Small Area Estimation of Health Insurance Coverage From the Current Population Survey's Annual Social and Economic Supplement and the Survey of Income and Program Participation[1]

Robin Fisher and Joanna Turner
U.S. Census Bureau, HHES-SAEB, FB 3 Room 1451
Washington, DC, 20233 Phone: (301) 763-3193, email: robin.c.fisher@census.gov

Key Words: Small Area Estimates; Health Insurance Coverage; ASEC; SIPP

1. Introduction

The U.S. Census Bureau's Small Area Health Insurance Estimates (SAHIE) project of the Small Area Estimates Branch (SAEB) is developing model-based estimates of the number of people not covered by health insurance (i.e. who are uninsured) for states and counties. There is broad public interest in health insurance coverage issues. Generally, health insurance coverage statistics are available only through national household surveys, and the estimates from these surveys vary widely for a number of well-documented reasons (Lewis *et al.*, 1998). The Current Population Survey's (CPS) Annual Social and Economic Supplement (ASEC) and the Survey of Income and Program Participation (SIPP) are two commonly cited sources for health insurance statistics. Sample sizes are not large enough, however, that the surveys alone can produce sufficiently reliable state or sub-state estimates for many policy purposes.

Recent methodological developments, at both the U.S. Census Bureau and in the broader research community, offer new potential for developing estimates of various uninsured populations in small areas. The SAEB has played a significant role in this field, developing a program that produces income and poverty estimates at the state, county, and school district levels. The Small Area Income and Poverty Estimates (SAIPE) program constructs statistical models that relate income and poverty to various indicators based on administrative records, population estimates, survey, and decennial census data. These are then combined with direct estimates from the ASEC to provide estimates and standard errors for the geographic areas of interest.

This research is part of an ongoing effort to expand SAIPE knowledge and methodologies to the area of health insurance coverage. This paper extends the work of Fisher and Turner (2003) to model the ASEC and the SIPP jointly. The result is a bivariate estimate of health insurance coverage for each county. The first element of this estimate reflects the ASEC estimate of health insurance coverage while the second reflects that of the SIPP. This has a few advantages. First, there is information from each survey about each of the two estimates. Second, the user may have a preference about which of the two surveys is more appropriate for their purposes. Third, it allows for an analysis of the difference between the two surveys' measurements of health insurance coverage. For discussions about the differences between the two measures of health insurance coverage, see Bennefield (1996) and Bhandari (2004). The paper proceeds as follows: Section 2 describes our data sources; Section 3 describes the model; Section 4 discusses the model fit and estimation; preliminary results are provided in Section 5; we conclude and describe future plans in Section 6.

2. Data

We describe the variables we considered, although not all were included in the model, for various reasons discussed below.

a. CPS ASEC

The ASEC is conducted annually, the data are released in a timely manner, and it has a state-based design. The sample size is approximately 50,000 housing units in 1999; since then it has increased to about 100,000 housing units.

- **Log proportion insured.** This is the log of the ratio of the total insured to the civilian non-institutional population, measured by the ASEC; this is a three-year average of the three ASEC direct

[1] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on technical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

estimates, centered on the year of interest (for this paper, 1999), weighted by the number of households in sample. The ASEC sample is reweighted so each county's direct estimate is approximately unbiased for the number of insured for that county. This is denoted $INSHR_{i,asec}$ for county i. Note that, for every county with sample, there are insured people. Thus when the log proportion insured is calculated, there are no counties for which the response is undefined. There are 1,198 counties with sample in at least one of the three years, out of the total of 3,141 counties, with an average sample size of 123 households.

b. SIPP

The SIPP is a longitudinal survey, conducted monthly, and is designed for national, rather than state estimates. The sample size is smaller than CPS ASEC and varies from approximately 28,000 to 35,000 housing units depending on the year.

