Errors-in-Variables Model for County-level Poverty Estimation*

Robin Fisher
U.S. Census Bureau
Housing and Household Economic Statistics Division
Room 1462, Building 3
robin.c.fisher@census.gov

December 16, 2003

Abstract

The Small Area Income and Poverty Estimates (SAIPE) program at the U.S. Census Bureau has a model for poverty which relates direct estimates from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey to various Administrative Records (AR) and the last decennial census through a regression model with random effects.

In this paper a Hierarchical Bayes (HB) model is described. Various data are modeled as functions of poverty. Further, the variances of the data sources, conditioned on poverty, are modeled explicitly. This avoid the well-known problems associated with regressions when the predictors are measured with error. Further, the model is easily extended to handle poverty-related data from other surveys or AR, such as the American Community Survey or data from the National School Lunch Program (NSLP). The model is parameterized in an easily interpreted way, which should make it easier to understand the relationships between the variables. The model is applied to the estimation of poverty for SAIPE and there is some discussion of the fit of the model and the interpretation of the parameters.

1 Introduction

The Small Area Income and Poverty Estimates (SAIPE) program at the U.S. Census Bureau has a model for poverty which relates direct estimates from the

*This paper reports the results of research and analysis undertaken by the 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 views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

Annual Social and Economic Supplement (ASEC) of the Current Population Survey to various Administrative Records (AR) and the last decennial census through a regression model with random effects. See U.S. Census Bureau (2003).

In the current SAIPE model, the AR are modeled as if they have a negligible random error, but this assumption has not been examined. Further, while one may believe that AR have negligible sampling error, it may be that the behavior of people in the various counties, given true poverty, may be modeled as a random effect. It is known that data from the decennial census long form has non-negligible variance, which itself varies among the counties. There is also interest in including new covariates which may have variable reliability or which may be absent in some counties. Examples include data from the American Community Survey (ACS) and data from the National School Lunch Program (NSLP). In any of these cases, it is desirable to model the variances in the model.

In this paper, a Hierarchical Bayes (HB) model is described in which the various measures of poverty available to SAIPE are modeled so their expectations are functions of poverty and their variances are modeled explicitly. This avoids the well-known problems associated with regressions when the predictors are measured with error. Further, the proposed model is easily extended to handle other poverty-related data from other surveys or AR. It also allows for an easily-interpreted parameterization of the model. The model, the data, and some simple applications of the model are also described. This includes the estimation and model fitting procedure, some results, the final estimates of the parameters, and the fit statistics. The parameters were constructed to be easily interpreted, so the success of that construction is discussed as well.

2 Data and Model

2.1 Data and Variable Definitions

A description of the model follows for the estimation of the log number of 5- to 17-year old children in families in poverty and all people in poverty in 1989. The log number in poverty will generically be refered to as LNP, modified by age group where necessary to avoid ambiguity. The SAIPE project used 1989 for the evaluation of the original SAIPE model for poverty; it is used here to preserve some comparability of the evaluations. The interpretations of some of the parameters are easier and more interesting when the relevant poverty definition is similar, and this motivates the experiment for LNP for all ages.

Detailed descriptions of the data are given by the U.S. Census Bureau (2003) and in other places. Brief definitions follow.

1. μ_i is the "true" LNP of 5-17 year old children or for all ages in county i.
2. X_{0i} is the ASEC direct estimate of the LNP in county i.
3. X_{1i} is the log number of children in poverty or total exemptions indicating poverty.

4. X_{2i} is the log number of Food Stamp (FS) recipients.

5. X_{3i} is the log number of nonfilers in county i. This is log (total population-total exemptions).

6. X_{4i} is the LNP measured by the previous decennial census.

7. $X_{cen,i}$ is the LNP measured by the current decennial census.

8. Z_i is the population of the 5-17 year old children or total population in county i.

2.2 General Description of the Model

Assume that there are M small areas; the goal is to estimate the quantities $\mu_i, i = 1, ..., M$, which are, in the present application, the LNPs for small areas indexed by i. There are J random variables, each of which is observed for some of the small areas; denote the jth random variable for small area i as X_{ij} , $i = 1, ..., M$, $j = 0, ..., J$. Conditioned on μ_i and some parameters, the X_{ij}'s have a normal distribution.

$$X_{ij}|\mu_i, \theta_j, V_{ij} \sim N(\theta_{j1}\mu_i + \theta_{j2}, V_{ij}).$$

The θ-parameters describe the relationship of the conditional expectation of μ_i and X_{ij} given μ_i. If X_{ij} is conditionally unbiased for μ_i, $\theta_{j1} = 1$ and $\theta_{j2} = 0$. Make this assumption for $j = 0$; this means that the ASEC, when available, is unbiased for the true LNP in the county. Consider the X_i's as estimators for LNP; then θ_{j1} and θ_{j2} are interpretable as bias parameters. That is their reference in this paper. This assumption also makes the θ-parameters identifiable. The variance parameter, V_{ij}, will be modeled separately for variable j. The parameters, μ_i, $i = 1, ..., M$ have a normal distribution, conditioned on other parameters η and v_μ, given by

$$\mu_i|\eta, v_\mu \sim N(\nu_i, v_\mu),$$

where $\nu_i = \eta + Z_i$. The parameter η can be interpreted as the national log poverty rate, and has the distribution

$$\eta = N(m_\eta, V_\eta).$$

Models for the variance terms V_{ij} appear in the list below. This term extends the idea of the sampling variance in ASEC to the AR; for AR, one might think of this variance as the sum of a sampling error variance and a random effect variance. The random effect variance describes the variability of the behavior of people filing tax forms or deciding whether to participate in food stamp programs.

1. The variance of the direct ASEC LNP estimate in county i is inversely proportional to the square root of the sample size.

$$V_{i0} = v_0(k_i^{-\frac{1}{2}}).$$

This is the model in the official SAIPE estimates (U.S. Census Bureau 2003); this form for the variance model was arrived at empirically.

2. The variance of the log number of exemptions for children in poverty or total exemptions given μ_i is inversely proportional to a function of the population: $V_{i1} = V_1/f(N_i)$. Here let $f(N_i) = N_i^{\frac{1}{2}}$.

3. The variance of the log number of FS recipients is constant: $V_{i2} = V_2$.

4. The variance of the log number of nonfilers is inversely proportional to a function of the population: $V_{i3} = V_3/f(N_i)$. In these experiments let $f(N_i) = N_i^{\frac{1}{2}}$. This allows for decreasing conditional standard deviations of the log tax variables or, approximately, decreasing coefficients of variation (cv), as the population increases. This function is motivated by exploratory analysis, but there is evidence, reported below, that it does not increase quickly enough in N_i.

