County and State Health Insurance Coverage Estimation

Robin Fisher

July 18, 2005

1 Introduction

The Small Area Health Insurance Estimates (SAHIE) program at the U.S Census Bureau is conducting research regarding the model-based estimation of insurance coverage for counties and has produced experimental estimates. There is widespread interest in estimates of health insurance at geographic levels below the state level. While there are several states that produce estimates of insurance coverage for counties within their boundaries, the estimates described here are novel in that they use a uniform method over all of the counties in the U.S. They also make use of several related administrative records, some of which are unavailable to the public.

Fisher and Turner(2003) model health insurance coverage, measured by the Annual Social and Economic Supplement to the Current Population Survey (CPS ASEC), as a function of several variables, including administrative records and demographic variables. The primary interest in that paper is to estimate insurance coverage as the CPS ASEC defines it. The Survey of Income and Program Participation (SIPP) was added as a correlated response by Fisher and Turner (2004) with two goals: first, to improve the precision of the estimated CPS ASEC insurance coverage, and, second, to provide estimates of insurance coverage as defined by SIPP. The benefits for the CPS ASEC insurance coverage were small and interest in the SIPP coverage was small, so development of that model has been deferred since the model was complex enough to make fitting a much slower process. One deficiency of the model in Fisher and Turner(2003), labeled FT in this paper, is the variance estimation. There is evidence in that paper that the variances, as modeled, do not fit the data that well. Improvements to the variance model are incorporated here.

Since the 2003 and 2004 papers, the models have been presented to several organizations including the State Data Centers (SDC) and the Federal and State Cooperative Program for Population Estimates (FSCPE). These groups had several suggestions, mostly for new covariates. We have also been considering the interpretability of the covariates and their coefficients and have concluded that a different set is indicated.

This paper presents the methods and covariates as they have been developed for the official release of the the experimental estimates. It is organized this way. In Section 2, I describe the structure of the model. Section 3 describes the predictors and response variables in the models. Section 4 describes some of the methods we use to evaluate the models. Section 5 continues that discussion with a discussion of the external evaluations. Section 6 notes that sometimes the covariates are missing for counties and describes the remedy. Section 7 describes the methods to ensure additive consistency between the geographic levels in these estimates. Section 8 shows results for the proposed models and shows how the SAHIE project has tried to use the suggestions of the professional organizations. Finally I conclude in Section 9.

2 Models

As the models have increased in complexity we have developed new software to fit them. Recent developments include the ability to fit flexible variance models, especially for the sampling error variance, which depend on the mean and increased flexibility with respect to the fitting of the different conditional distributions.

2.1 Structure

The models here have the same structure as that in FT, with modifications. The modeling of the sampling error variance has been a topic of ongoing research, and that is where most of the changes have been. Here I describe the model.

There is a quantity of interest for county, μ_i, here the log of the 'true' insured rate, p_i, though other transformations may be considered. Conditioned on some observable covariates and some parameters β and v_u, it has a truncated normal distribution,

$$
\begin{aligned}
f(\mu_i|\beta, v_u) &= \frac{c}{\sqrt{2\pi v_u}} \exp(-\frac{1}{2}\frac{(\mu_i - \mathbf{X}_i\beta)^2}{v_u}) & \mu_i < 0, \\
& \quad 0 & o.w.
\end{aligned}
$$

Here, $c^{-1} = \frac{1}{\sqrt{2\pi v_u}} \int_{\mu<0} \exp(-\frac{1}{2}\frac{(\mu_i - \mathbf{X}_i\beta)^2}{v_u})d\mu$. This ensures $P(p_i > 1|\mathbf{y}) = 0$.

The difference $u_i = \mu_i - \mathbf{X}_i\beta$ is referred to alternately as the *model error* or the *random effect*, and the parameter v_u is the *model error variance* or the *random effect variance*. The latter term is preferred here, though the former is a common interpretation. The quantity μ_i is not itself observable; what is observable is Y_i, the three-year average direct estimate of the health insurance rate from the CPS ASEC, Y_i, which has a normal distribution, conditioned on μ_i and a variance parameter $v_{\epsilon,i}$,

$$
Y_i \sim N(\mu_i, v_{\epsilon,i}). \tag{1}
$$

where $v_{\epsilon,i} = f(v_\epsilon, k_i, \mu_i)$; the form of the function f is a modeling assumption; two possibilities are given below. The quantity $\epsilon = Y_i - \mu_i$ is associated with the sampling error of the CPS ASEC.

FT considers the model of the form

$$v_{\epsilon,i} = \frac{v_\epsilon}{k_i^{0.5}}. \tag{2}$$

Here k_i is defined as the sum of the numbers of housing units in sample in the three years in the three-year average; I will refer to it as the sample size. Diagnostics shown in FT indicate that this variance model fails to describe a dependence on the proportion insured, which leads to consideration of more general variance models.