- **Log proportion insured.** This is the log of the ratio of the total insured to the civilian non-institutional population, measured by the SIPP; this is a three-year average of the three SIPP direct estimates, centered on the year of interest (1999). The estimates for 1998 and 1999 are from the 1996 panel. Due to funding constraints, the 2000 panel only covers ⅔ of the year, so for the purposes of this research we used the 2001 panel for our 2000 estimates. The SIPP sample is reweighted so each county's direct estimate is approximately unbiased for the number of insured for that county (Rottach, 2004). This is denoted $INSHR_{i,sipp}$ for county i. There are 736 counties with sample in at least one of the three years, with an average sample size of 123 households. Note that there are only four counties with SIPP sample but no insured people; the contribution of these counties was neglected since their estimate is undefined on the log scale.

c. Medicaid

States submit their eligibility and claims data quarterly to the Centers for Medicare and Medicaid Services (CMS) through the Medicaid Statistical Information System (MSIS). These data also contain the number of State Children's Health Insurance Program (SCHIP) recipients. However, not all states report data for SCHIP recipients. 1999 was the first year for reporting these data and states were not required to follow the reporting requirements as strictly as in later years (Centers for Medicare and Medicaid Services, 2003).

- **Log proportion eligible for Medicaid by various age and race/ethnicity categories.** Groups of particular interest are children (denoted $medicaid0\_17_i$), adults ages 35 to 64 (denoted $medicaid35\_64_i$), Hispanics (denoted $medicaidhisp_i$), and Blacks (denoted $medicaidblk_i$). We added 1.0 to the number of eligibles before forming the log proportion for children, Hispanics, and Blacks. An individual is considered eligible if they were covered by Medicaid for at least one day during the quarter. We counted an individual as eligible if they received full benefits or received benefits through a SCHIP expansion program. Due to data quality issues we used the second quarter of calendar year 1999 data; further research about how to use the Medicaid data is underway.

d. Internal Revenue Service (IRS)

This is information from tax returns aggregated to state and county levels. The total number of exemptions attributed to a return includes the filer, the spouse of the filer, and the number of child exemptions for the household.[2] Income year 2000 tax data is used, since this is the first year these variables were available to us.

- **Log IRS proportion between multiples of the Federal Poverty Threshold (FPT).** This is the log of the fraction of exemptions on tax returns living in households with money income between two proportions of the FPT, say p_1 and p_2. This is denoted $tax_i(p_1, p_2)$ for county i. Available values for p_k are 0%, 50%, 100%, 130%, 200%, 300%, and infinity. An alternative summary to the proportions between multiples of the FPT follows.

[2] For more details see http://www.census.gov/hhes/www/saipe/techdoc/inputs/taxdata.html.

- **IRS moment of the log ratios of individuals' family income to their Federal Poverty Threshold (FPT).** This is

$$LIPR_i(r) = \sum_{j}^{N_j} \left[\ln \left(\frac{inc_{ij}}{FPT_{ij}} \right) \right]^r ,$$

r =1, 2, 3, 4. FPT_{ij} is the Federal Poverty Threshold for the family of person *j* in county *i*. Family money income of that person is inc_{ij}. These moments contain information about the shape of the income distribution. There is evidence of a relationship between income relative to the FPT and insurance coverage at the state level (Fisher and Campbell, 2002).

e. Census 2000

Several variables tabulated from Census 2000 were considered as predictors, in particular the log of the total population (denoted $totpop_i$), log proportions in several age categories, log proportion Hispanic (denoted $hisp_i$), and log proportions in various race categories.

f. Food Stamp Program

The food stamp program is a low-income assistance program that is uniform in eligibility requirements and benefit levels across states, with the exception of Alaska and Hawaii, where benefit levels are higher. The U.S. Census Bureau receives counts of the number of people participating in the food stamp program from the United States Department of Agriculture, Food and Nutrition Service.[3]

- **Log number of recipients.** By county, this is the log of the number of individuals participating in the food stamp program in the month of July in 1999, denoted as $foodstamp_i$.

g. USDA Urban Influence Codes (UIC)

The Economic Research Service (ERS) of the United States Department of Agriculture (USDA) urban influence codes for 2003 divide the counties in the United States into twelve groups based on metro/non-metro status, adjacency to metro area, and largest city size.[4] This variable was considered on the hypothesis that there are effects of "urbanness" beyond those described by the other variables. There may be differences, for example, in access to programs or culture that may effect participation in health insurance programs. The variables are denoted as UIC_{im}, where m = 1 to 12; the groups of particular interest are defined as follows:

- UIC_{i1}: In large metropolitan area of 1+ million residents
- UIC_{i2}: In small metropolitan area of less than 1 million residents
- UIC_{i3}: Micropolitan area adjacent to large metropolitan area
- UIC_{i5}: Micropolitan area adjacent to small metropolitan area
- UIC_{i8}: Micropolitan area not adjacent to a metropolitan area.

h. Innovation Codes

The Urban Institute's "Assessing the New Federalism" project (Holahan and Pohl, 2002) divides the 50 states (excluding the District of Columbia) into four categories based on their level of innovation in providing health insurance coverage to low-income Americans. We have collapsed their categories into two groups: states that have gone beyond the minimum requirements of the law and those that have not.