5. Two models are considered for the variance of the LNP from the previous and current decennial censuses.

   - Decennial Census Variance Model 1: The total variance model for the decennial census LNP borrows the form of the census generalized variance function (GVF), transformed to describe the squared coefficient of variation, which is approximately the variance on the log scale. This model simplifies the function, so parameters specific to states and places with various sampling intervals are neglected. For a detailed description of the census GVF, see U.S. Census Bureau (1992).

$$V_{4i} = \gamma_1(\frac{1}{X} - \frac{1}{N}) + v_t.$$

   Note that, if the GVF is correct, γ_1 should be indexed by the counties and, for county i, $\gamma_{1i} = 5DEF_i$, where DEF_i contains factors for the sampling intervals of the county and a factor for the state. The multiplier 5 is part of the definition of the GVF and is the product of the sampling interval and the finite population correction. The extra variance term v_t represents an added variance for the passage of time between the decennial census and the year of interest. The analogous form is used for the current decennial census, but the prior distribution for v_t is chosen so nearly all of the probability mass is near zero.

- Decennial Census Variance Model 2: The total variance for the previous decennial census LNP is equal to the variance derived from decennial census GVF plus a term to represent the change in the real LNP between 1979 and 1989.

  $$V_{4i} = \gamma_1 GVF_i + v_t,$$

  where GVF_i is the variance from the decennial census GVF, transformed to apply to the log scale. This is approximately the squared coefficient of variation for the decennial census number in poverty. Again, there is an additive term to represent a random effect for the passage of time.

  This model derives from the assumption that the transformed decennial census GVF is correct to a multiplicative constant and that the movement of the LNP through time is a diffusion process. Both of these assumptions are of course suspect, but model checks, described below, do not identify this as a failure in the model.

The direct estimate from the current decennial census, given the true LNP, has a normal distribution.

$$X_{cen,i}|\mu_i, \theta_{cen,1}, \theta_{cen,2}, V_{cen,i} \sim N(\theta_{cen,1}\mu_i + \theta_{cen,2}, V_{cen,i}).$$

Modeling assumptions for $\theta_{cen,1}$, $\theta_{cen,2}$, and $V_{cen,i}$ are analogous to those for the previous decennial census. This paper is motivated by an interest in the application of these models, in the relationship of the conditional expectations of the decennial census and ASEC, and in the worthiness of these models for estimation of LNP in counties. It is therefore worthwhile to fit the models in a census year and condition on the current decennial census as well as the previous decennial census to estimate the bias parameters for the current decennial census. It would also be worthwhile to fit it without conditioning on the current decennial census and do evaluations analogous to some of those reported by the National Academy of Sciences (NAS) panel charged with the evaluation of the SAIPE estimates of poverty (NAS, 2000), and to see how the model fares at the end of a decade. One can also use these models, even conditioning on the current decennial census, to say something about the performance of the relative differences.

There are three experiments described in this paper: Experiment 1 has the decennial census variance model 1 and fit to the LNP for school-age children. Experiment 2 uses census variance model 2 on school-age children. Experiment 3 uses census variance model 1 on poverty for all ages.

2.3 Prior Distributions

The unknown parameters are assigned prior distributions as follows.

1. The θ_{ij}, $j > 0$, $i = 1, 2$, parameters are normally distributed.

$$\theta_{ij} \sim N(\mu(\theta_{ij}), v(\theta_{ij})).$$

2. The proportionality constant for the ASEC variance, v_0, has an inverted gamma distribution

$$v_0 \sim I\Gamma(sh(v_0), sc(v_0)),$$

where $I\Gamma$ denotes the inverted gamma distribution and $sh(.)$ and $sc(.)$ represent the shape and scale parameters respectively. In this parameterization, the mean is $\frac{sc(v_0)}{sh(v_0)-1}$ and the variance is $\frac{sc(v_0)}{(sh(v_0)-1)^2(sh(v_0)-2)}$. The precision, $\frac{1}{v_0}$, has a gamma distribution with mean $\frac{sh(v_0)}{sc(v_0)}$ and variance $\frac{sh(v_0)}{sc(v_0)^2}$. The inverted gamma distribution for the variance is equivalent to the gamma distribution for the precision, and is the conjugate prior for this parameter in the normal distribution. This improves the performance of the Monte Carlo procedure for forming the estimates described in Section 3.

3. The variance of μ_i given ν_i also has an inverted gamma distribution.

$$v_\mu \sim I\Gamma(sh(v_\mu), sc(v_\mu)).$$

Here, the variance of the true LNP is approximately the same as the cv^2 of the number in poverty, given the mean parameter. In our model the population is a fixed quantity, so this is approximately equal to the cv^2 of the poverty rate given the overall poverty rate. This, taken as a global parameter, is bounded as a practical matter so that the LNP does not exceed the log population very often.

4. The variance parameters of the X-variables have inverted gamma distributions.

$$v_j \sim I\Gamma(sh(v_j), sc(v_j)).$$

5. The additive constant for the previous decennial census variance has an inverted gamma distribution,

$$v_t \sim I\Gamma(sh(v_t), sc(v_t)).$$

6. The additive constant for the current decennial census variance has an inverted gamma distribution,

$$v_{cen} \sim I\Gamma(sh(v_{cen}), sc(v_{cen})).$$

Table 1. Models and Prior Parameters.

	Experiment 1	Experiment 2	Experiment 3
Decennial Census Variance Model	2	1	2
m_γ, v_γ	-1.9, 0.25	-1.9, 0.25	-1.9, 0.25
$\mu(\theta_{j1})$, $v(\theta_{j1})$	1.0,1.0	1.0,1.0	1.0,1.0
$\mu(\theta_{j2})$, $v(\theta_{j2})$	0.0, 2.0	0.0, 2.0	0.0, 2.0
$sh(v_0)$, $sc(v_0)$	0.0, 0.0	0.0, 0.0	0.0, 0.0
$sh(v_m)$, $sc(v_m)$	1.0, 1.0	1.0, 1.0	1.0, 1.0
$sh(v_1)$, $sc(v_1)$	3.6, 0.716	3.6, 0.716	3.6, 0.716
$sh(v_2)$, $sc(v_2)$	3.6, 0.716	3.6, 0.716	3.6, 0.716
$sh(v_3)$, $sc(v_3)$	3.6, 0.716	3.6, 0.716	3.6, 0.716
$sh(v_t)$, $sc(v_t)$	10.0, 1.0	10.0, 1.0	10.0, 1.0
$sh(v_{cen})$, $sc(v_{cen})$	1000.0, 10.0	10000.0, 1.0	1000.0, 1.0
$\mu(\gamma_g)$, $v(\gamma_g)$	5.0, 1.0×10^{-6}	1.0, 5.0	5.0,1.0×10^{-6}
$\mu(\gamma_c)$, $v(\gamma_c)$	5.0, 1.0×10^{-6}	1.0, 5.0	5.0,1.0×10^{-6}

These are the prior parameters for each of the three models. In this notation, $\mu(\theta)$ refers to the mean of θ, $v(\theta)$ is the variance of θ, $sh(v)$ is the shape parameter and $sc(v)$ the scale parameter for v in an inverted gamma.