A variance model that contains the above as a special case as well as an approximation of one derived from the assumption that the variance should be a multiple of a binomial variance is

$$v_{\epsilon,i} = v_\epsilon(\frac{1-p_i}{p_i})^\gamma \frac{1}{k_i^\alpha}. \tag{3}$$

This model is not actually used for any of the results in this paper. The software, which will allow the fitting of this model is in the testing stages, so soon this model can be examined. An approximation is used for some of the models here. The variance function is

$$v_{\epsilon,i} = v_\epsilon(\frac{1-\tilde{p}_i}{\tilde{p}_i})^\gamma \frac{1}{k_i^\alpha}. \tag{4}$$

Here, $\tilde{p}_i$ is the posterior mean proportion insured from a previous run of the model.

For brevity, I denote the parameters $(\beta, v_u, v_\epsilon, \gamma, \alpha)$ as θ.

2.2 On the Log Transformation

The insured, as opposed to the uninsured, rate was chosen because there are many counties with no uninsured in sample, but few (in the study years) with no insured. In models estimated on the log scale, counties with zero-valued direct estimates are problematic in that the response is undefined. In the county-level Small Area Income and Poverty Estimates (SAIPE) program's models (Fisher 1997, NAS 2000), the direct estimates for these counties are not used, leading to a well-discussed bias in the estimates. It is also the case that several of the predictors are measures of insurance coverage of one group or another, so the formulation in terms of insurance coverage is more natural. In the data considered here, no counties have sample without any insured, and only 1 has sample without insured children in sample.

The log transformation was chosen on the basis of mild empirical evidence that that transformation is better-fitting than the linear transformation. The log model also has an attractive interpretation, in that the insured rate can be described as a product of ratios. The logistic transformation may be useful, though we would face the problem of undefined responses if it were applied to the direct estimates from CPS ASEC. It would certainly be possible to let $E(\mu_i|\beta) = X_i\beta$ while the expectation of Y_i is the inverse logistic transformation of μ_i. We don't explore that in this paper, though we still consider it a viable alternative model and it is under active consideration.

2.3 Priors

This is a Bayesian model, so we must define priors. In this paper, components of θ are independent in their prior distributions. They are defined to introduce small amounts of information. We have information about some parameters and better prior distributions may be chosen for them, but that has not yet been developed.

It would be reasonable to use prior years' data to form more-informative prior distributions about the current year's parameters. That will be studied in more detail in the future; in the current models, some of the administrative records variables are available for the first time for the years to which we fit the data.

3 Predictors and Response

FT contains a catalog of some of the covariates available. Since that paper was written, we presented results to members of the State Data Centers (SDC) and the Federal and State Cooperative Program for Population Estimates (FSCPE) [1]. Members of both of these organizations had suggestions for predictors:

- American Indian and Alaska Native (AIAN)
- Number of Firms
- Employment
- Proportion of the population aged 65 or more years

We have AIAN as a standard population estimate. Number of firms and employment come from the County Business Patterns (CBP) data, which are described in FT. The CBP data have some problems, notably the exclusion of some large

[1]The Census Bureau works with state-level organizations in the process of producing population estimates and in making data broadly available to the public. The FSCPE members work in cooperation with the Population Estimates Branch to produce subnational population estimates and participate in the review of those data. The State Data Center (SDC) Program is a cooperative program between the states and the Census Bureau that makes data available locally to the public through a network of state agencies, universities, libraries, and regional and local governments. The FSCPE and SDC members tend to have extensive experience and understanding of census data products and their respective local state data. More information is available at their web sites: http://www.census.gov/population/www/coop/fscpe.html and http://www.census.gov/sdc/www/

interesting groups such as state and federal government employees. We now have the Quarterly Census of Employment and Wages (QCEW) data, which should be better in this respect, but we have not had the chance to investigate them yet.

The proportion of people 65 years old or older is motivated by the idea that nearly every citizen is eligible for Medicare once they reach that age. That variable has turned out to be important for total insured, but not significant for child insured. In addition to the suggestions from the SDC and the FSCPE members, consideration of the interpretability and 'face validity' of the models has led to consideration of other sets of predictors than those in FT.

3.1 Predictors, Listed

For reference, a list of most of the predictor variables we have tried follows.

Log Population Log of the resident population in the age group of interest.

Log Hispanic Rate Log of the ratio of the number of Hispanics to the total resident population.

Log Food Stamp Rate Log of the of the ratio of the number of food stamp recipients to the number of tax exemptions associated with people in families with incomes less than 130% of the Federal Poverty Threshold (FPT).

Log IPR(a, b) Log of the proportion of the resident population with Income to Poverty Ratios (IPRs), defined as the ratio of the individual's family's income to its FPT, between $aFPT$ and $bFPT$.

Log Medicaid Rate Log of the proportion of the total resident population receiving Medicaid.