3. Model

This model generalizes that of Fisher and Turner (2003). The model for the log insured rate for county *i* is $INSHR_i^T = X_i \beta + u_i + \varepsilon_i$, where $INSHR_i^T$ is the pair of direct estimates from the surveys, X_i is the vector, of length *d*, of covariates for that county, β is a $d \times 2$ matrix where column 1 is the vector of regression coefficients for the ASEC, and column 2 is for the SIPP. The random effects term, u_i, and the sampling error term, ε_i, have bivariate normal distributions $N_2(0, V_u^{1/2} C_\varepsilon V_u^{T/2})$ and $N_2(0, V_{\varepsilon i}^{1/2} C_\varepsilon V_{\varepsilon i}^{T/2})$, respectively. Here, $V_{\varepsilon i}$ is the diagonal matrix with $[V_{\varepsilon i}]_{11} = \frac{v_{\varepsilon 1}}{k_i^\gamma}$ and $[V_{\varepsilon i}]_{22} = v_{\varepsilon 2} \tilde{v}_i$, where $\tilde{v}_i$ is an independent

[3] For more details see http://www.census.gov/hhes/www/saipe/techdoc/inputs/foodstmp.html.

[4] For more details see http://www.ers.usda.gov/briefing/Rurality/UrbanInf.

estimate of the variance of the SIPP direct estimate (Rottach, 2004) and V_u is the diagonal matrix where the diagonal elements may differ. k_i is the ASEC sample size. (We will see that this assumption fits imperfectly in a following section.) The correlation matrices C_u and C_ε are the general correlation matrix, where the off-diagonal element is ρ, and the identity, respectively; the latter follows from the assumption that the ASEC and the SIPP are independent.[5] For brevity we will denote the parameters $(\beta, v_{u1}, v_{u2}, \rho, v_{\varepsilon 1}, v_{\varepsilon 2})$ as θ and $X_i\beta + u_i$ as μ_i. The parameter μ_i, then, has two components, $\mu_{i,ASEC}$ and $\mu_{i,SIPP}$. The underlying discreteness of the ASEC sample, which may be important when the proportion of interest is close to zero or one and the sample size is small, makes the normality assumption for the sampling error particularly suspect.

Model 1 is defined by the above considerations and we have specified it as

$$E(INSHR_{i,s} \mid \theta) = \beta_{0s} + \beta_{1s} medicaid0\_17_i$$
$$+ \beta_{2s} medicaid35\_64_i + \beta_{3s} medicaidhisp_i$$
$$+ \beta_{4s} medicaidblk_i + \beta_{5s} tax_i(100{,}130)$$
$$+ \beta_{6s} tax_i(200{,}300) + \beta_{7s} totpop_i + \beta_{8s} hisp_i$$
$$+ \beta_{9s} foodstamp_i.$$

Here, s indexes the survey: $s \in \{ASEC, SIPP\}$. If $\rho = 0$ and $\gamma = 0.5$, the portion of the model with $s = ASEC$ is the same as the model of Fisher and Turner (2003). The last two terms were inadvertently omitted in that paper, but not in their model. This choice of γ was taken from similar models for poverty (Fisher and Asher, 2000 and Asher and Fisher, 2000).

Two other sets of variables improved the model fitting statistics sufficiently over the previous model (Fisher and Turner, 2003): the USDA urban influence codes and the Urban Institute innovation codes. There are currently no plans for the regular production of the innovation codes. Since state legislation may change quickly and the effect of the innovation codes is small, we disqualified them from consideration for a production model, though insight might be gained by considering them in research versions of the model. The urban influence codes are produced as a function of the decennial census and describe characteristics which might be expected to remain informative throughout the decade. They were introduced into the model as indicators for the categorical variables; no attempt was made to order the codes. Using weighted least squares regression techniques, where the weights are derived from the variances in model 1, a subset of the urban influence codes was chosen on the basis of the AIC and C_P, namely {1,2,3,5,8}. This set seems significant using F-statistics based on the above regression method, but failed to be when the results of the hierarchical Bayesian analysis were used .