The additive constant for the decennial census variance of the LNP in the current decennial census remains in the equation formally, but the prior distribution may constrain it to small values. The parameter v_t can be thought of as a cv^2 for a random multiplier applied to the true current LNP to get the LNP in the year of the previous decennial census. In this case, that is 1980.

7. The parameters γ_1 and γ_2 in variance models 1 and 2 have normal prior distributions.

$$\gamma_j \sim N(m(\gamma_j), v(\gamma_j)).$$

The prior parameters are presented in Table 1. In experiments 1 and 3, the prior distributions for the decennial census variance have very low variance and expectations equal to 5.0. This is the same as treating the base formula for the decennial census variance as true. There was no reason to anticipate this would fit well, but the evidence for rejecting this model is weak, as reported in the model-fitting section. Setting the prior expected values for v_{cen} close to zero reflects an interpretation of that variable as the contribution to the variance of the decennial census which arises by getting out of date; the current decennial census is not out of date at all, so this variance contribution is small.

Note: Parameter Descriptions.

Parameters	Description
m_γ, v_γ	Mean and Variance for National LNP
$\mu(\theta_{j1})$, $v(\theta_{j1})$	Mean and Variance for The Multiplicative Bias Parameter
$\mu(\theta_{j2})$, $v(\theta_{j2})$	Mean and Variance for The Additive Bias Parameter
$sh(v_0)$, $sc(v_0)$	Shape and Scale for the ASEC Variance Parameter
$sh(v_m)$, $sc(v_m)$	Shape and Scale for the Random Effect Variance for LNP
$sh(v_1)$, $sc(v_1)$	Shape and Scale for the Tax LNP Variance Parameter
$sh(v_2)$, $sc(v_2)$	Shape and Scale for the Log FS Variance Parameter
$sh(v_3)$, $sc(v_3)$	Shape and Scale for the Log Nonfiler Variance Parameter
$sh(v_t)$, $sc(v_t)$	Shape and Scale for the Additive Variance Parameter of the Previous Decennial Census
$sh(v_{cen})$, $sc(v_{cen})$	Shape and Scale for the Additive Variance Parameter of the Current Decennial Census
$\mu(\gamma_g)$, $v(\gamma_g)$	Mean and Variance for the Multiplicative Variance Parameter of the Previous Decennial Census
$\mu(\gamma_c)$, $v(\gamma_c)$	Mean and Variance for the Multiplicative Variance Parameter of the Current Decennial Census

2.4 Relationship of this Model to Regression Models

Condition the ASEC on the other data and the parameters besides μ to get the following model. The distribution of the LNP given the data except for the ASEC direct estimate is

$$p(\mu_i|\mathbf{X}_{i(-0)}) \sim N(r_i, v_i),$$

where (-0) indicates that X_{0i} has been left out of the vector. The mean, r_i, is

$$\begin{aligned} r_i &= \sum_j \frac{\theta_{j1}^2 \tau_{ij}}{\sum_k \theta_{k1}^2 \tau_{ik} + \tau_m} \frac{X_i - \theta_{j2}}{\theta_{j1}} \\ &+ \frac{\tau_m}{\sum_k \theta_{k1}^2 \tau_{ik} + \tau_m} (\eta + log(N_i)), \end{aligned}$$

where $\tau_{sub} = 1/v_{sub}$ for all subscripts sub, and

$$v_i = (\sum \theta_{j1}^2 \tau_{ij} + \tau_m)^{-1},$$

so

$$p(X_{0i}|\mathbf{X}_{i(-0)}) \sim N(r_i, v_i + v_0/\sqrt{k_i}).$$

The mean, r_i, can be rewritten

$$\begin{aligned} r_i &= \sum_j \frac{\theta_{j2}\tau_{ij}X_i}{\sum_k \theta_{k1}^2\tau_{ik} + \tau_m} + \frac{\tau_m}{\sum_k \theta_{k2}^2\tau_{ik} + \tau_m}(\eta + log(N_i)) \\ &\quad - \sum_j \frac{\theta_{j1}\tau_{ij}\theta_{j2}}{\sum_k \theta_{k1}^2\tau_{ik} + \tau_m} \\ &= \sum \beta_{j=1}^J X_{ij} + \beta_0. \end{aligned}$$

When $\tau_{ij} = \tau_j$, this is the regression model, except for the common dependence of $\mathbf{V}$ and β on $\mathbf{b}$ and τ. If $\frac{\tau_{ij}}{\tau_m} \to \infty$, those terms act more like regular regression parameters, so multivariate regressions are also limiting cases. Then $\beta_j = \frac{\theta_{j1}}{\sum \theta_{k1}^2}$ for $j > 0$ and $\beta_0 = \frac{\sum_j \theta_{j1}\theta_{j2}}{\sum_k \theta_{k1}^2}$.

3 Estimation

The Metropolis sampler (see, for example, Gelfand 1995) is an appropriate tool to calculate the joint posterior distribution of the model. The distribution of a random variable and that of its parent parameters is usually a conjugate pair in this model, so derivation of the full conditional distributions in these cases is straightforward, and these steps are ordinary Gibbs steps. The parameters in the variance model for the decennial census had more complicated distributions and were simulated with Metropolis steps. The program was written in C using the RANLIB library of random number generating functions (Brown *et al.* 1993). The simulation was run in each experiment for 1,000,000 iterations, thinned to 10,000.

3.1 Model Fit

There is a heavy reliance on posterior predictive p-values (PPP-values) for evaluating model fit. For a discussion of PPP-values, see Gelfand (1995) or Meng (1994). The interpretation of these is similar to classical p-values; the lack of fit of the data to the posterior predictive distribution (the distribution of new data analogous to the observed data) is measured by the probability mass of values more extreme than the observed data. Examples of these p-values in small-area estimation are presented by You *et al.* (2000) and Fisher and Campbell (2002). The PPP-values are defined as

$$p = P(T(\mathbf{X}_{rep}, \xi_{rep}) < T(\mathbf{X}_{obs}, \xi_{rep})|data),$$

where $T(.)$ is some function, chosen to evaluate some interesting aspect of the model. One example is

$$T(x, \xi) = x,$$

so that p represents the probability that a new observation from the posterior is less than the observed value. One can also define the overall PPP-value, which is the probability that a randomly chosen county has $T(\mathbf{X}_{rep}, \xi_{rep}) < T(\mathbf{X}_{obs}, \xi_{rep})$. In each case, values of p close to zero or one suggest some problem with the model in terms of the location of the response. A well-fitting model should exhibit the following properties.

