ_1 do not represent accurate posterior distributions due to an error in the random number generation for the candidate generating functions used to produce that paper's results. The impact of that error on posterior distribution of the p_i's, however, was minimal. Results in this paper for Model 1 should supersede results given in Fisher and Asher (1999); further information will appear in an addendum to that paper.

Future work will include formulating a discrete scaling parameter function, removing the need for an approximated pdf.

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.

Dickey, D. and Fuller, W. (1981). Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root. *Econometrica*, **49**, 1057-72.

Fay, R. E. and Herriot, R. (1979). Estimates of Income for Small Places: An Empirical Bayes Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, **78**, 269-277.

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). Bayesian Hierarchical Modeling of U.S. County Poverty Rates. See www.census.gov/hhes/www/saipe/asapaper/FisherAsher99.pdf

Fisher, R. and Asher, J. (2000). 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.

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.

U.S. Bureau of the Census (1995). Current Population Reports, Series P60-188, *Income, Poverty, and Valuation of Noncash Benefits: 1993, Appendix E.* Washington, DC: U.S. Government Printing Office, E1-E9. See Table E-7.

U.S. Bureau of the Census (2000). *Current Population Survey, Technical Paper 63, Design and Methodology.* Washington, DC: U.S. Government Printing Office.

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.