The hierarchical Bayesian model parameters were also estimated. The conditional expectation for model 2 follows.

$$E(INSHR_{i,s} \mid \theta) = \beta_{0s} + \beta_{1s} medicaid0\_17_i$$
$$+ \beta_{2s} medicaid35\_64_i + \beta_{3s} medicaidhisp_i$$
$$+ \beta_{4s} medicaidblk_i + \beta_{5s} tax_i(100{,}130)$$
$$+ \beta_{6s} tax_i(200{,}300) + \beta_{7s} totpop_i + \beta_{8s} hisp_i$$
$$+ \beta_{9s} foodstamp_i + \beta_{10s} UIC_{i1} + \beta_{11s} UIC_{i2}$$
$$+ \beta_{12s} UIC_{i3} + \beta_{13s} UIC_{i5} + \beta_{14s} UIC_{i8}.$$

The prior distributions were specified as follows:

- $\beta_{i,s} \sim N(0{,}1000)$
- $v_{u,s} \sim \Gamma(0.1{,}1)$
- $v_{\varepsilon,ASEC} \sim \Gamma(0.1{,}1)$
- $v_{\varepsilon,SIPP} \sim \Gamma(10.0{,}10.0)$
- $\rho \sim U(-1{,}1)$.

The notation $\Gamma(\alpha, \beta)$ denotes the gamma distribution with mean α / β .

4. Model Fit and Estimation

Candidate models were chosen by examining scatter plots and other exploratory methods. Other research has also shown the utility of various versions of the predictors we chose. (See Fisher and Campbell (2002); Popoff, O'Hara, and Judson (2002); Popoff, Judson, and Fadali (2001); Lazerus *et al.* (2000); and Brown *et al.* (2001).)

[5] The Census Bureau selects the samples for each survey so there is no overlap, though there is an attempt to have overlap in the counties included in each to reduce field costs.

We rely on plots and posterior predictive p-values (PPP-values) to check the fit of the model. Given a function of the data $\mathbf{y}$ and the parameters θ, namely $T(\mathbf{y},\theta)$, the PPP-value is p($T(\mathbf{y}_{rep},\theta_{rep}) > T(\mathbf{y}_{obs},\theta_{rep})$). Here the subscript *obs* indicates the actual observed value while the subscript *rep* denotes a realization from the posterior distribution. (More detail is available in Gelfand (1998) and Gelman and Meng (1998).) Generally, PPP-values near zero or one indicate failures of the model to explain the data. We concentrate on PPP-values based on the following three defining functions:

- $T_1(\mathbf{y},\theta) = y_{is}$
- $T_2(\mathbf{y},\theta) = (y_{is} - X_i\beta_s)^2$
- $T_{31}(\mathbf{y},\theta) = \sum_i \frac{(y_{i1} - X_i\beta_1)^2}{(v_{u1} + v_{\varepsilon 1}/k_i^{1/2})}$
- $T_{32}(\mathbf{y},\theta) = \sum_i \frac{(y_{i2} - X_i\beta_2)^2}{(v_{u2} + v_{\varepsilon 2}\tilde{v}_i)}$.

Here i is the index for some county, s the index of a survey, and T_{31} and T_{32} apply to ASEC and SIPP, respectively. The first two functions give indications of the fit of the model with respect to the expectation and total variance, respectively, by county. We also summarize the resulting PPP-values by taking the mean across the counties to measure the overall fit of the models with respect to expectation and variance. The third is a measure of the overall goodness of fit. You *et al.* (2000) use this measure in their small-area estimation of unemployment.

We use Markov Chain Monte Carlo (MCMC) methods to sample from the posterior distribution of θ and the county log insured rate, and to evaluate the model. The implementation is a Metropolis algorithm written in GNU Fortran 77 (Brown and Lovato, 1993). The 'true' county log insured rates were integrated out so the joint posterior distribution of θ is $g(\theta\,|\,\mathbf{y}) = f(\mathbf{y}\,|\,\theta)p(\theta)$. The parameters were updated in two subsets; the coefficients composed one subset and the variance parameters composed the other. Then the county log insured rates, μ_i, were updated individually in a Gibbs step.