- The posterior predictive probabilities should not be close to zero or one.
- The posterior predictive probablities should not exhibit a trend relative to other variables. In this respect, the posterior predictive p-values may be examined in a way similar to residual plots.

With the above definition for T, extreme predictive p-values indicate that a replicated observation is too high or too low most of the time. That is, the observed datum is far in the tails of the marginal distribution of the modeled distribution which should then be viewed with distrust. In all three experiments, the overall PPP-values fall in the interval (0.47,0.49), which is comfortably far from zero or one.

Some of the plots of the PPP-values against the logs of the posterior mean numbers in poverty are presented. They are similar for all three experiments so only those for Experiment 1 are shown.

The plot in Figure 1 for the ASEC PPP-value shows little dependence on the LNP, except for some curvature at the lower end which may be illusory. On the other hand, it may be an effect of the censoring of the no-poverty-in-sample counties. The counties with PPP-values equal to zero are counties with no people in poverty in sample.

The plot for log poverty exemptions in Figure 2 shows a clear relationship between the spread of the PPP-values and the log posterior mean poverty. For larger posterior mean LNP, the PPP-values for the log child poverty exemptions spread, tending toward the extreme for the largest counties. The same trend is visible, but less obviously, in the plot for nonfilers, X_{3i}. This suggests that the model for the variability does not fit well in some respect. This is consistent with the PPP-value plots, presented in Figure 5, for the variances of the tax variables, which follow. There seems to be a more complicated situation in the plot for FS, seen in Figure 3; for smaller counties, the PPP-values tend toward the extremes, while they tend toward zero for larger counties, suggesting that food stamp usage is underestimated for large counties. This should be examined further, since a systematic underprediction here may indicate a conditionally biased part of the model or perhaps a curvilinear relationship between log-food stamps and log-poverty.

The decennial census plots in Figure 4 are striking mostly for their difference. Even though the models share the same form, the plot for the previous decennial census shows no visible relationships between the PPP-value and the log posterior mean number in poverty at all, while the variability in the current decennial census plot clearly decreases with increasing log posterior mean in poverty.

Another possibility for $T(x, \xi)$ is

$$T(x, \xi) = (x - E(x|\xi))^2,$$

so overall values of p close to zero or one indicate a problem in the variance model. In this case, all of the overall PPP-values are in the interval (0.47,0.51), which is not itself indicative of a failure in the model with respect to variances.

Consider the plot of the PPP-values for the sampling variance of the log number of poverty exemptions in Figure 5. The plots in all of the experiments are similar, showing that the variances of replicated observations tend to be too big for places with more people in poverty, suggesting that the tax variance function $f(.)$ should increase more quickly with the size of the county. This will be addressed in future research.

Figure 6 shows plots of the PPP-values for the decennial census variances; the plots for the other two experiments are similar to these, and so are not shown, but the plots for the two decennial censuses are different from each other. The plot for the previous decennial census shows PPP-values fairly uniformly distributed over the interval, without a trend, except for a compression of values at the low end. The 1990 decennial census PPP-values, on the other hand, while generally centered around 0.51, show much less variation in the large counties. Perhaps the normal model is failing with respect to higher moments, or the diffusion effect, which is negligible in the 1990 decennial census, has a more normal distribution.

3.1.1 Relative and Absolute Relative differences from the Decennial Census

Historically, in the SAIPE project, the mean relative difference across counties between the estimates from the model and the decennial census estimates in 1990 has been an important statistic for judging quality of a model. The definition of the relative difference between x_1 and x_2 is

$$RD(x_1, x_2) = \frac{x_1 - x_2}{x_2}.$$

The absolute relative difference, $ARD(x_1, x_2)$, is the absolute value of $RD(x_1, x_2)$. The mean of these quantities, applied to some estimates $\hat{\mu}_i$ and decennial census estimates $X_{cen,i}$ are $MRD = \frac{1}{M}\sum_i RD(\hat{\mu}_i, X_{cen,i})$ and $MARD = \frac{1}{M}\sum_i ARD(\hat{\mu}_i, X_{cen,i})$.

Models with MRDs and MARDs close to zero were preferred, as were those where the MRD varied little with respect to other variables. MRDs with large negative or positive values were considered indicators of relative bias; large MARDs were considered indicators of high relative error. Further, if the MRD was high for a subset of counties distinguished by some other variable, the suspicion was that the estimates were biased for those counties. The interpretation of the MRD and MARD as measures of lack of accuracy and precision depends on the assumption that the decennial census estimates in poverty are both accurate and precise. That assumption is questionable for precision of small counties and for the accuracy of all counties, since the θ-parameters for the decennial census are unlikely to be zero and one, especially given the estimates. The models fit here give us an opportunity to measure certain kinds of failure in that assumption. This model has the assumption that the decennial census LNP in 1990 is affine in the true LNP.

It is also useful to examine the expectation of the RD under the model. Consider the following transformation of the RD between decennial census and true LNP for county i.

$$\begin{aligned} q &= log(RD(\mu_i, X_{cen,i}) + 1) \\ &= \mu_i - (\theta_{cen,1}\mu_i + \theta_{cen,2} + Z\sqrt{v_{cen,i}}) \\ &= (1 - \theta_{cen,1})\mu_i - \theta_{cen,2} - Z\sqrt{v_{cen,i}}, \end{aligned}$$

where Z is a random variable with mean zero and variance one. In these models there is also the assumption that it is normally distributed, but that assumption is not necessary for the following. It follows that

$$\exp(q) = RD(\mu_i, X_{cen,i}) + 1.$$

The first terms of the Taylor expansion give

$$\exp(q) = \exp(q_0) + (q - q_0)\exp(q_0) + \frac{1}{2}(q - q_0)^2 \exp(q_0) + o_p(q - q_0)^2.$$