The advantages of this method include the fact that the posterior moments of a wide class of transformations can be calculated with little additional effort, facilitating the calculation of the posterior moments on the linear scale. Further, there is no need to approximate the standard error, as is necessary for the estimated best linear unbiased predictor (EBLUP). See, for example, Datta and Lahiri (2000). Finally, the fact that the set of counties without sample is different in ASEC and SIPP is handled by treating the direct estimates as unobserved variables in the model and generating them in the MCMC in exactly the same way.

5. Results

Parameter estimates and fit statistics are presented for the two models. Table 1 shows the overall posterior predictive p-values for the two models. For each model, the PPP-values are close to 0.5, so the models are not rejected. They are also close to each other, so are no help in discriminating between the models.

Table 1. PPP-values for three defining functions, two surveys, and two models; note model 2 has the same predictors as model 1, but also includes the UIC variables.

$T(\mathbf{y},\theta)$	Model 1	Model 2
$T_1(\mathbf{y},\theta) = y_{i,ASEC}$	0.49	0.49
$T_2(\mathbf{y},\theta) = (y_{i,ASEC} - X_i\beta_{ASEC})^2$	0.52	0.52
$T_{31}(\mathbf{y},\theta) = \sum_i \frac{(y_{i1} - X_i\beta_1)^2}{(v_{u1} + v_{\varepsilon 1}/k_i^{1/2})}$	0.45	0.47
$T_1(\mathbf{y},\theta) = y_{i,SIPP}$	0.51	0.51
$T_2(\mathbf{y},\theta) = (y_{i,SIPP} - X_i\beta_{SIPP})^2$	0.50	0.50
$T_{32}(\mathbf{y},\theta) = \sum_i \frac{(y_{i2} - X_i\beta_2)^2}{(v_{u2} + v_{\varepsilon 2}\tilde{v}_2)}$	0.46	0.42

Evaluation of the models proceeded by examining the plots of PPP-values against several other variables: the predictors, the population density, the innovation variable, the urban influence codes, and states. In the case of states, Alaska stood out; the mean PPP-value for Alaska counties was about 0.85 and, while there are only ten counties in that state, it appears that Alaska may not be described by these models well. Paying special attention to the PPP-values for the variance, we also examined the log of the ASEC sample size and the external variance estimate for SIPP, $\tilde{v}_i$. The plot of the PPP-values for ASEC versus the ASEC sample size shows a positive relationship, possibly indicating a weakness in the variance function. One hypothesis that seems consistent with the observation is that the variance function, $\frac{v_{\varepsilon 1}}{k_i^{\gamma}}$, $\gamma = 1/2$, decreases too slowly as a function of the sample size, k_i. Refitting the single-term model with $\gamma = 3/4$ leads to a plot with a slight negative relationship. Work proceeds to fit a model for that variance function. One possibility is to allow γ to vary, define a prior distribution for it, and fit the enlarged model. The plot of the PPP-values for the SIPP variance versus $\tilde{v}_i$ may show a negative relationship, though the effect is subtle. This will require more study.

The proportion uninsured is the variable of interest for many people. Simply subtracting the proportion insured from 1.0 forms these estimates. The posterior coefficients of variation (*cv'*s) of the estimates for SIPP uninsured rates are higher than those for the ASEC rates. Part of the reason is that the posterior variances are about twice as large, which follows from the fact that $v_{u,SIPP}$ is stochastically larger than $v_{u,ASEC}$. This may be a consequence of a misspecified variance model or it may really be the case that the insured rate as measured by the SIPP is not as highly correlated with the predictors as is the ASEC insured rate. This also requires further study. Two methods are available for that study. First, external variance estimates for the ASEC insured rates can be formed and used in a variance model similar to that in use for the SIPP. The other option is to try various models for the ASEC and SIPP variance functions.

The constant model error variance is historically useful in the SAIPE project, but various plausible assumptions lead to other forms. One appealing choice is to assume the true number of insured people is binomial random variable or a mixture of binomial random variables.

The PPP-values were also plotted against several variables related to the evaluations of the SAIPE models (see National Research Council, 2000). Variables derived from Census 2000 are: percent Hispanic, population size, percent of people living at rural addresses, percent of people living in group quarters, percent Black, and poverty. Also included are Metro/NonMetro and Census Division.

Each variable was used to divide the counties into categories of roughly equal sizes, then box plots were formed. In no case did the models seem to fail in the sense that there were no categories with PPP-values far from 0.5 and, when the categories were ordered, as with percent Hispanic, there was no apparent relationship between the PPP-values and the categories.

The two models' plots were visually identical. Also note that $P(f(data|\theta, \mathit{model\ 2}) > f(data|\theta, \mathit{model\ 1})) = 0.46$. This is a Bayesian version of the likelihood ratio test, and shows that once the considerable variability associated with the estimation of variances is included, the ability to distinguish between the models vanishes.

Tables 2a and 2b show the posterior means and standard deviations (SDs) for the ASEC and SIPP for model 1 and model 2.

Table 2a. Posterior means of the variance parameters in model 1 and model 2; note model 2 has the same predictors as model 1, but also includes the UIC variables.

	Posterior Mean	
Parameter	Model 1	Model 2
$v_{u,ASEC}$	9.3×10^{-5}	8.6×10^{-5}
$v_{u,SIPP}$	1.1×10^{-3}	1.1×10^{-3}
$v_{\varepsilon,ASEC}$	4.4×10^{-2}	4.3×10^{-2}
$v_{\varepsilon,SIPP}$	1.00	0.99
ρ	0.18	0.15

Table 2b. Posterior standard deviations of the variance parameters in model 1 and model 2; note model 2 has the same predictors as model 1, but also includes the UIC variables.

	Posterior Standard Deviation	
Parameter	Model 1	Model 2
$v_{u,ASEC}$	1.0×10^{-4}	1.0×10^{-4}
$v_{u,SIPP}$	1.9×10^{-4}	2.1×10^{-4}
$v_{\varepsilon,ASEC}$	2.6×10^{-3}	2.7×10^{-3}
$v_{\varepsilon,SIPP}$	3.9×10^{-2}	9.9×10^{-2}
ρ	0.51	0.52

The posterior mean for the SIPP sampling variance is close to 1.0 in both models, as it should be, since the prior distribution is strong around 1.0. In model 1, $P(\rho > 0.0 | data) = 0.65$. In model 2, $P(\rho > 0.0 | data) = 0.63$. This is only weak evidence that the random effect is correlated at all.

Tables 3a and 3b show the average posterior means, SDs, and *cv*'s for all of the counties.

Table 3a. Average posterior means, standard deviations, and coefficients of variation of uninsured rates for model 1.

	Posterior Mean	Posterior SD	Posterior *cv*
ASEC uninsured rate	0.15	0.012	0.089
SIPP uninsured rate	0.13	0.024	0.24

Table 3b. Average posterior means, standard deviations, and coefficients of variation of uninsured rates for model 2.

	Posterior Mean	Posterior SD	Posterior *cv*
ASEC uninsured rate	0.15	0.013	0.094
SIPP uninsured rate	0.13	0.026	0.25

The posterior *cv*'s for the ASEC uninsured rates are higher than those reported by Fisher and Turner (2003). One possibility is that the presence of the SIPP, with its correlated random effect and external sampling variance estimates, is affecting the posterior distribution of the county insured rates. Fisher and Turner (2003) expressed concern that the decomposition of the variance into random effect variance and sampling error variance was entirely dependent upon, and somewhat sensitive to, the variance models. The posterior variance depends directly on the variance model and parameters as well, so that sensitivity is inherited. The expectation is that the presence of an independent estimate of variance would improve the estimation of the variance terms and, therefore, the posterior variances.

6. Conclusion

We have presented a method to include both SIPP and ASEC into a single model in a way that allows the information from each to contribute to the reliability of all of the parameters and the county-level insurance coverage rates. This also has the effect that uninsured rates corresponding to each survey can be estimated. Posterior coefficients of variation for the ASEC uninsured rates average about 9.0 percent, while those for the SIPP are about 25 percent for model 1.

Additional variables, relative to those in Fisher and Turner (2003) were considered, namely innovation codes and urban influence codes. While the urban influence codes were significant in the exploratory regression model with the variances from the first model, they were not significantly different in the Bayesian model in any of the comparisons considered. Considering the fact that the innovation and urban influence codes are not produced more frequently than decennially, it does not appear that including them in the model for production will be fruitful.