The expectation conditioned on some parameters ξ is

$$\begin{aligned} E(RD(\mu_i, X_{cen,i}) + 1|\xi) = \exp(q_0) + E(q - q_0|\xi)\exp(q_0) \\ + \frac{1}{2}V(q|\xi)\exp(q_0) + o(V(q|\xi)), \end{aligned}$$

where, by assumption, $q_0 = E(q|\xi)$. Now say ξ is μ and θ, the relevant parameters. Then

$$E(RD(\mu_i, X_{cen,i}) + 1|\mu, \theta) \quad = \quad (1 + \frac{1}{2}V(z|\mu, \theta))\exp(E(z|\mu, \theta)) + o(V(Z\sqrt{v_{cen,i}}))$$

so

$$E(RD(\mu_i, X_{cen,i}) + 1|\mu, \theta) = (1 + \frac{1}{2}v_{cen,i})\exp((1 - \theta_{cen,1})\mu_i - \theta_{cen,2}) + o(v_{cen,i}). \quad (1)$$

If $\theta_{cen,1} = 1.0$ and $\theta_{cen,2} = 0.0$, then

$$E(RD(\mu_i, X_{cen,i})) \quad = \quad \frac{1}{2}v_{cen,i} + o(v_{cen,i}).$$

Neglecting the error term, the expected RD is one-half the sampling variance of the decennial census LNP. This number is approximately one-half the relative variance of the decennial census direct estimate of poverty, which can be high for counties with little sample and decreases for counties with more. The real point is that even when the decennial census is unbiased for the ASEC LNP, the RD is nonzero and is not flat.

Table 2 shows the MRD of the posterior means, $\bar{\mu}$ of the county LNP and the decennial census Estimates LNP, the posterior mean of the MRD of the LNP and the decennial census LNP, and the expectation of the MRD using equation (1) above for each of the decennial census variance models.

Table 2. Mean Relative Differences (MRD) From the Decennial Census.

	MRD($\bar{\mu}$,$\mathbf{X}_{cen}$)	$MRD(\mu, \mathbf{X_{cen}})$	$E(MRD)$
Experiment 1	-0.117	-0.117	-0.126
Experiment 2	-0.107	-0.107	-0.114
Experiment 3	-0.174	-0.175	-0.177

The MRD estimates in all three experiments are somewhat less than zero, especially for Experiment 3. The consistency with their theoretical expectation, $E(MRD)$, in the third column, suggests the MRD values are explained by the model and the posterior distributions for θ, at least in the presence of the available data.

The RD can be used to construct posterior predictive p-values as well; consider the function

$$T(x, \xi) = \frac{\exp(\mu_i) - Y_{d,i}}{Y_{d,i}}.$$

Table 3. Overall Relative Difference Predictive p-values.

Experiment 1	0.51
Experiment 2	0.50
Experiment 3	0.51

The nicely moderate PPP-Values for the MRD in Table 3 show that the model explains the overall relative differences between the posterior mean number in poverty and the decennial census number in poverty. Figure 7 shows the PPP-values plotted against the log-posterior mean numbers of people in poverty. There is no clear trend for the level of PPP-values though there is, again, heterogeneous variability.

The conclusions of this section are (1) that the measurement differences between ASEC and decennial census are unknown but likely not zero and (2) given that and the effects of the variance, the RD and the ARD can not be expected to be flat or centered on zero. Under the assumptions of this model (including that of ignorability of the censoring of no-poverty-in-sample counties) the RD is closer to zero than its approximated expectation and behaves much as the model predicts, as measured by PPP-values.

3.2 Estimates

The posterior means and standard deviations of the parameters are presented in Table 4. Experiment 3, where the variables are all related to total poverty, is most easily interpreted. The multiplicative bias parameters for the log tax data have posterior means around 0.85, which is several standard deviations away from 1.0. This indicates a downward concave function of tax poverty relative to the true ASEC poverty. In counties with relatively large populations of people in poverty, fewer of those people in poverty appear as such in the taxes, and, in those counties, the rate of nonfilers also decreases. While this may reflect a difference in the populations of counties with various levels of poverty, one must also recognize the possibility of an effect of censoring the counties without poverty in sample. Since the probability of having no poverty in sample and, therefore, having an undefined ASEC direct estimate of LNP, decreases with increasing poverty rates, the model must be misspecified; the magnitude and nature of the effect is still not clear, though Maiti and Slud (2002) study it with simulation experiments. It is possible to explicitly model the censoring mechanism in an HB model, and that research is underway.

To examine the conditional expectations of the x_{ij}, $j > 0$, on the linear scale, consider the quantity

$$r_{ij} = \exp((\hat{\theta}_{0,j} - 1)\hat{\mu}_i + \hat{\theta}_{1,j}),$$

where the hat denotes the posterior mean and i indexes counties. This is the ratio of the posterior means of the expectations of the covariates to those of the ASEC, converted to the linear scale; if the AR and decennial census measure poverty, these are the ratios of those poverty measure to that of the ASEC. Table 5 shows the averages of the ratios across counties.

The results for Experiments 1 and 2 are similar. The numbers of exemptions for people in poverty are about 80 percent larger than the ASEC direct estimates. One would expect there to be more child poverty exemptions, since the universe for this classification is nominally children under 21 years of age. Recall that FS participation counts all ages, so there are 2.45 times as many estimated total FS participants as school-aged children in poverty. Similarly, there are perhaps five times as many nonfilers as there are school aged children in poverty. Since the universe for nonfiler is *all people* rather than school-aged children, it is not very surprising that there are so many more. Finally, the decennial censuses, which do measure similarly (but not identically) defined age groups, show about 13 to 15 percent more poverty than the ASEC, consistently between the decennial censuses. In experiment 3, the poverty-related measurements are for all ages, so they are more comparable than those for the school-aged children. Also note the numbers of poverty exemptions and non-filers are 29 percent and 22 percent larger than the ASEC people in poverty, respectively. The bias parameters for the tax poverty and the decennial census poverty are fairly similar.

Table 4. Posterior Means and Standard Deviations of The Parameters.