Future work includes the perennial search for new predictors in the regression and for improved variance functions. The work described here is for the estimation of insurance coverage for all people; future research will involve the estimation for various sub-groups such as children, and children in families with incomes less than or equal to 200 percent of the federal poverty threshold. The last quantity is of interest because of the State Children's Health Insurance Program, which targets this group specifically.

References

Asher, Jana and Fisher, Robin (2000), "Alternate CPS Sampling Variance Structures for Constrained and Unconstrained County Models: Additional Research," accessed May, 2004, at http://www.census.gov/hhes/www/saipe/techrep/tech.report.4.pdf.

Bennefield, Robert L. (1996), "A Comparative Analysis of Health Insurance Coverage Estimated: Data from CPS and SIPP," SIPP Working Paper No. 218.

Bhandari, Shailesh (2004), "People with Health Insurance: A Comparison of Estimates from Two Surveys," SIPP Working Paper No. 243.

Brown, Barry W. and Lovato, J.; 1993; RANLIB random number generation library; <http://www.netlib.org/random/ranlib.c.tar.gz>; (accessed: May 2003).

Brown, Richard E., Meng, Ying-Ying, Mendez, Carolyn A., and Yu, Hongjian (2001), *Uninsured Californians in Assembly and Senate Districts, 2000*, UCLA Center for Health Policy Research, Los Angeles, CA.

Centers for Medicare and Medicaid Sercices; 2003; "MSIS Caveats and Data Limitations;" http://www.cms.hhs.gov/medicaid/msis/caveats.asp; (accessed: June 2004).

Datta, G.S. and Lahiri, P. (2000), "A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictors in Small Area Estimation Problems," Statistica Sinica, v. 10 pp. 613-627.

Fisher, Robin and Asher, Jana (2000), "Alternate CPS Sampling Variance Structures for Constrained and Unconstrained County Models," accessed May, 2004, at http://www.census.gov/hhes/www/saipe/techrep/tech.report.1.revised.pdf.

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

Fisher, Robin and Turner, Joanna (2003), "Health Insurance Estimates for Counties," *2003 Proceedings of the Section on Survey Research Methods*, Alexandria, VA: the American Statistical Association.

Gelfand, Alan E (1998), "Model Determination Using Sampling-Based Methods," In *Markov Chain Monte Carlo in Practice*, (eds W.R. Gilks, *et al.*), pp. 145-161.

Gelman, A. and Meng, X. (1998), "Model Checking and Model Improvement," In *Markov Chain Monte Carlo in Practice*, (eds W.R. Gilks, *et al.*), pp. 189-201.

Holahan, John and Pohl, Mary Beth (June 2002), "States as Innovators in Low-Income Health Coverage," The Urban Institute, Assessing the New Federalism: Discussion Paper 02-08, Washington, D.C.

Lazarus, W., Foust, B., and Hitt, B. (2000), *The Florida Health Insurance Study Volume 6: The Small Area Analysis*, State of Florida, Agency for Health Care Administration, Tallahassee, FL.

Lewis, Kimball, Ellwood, M., and Czajka, J. (July 1998), "Counting the Uninsured: A Review of the Literature," The Urban Institute, Assessing the New Federalism: Occasional Paper No. 8, Washington, D.C.

National Research Council (2000), *Small Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology,* Panel on Estimates of Poverty for Small Geographic Areas, Constance F. Citro and Graham Kalton, editors. Committee on National Statistics. Washington, DC: National Academy Press.

Popoff, Carole, Judson, D.H., and Fadali, Betsy (2001), "Measuring the Number of People without Health Insurance: A Test of a Synthetic Estimates Approach for Small Areas Using SIPP Microdata," presented at the 2001 Federal Committee on Statistical Methodology Conference.

Popoff, Carole, O'Hara, B., and Judson, D.H. (2002), "Estimating the Proportion of Uninsured Persons at the County Level: Exploring the Use of Additional Covariates in a Synthetic Estimates System," *2002 Proceedings of the Section on Government and Social Statistics*, Alexandria, VA: the American Statistical Association.

Rottach, Reid (2004), "Direct County Level Estimates Using Data from the Survey of Income and Program Participation," Unpublished U.S. Census Bureau technical document.

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 Proceedings of the Section on Government and Social Statistics*, Alexandria, VA: the American Statistical Association.