	Description	Exp 1	Exp 2	Exp 3
θ_{11}	Multiplicative Bias Parameter, Log Tax Poverty.	0.85, 0.008	0.86, 0.006	0.85, 0.007
θ_{21}	Multiplicative Bias Parameter, Log FS[a]	1.01, 0.007	1.02, 0.008	1.02, .009
θ_{31}	Multiplicative Bias Parameter, Log Nonfilers	0.96, 0.01	0.97, 0.009	0.99, 0.009
θ_{41}	Multiplicative Bias Parameter, 80 Decennial Census LNP[b]	0.87, 0.007	0.89 0.007	0.85, 0.007
θ_{12}	Add Bias Parameter, Log Tax Poverty	1.55,0.06	1.50, 0.05	1.43, 0.07
θ_{22}	Add Bias Parameter, Log FS	0.82, 0.07	0.79, 0.06	-0.63, 0.84
θ_{32}	Add Bias Parameter, Log Nonfilers	1.88, 0.08	1.81, 0.07	0.31, 0.09
θ_{42}	Add Bias Parameter, 80 Decennial Census LNP	1.00, 0.06	0.86, 0.06	1.34, 0.07
$\theta_{cen,1}$	Multiplicative Bias Parameter, 90 Decennial Census LNP	0.92, 0.007	0.93, 0.006	0.91, 0.007
$\theta_{cen,2}$	Add Bias Parameter, 90 Decennial Census LNP	0.68, 0.06	0.60, 0.05	0.95, 0.07
v_0	ASEC Variance Parameter.	5.70, 0.26	5.57, 0.25	3.07, 0.13
v_m	Variance of $\mu_i\|\eta$	0.30, 0.009	0.28, 0.008	0.23, 0.007
v_1	Tax Poverty Variance Parameter.	4.6, 0.13	4.3, 0.13	9.21, 0.25
v_2	FS Variance Parameter.	0.12, 0.004	0.11, 0.004	0.15, 0.004
v_3	Nonfiler Variance Parameter.	33.2, 0.86	31.8, 0.82	50.91, 1.32
v_t	80 Decennial Census LNP Variance Parameter.	0.08, 0.010	0.06, 0.003	0.04, 0.001
v_{cen}	90 Decennial Census LNP Variance Parameter ($\times 10^{-3}$)	1.0,0.03	0.1, 0.01	1.0,0.04
$\gamma_{1,80}$	80 Decennial Census Variance Parameter	NA	1.83, 0.25	NA
$\gamma_{1,90}$	90 Decennial Census Variance Parameter	NA	2.19, 0.13	NA

[a] Number of Food Stamp Recipients
[b] Log Number in Poverty

Table 5. Ratios of Covariate measures of Poverty to that of the ASEC.

j	Measure	Exp. 1	Exp. 2	Exp. 3
1	Poverty Exemptions	1.81	1.79	1.30
2	Food Stamps	2.48	2.48	0.60
3	Nonfilers	5.18	5.11	1.23
4	Previous Decennial Census	1.15	1.13	1.16
cen	Current Census	1.16	1.14	1.24

Examination of the θ-parameters for the decennial census LNPs show that the ratio of the expectation of the decennial census measurement of LNP to that of the ASEC increases with the LNP and averages about 1.23. This positive slope is consistent with the plot of the relative difference from the decennial census shown versus size in Figure 8. This and the moderate PPP-values and the plots of the PPP-values for the relative differences lead to the conclusion that the relative difference is about what it should be under the model and the decennial census' value as a "gold standard", against which other estimates may be compared, is limited, at least using the RD directly.

One important flaw in this argument is that the effect of the censoring of counties with no people in poverty in sample is not well understood. Depending on the censoring mechanism, this may have an important effect on these comparisons. Indeed, if, in counties with smaller samples, counties with fewer people in poverty were more likely to have zero people in poverty and, therefore, to be dropped, then we might expect the model to show a lower decennial census-to-ASEC ratio for small counties and a higher one for large-sample counties. This is in fact contrary to these results.

Turning our attention to the variance parameters, the FS variance parameter is about 0.11 in the experiments with school aged children and about 0.15 in that with all people in poverty. Since this is the log scale, this is approximately the cv^2 of the FS, or a cv of about 0.33 and 0.38, respectively. There may be some error in the measure of FS given the true number of FS participants, but it seems reasonable to suspect there is a comparatively large random effect in the true participation given poverty. Modeling of the relationship of FS participation to other variables may yield some benefits.

The posterior standard deviations of the number of people in poverty in the three experiments, divided by the posterior means, are shown in Table 6.

The only differences between Experiment 1 and 2 are in the model for the decennial census variances, namely the prior distribution for the constant term, the form of the GVF-type term, and the prior distribution for its parameter. Since the prior mean for the constant term is the smaller of the two in Experiment 2, one has to conclude that the formulation for the GVF-like term in Experiment 1 vs. that in Experiment 2 has the larger effect. The posterior means for the multiplier for the GVF in Experiment 2 are larger than one, so the estimated

Table 6. Average Ratio of the Posterior Standard Deviation to the Posterior Mean

Experiment 1	0.09
Experiment 2	0.12
Experiment 3	0.06

information in the second model is lower. These ratios compare favorably to the *cvs* in the SAIPE mixed effects model, which are 0.13 and 0.12 for school-aged related children in poverty and total people in poverty, respectively.

4 Conclusions

4.1 More Research

Research on these models will proceed in the following ways.

1. Extend the ASEC model into a mixture, where one component is as modeled as in this paper, while the other concentrates all of its probability at zero on the linear scale, where the probability of that component is some function of the poverty rate and the sample size. Alternately, one can employ a probabilistic imputation of the counties with no people in poverty in the ASEC sample, conditioning on the event they fall below a cut-off. This is convenient in a Markov Chain Monte Carlo method.

2. In an approach similar to that in the first item, the approximate scaled binomial model (Fisher and Asher, 1999) could be used in the ASEC sampling model. Their use of this method seemed to yield a good fit and estimates with good variances, but the properties of the method should be more fully established before it is used in production.

3. The choice of prior distributions here is a little strange, especially for the FS variance, though these priors probably contribute little information. Better priors could be formulated on the basis of other knowledge and the sensitivity to those prior distributions can be tested.

4. Include other variables of interest. A current example is health insurance coverage (HI). In that case there are several measurements of insurance status from AR and surveys. Given that there is a relationship between HI and income, information about each variable could inform us about the other.

5. The FS participation rates could be modeled as a function of other covariates. As mentioned in section 3.2, there is a large error associated with FS conditioned on LNP. It may be profitable to model this effect, which could allow us to get more information from FS.

6. This model has the assumption of independence between the two tax variables. It would be straightforward to test this assumption with a PPP-value constructed around a correlation or similar measure of dependence. Then a modification of the model might be appropriate.

5 References

Brown, Barry W., Lovato, J., Russel, K.,(1993), RANLIB random number generation library, available at http://www.netlib.org/random/ranlib.c.tar.gz

Fisher, Robin and Campbell, Jennifer (2002) Health Insurance Estimates for States, *2002 American Statistical Association Proceedings of the Section on Government and Social Statistics*

Gelfand, Alan E.,(1995), Model Determination using Sampling-Based Methods. In *Markov Chain Monte Carlo In Practice* (eds W.R. Gilks, S. Richardson, and D. J. Spiegelhalter),pp.145-161

Maiti, T. and Slud, E.V. (2002). Comparison of Small Area Models in SAIPE. Tech Rep. U.S. Census Bureau.

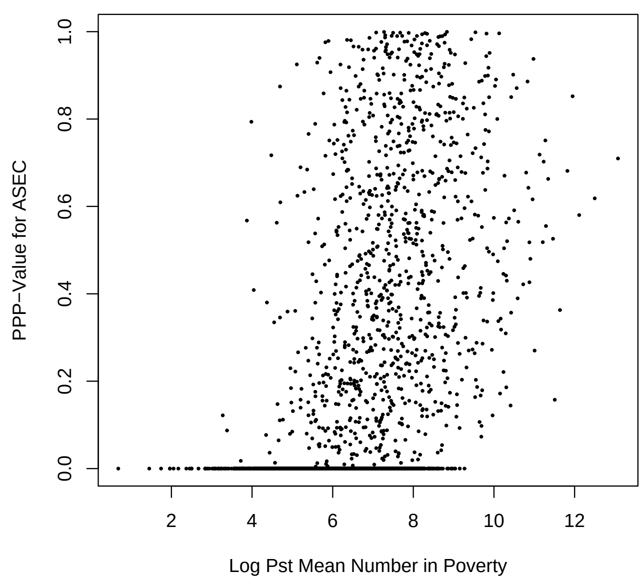
Meng, Xiao-Li, (1994), Posterior Predictive p-Values, *The Annals of Statistics*, 22, pp. 1142-1160

U.S. Census Bureau (2003), "Small Area Income and Poverty Estimates, Intercensal Estimates for States, Counties and School Districts", Available at http://www.census.gov/hhes/www/saipe.html

U.S. Census Bureau (1992), "1990 Summary Tape File 3 Technical Documentation on CD-ROM", available at http://factfinder.census.gov/metadoc/1990stf3td.pdf

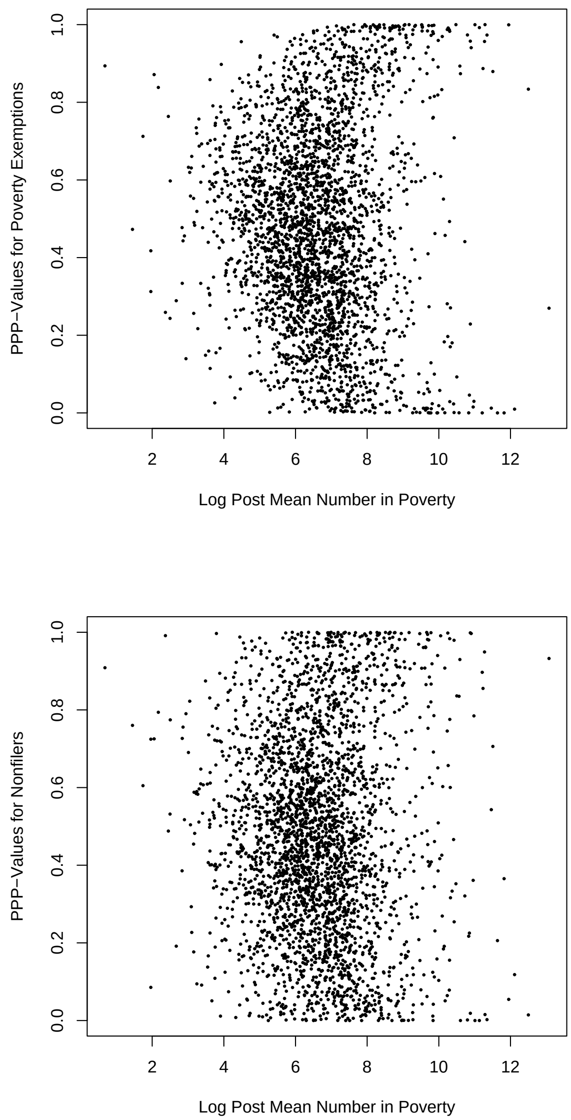
You, Yong, Rao, J.N.K., and Gambino, Jack (2000), "Hierarchical Bayes Estimation of Unemployment Rate for Sub-provincial Regions Using Cross-sectional and Time-series Data", *2000 American statistical Association Proceedings of the Section on Government and Social Statistics*

Figure 1. PPP-values for the Annual Social and Economic Supplement (ASEC) in Experiment 1.



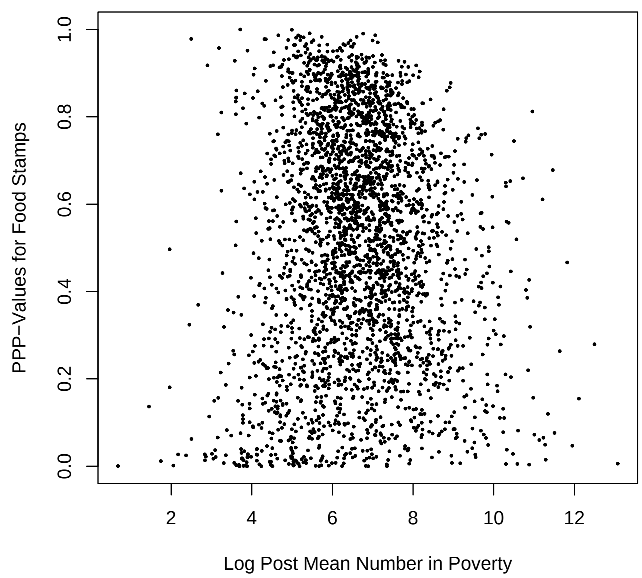
There is little evidence for a failure of the model except possibly at the lower end. The counties with PPP-values equal to zero have no poverty in sample.

Figure 2. PPP-values for Tax Variables in Experiment 1.



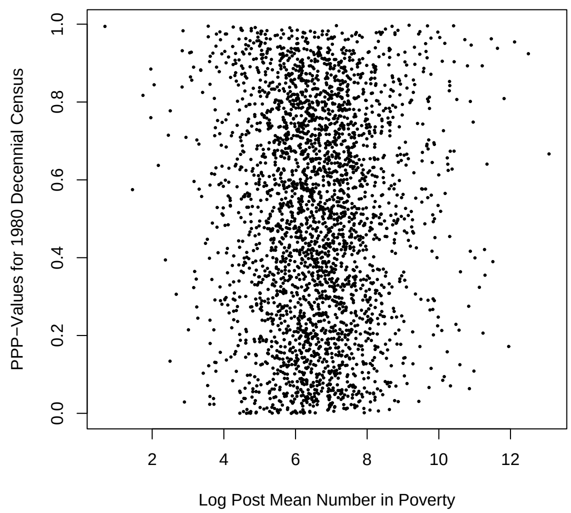
The PPP-values do not seem to indicate a failure of the model with respect to the expectation of the tax variables.

Figure 3. PPP-values for Food Stamps in Experiment 1.



There is some relationship between the PPP-value and log posterior mean poverty, but it is not an increasing or decreasing one.

Figure 4. PPP-values for the Decennial Censuses in Experiment 1.



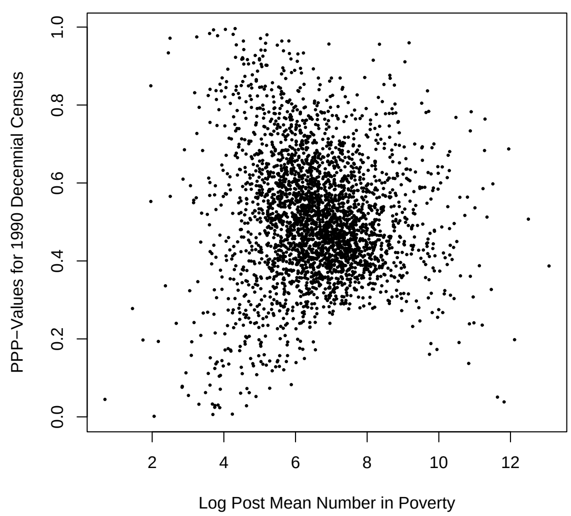
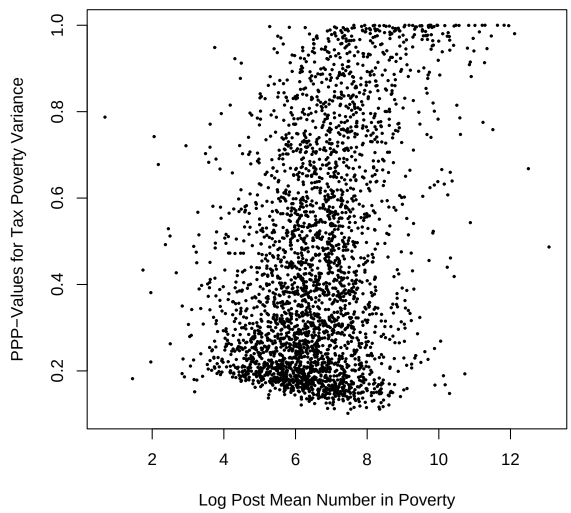


Figure 5. PPP-values for the Tax variances in Experiment 1.



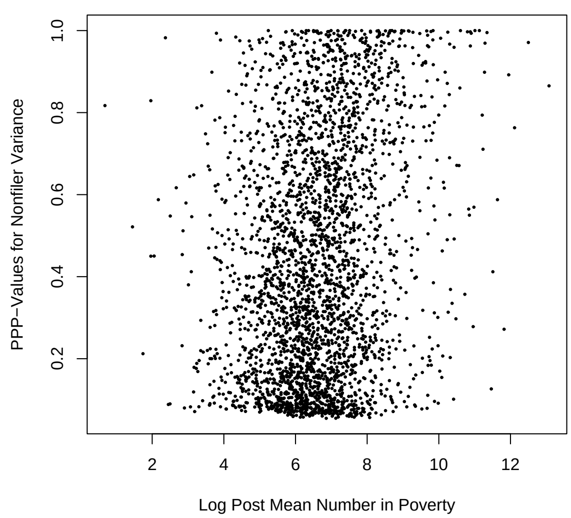
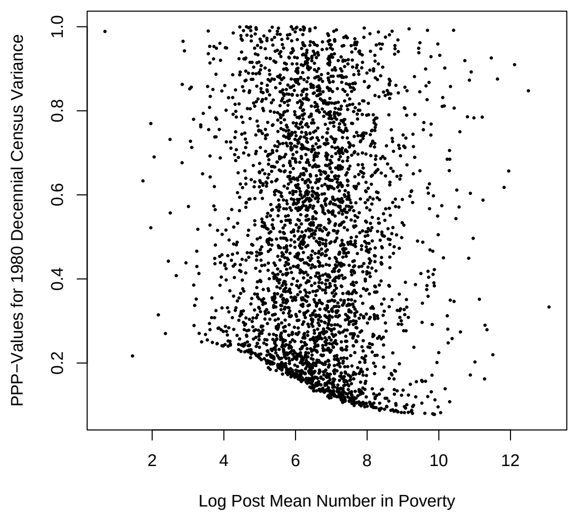


Figure 6. PPP-values for the Variances of the Decennial Censuses in Experiment 1.



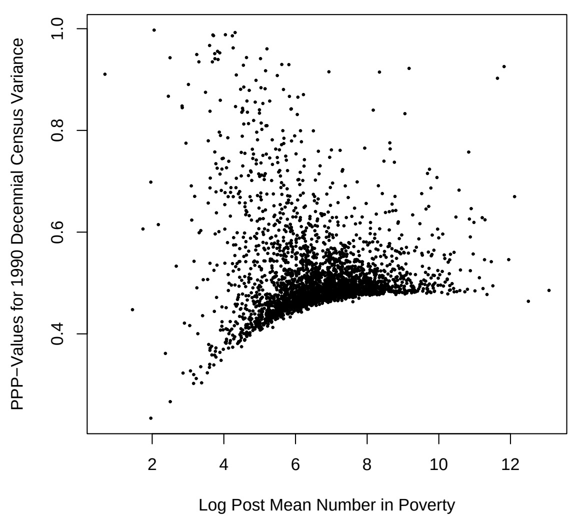


Figure 7 PPP-values for the Relative Difference with the Decennial Census in Experiment 1.

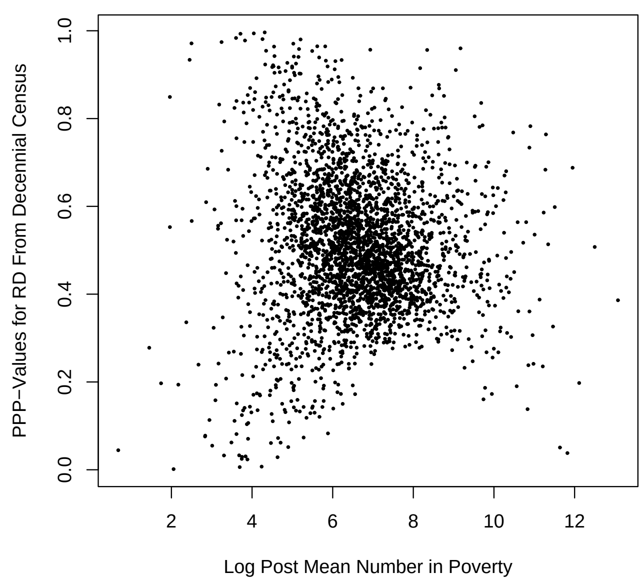


Figure 8. Relative Differences Between Posterior Mean poverty and Decennial Census in Experiment 3.