Log Child Medicaid Rate Log of the proportion of the resident population that is eligible for Medicaid as a child age 0 to 17.

Log Hispanic Medicaid Rate Log of the proportion of the total resident population that is Hispanic and receiving Medicaid.

Log Black Medicaid Rate Log of the proportion of the total resident population that is Black and receiving Medicaid.

Log Adult Medicaid 1 Log of the proportion of the total population that is aged 18-34 years and receiving Medicaid.

Log Adult Medicaid 2 Log of the proportion of the total population that is aged 35-64 years and receiving Medicaid.

Log AIAN Rate Log of the proportion of the total resident population that is American Indian or Alaska Native(AIAN).

Log Aged Rate Log of the proportion of the total resident population that is over the age of 64.

West Indicator Indicator for the event the county is in the West Census region.

South Indicator Indicator for the event the county is in the South Census region.

South-Hispanic Interaction Product of the indicator for the South Census region and the proportion Hispanic.

Mean Log IPR Mean of the log of the ratio of the Adjusted Gross Income (AGI) to the person's family's FPT, calculated from tax returns.

Variance Log IPR Variance of the log of the ratio of the Adjusted Gross Income (AGI) to the person's family's FPT, calculated from tax returns.

3.2 Response

The response Y_i is the log of a three-year weighted average of the direct CPS ASEC estimates of the insured rate, centered on the year of interest. Here the center year is 2001, which refers to the year 2000. The weight of a year in the average is the number of households in sample in that year. The sample size, k_i, is the sum of the numbers of households in sample in each the three years.

4 Evaluation of Model Fit

To judge model fit, we rely largely on posterior predictive p-values (PPPVs) and on plots of standardized residuals, as well as other plots. The PPPVs are defined as

$$p_j = P(T_j(y^{obs}, \theta^{rep}) < T_j(y^{rep}, \theta^{rep})) \quad (5)$$

$$= \int_{T_j(y>obs, \theta^{rep}) < T_j(y^{rep}, \theta^{rep})} p(y|\theta)p(\theta|data)d(y, \theta). \quad (6)$$

The T-functions used here follow.

1. $T_1(y, \theta) = y_i$
2. $T_2(y, \theta) = (y_i - \mu_i)^2$
3. $T_3(y, \theta) = \frac{(y_i - \mu_i)^2}{v_{\epsilon,i}}$

These are examined in the aggregate, and, for j=1,2, as plots for individual counties. In evaluating the SAIPE program's estimates, the National Academy of Sciences (2000) considered, among many other things, the distributions of standardized residuals among various subgroups defined by demographic variables. We produce, but do not show, analogous plots: plots of standardized residuals and PPPVs for the demographically-defined subgroups.

In an exploratory setting, it is expedient to use the results from an approximate weighted least squares (WLS) regression to evaluate predictor choices. Runs from the Bayesian model produce estimates (here, posterior means) of v_u and $v_{\epsilon,i}$ which can be used to calculate total variances, that is $v_u + v_{\epsilon,i}$, which can be inverted to provide weights in the WLS regression. If the models are close in some sense, the WLS regression will provide good approximations.

The WLS regression can be fit with Proc Reg (SAS Institute, 1999), and one statistic provided by Proc Reg is R^2. It would be nice, in spite of the well-known shortcomings of R^2, to have an interpretable version for use here. The R^2 from Proc Reg is clearly not directly interpretable as an R^2; its expression is

$$R^2 = \frac{SS_M}{SS_M + SS_E}, \tag{7}$$

but

$$E(SS_E) = \sum w_i(v_u + v_{\epsilon,i}), \tag{8}$$

where w_i is the weight in the WLS regression. In a perfectly fitting model, it may be that v_u gets small, but R^2 will not approach one, since $v_{\epsilon,i}$ is fixed with respect to the model fit and is positive. To get a more-easily interpreted measure of the fraction of "available" variance explained by the regression, we define

$$R_a^2 = \frac{SS_M}{SS_M + \sum w_i \hat{v}_u}. \tag{9}$$

When I report it, $\hat{v}_u$ will be the posterior mean of v_u from a run of the model being evaluated. R_a^2 is dependent and sensitive to the variance model, but it may be useful to those readers that like R^2 and seek a familiar statistic.

The FT variance function was inherited from the SAIPE program's variance model and has good fit when the PPPVs are plotted against k. Plots of the PPPVs or standardized residuals versus μ_i, the posterior mean log proportion insured, however, show a dependence on μ_i. The variance function in equation (3) is therefore considered; α and γ are estimated with their posterior means.

The exploration and evaluation of the variance model is carried out approximately for models with the form of equation (3). Our software is currently not set up to estimate the models with this form of $f_{Y_i}(y|\theta, \mu_i)$. We have been proceeding by calculating posterior means from the Bayesian model using the sampling variance model in equation (2), and using that in the variance function (4) to get new posterior means for the proportions insured, then doing another iteration by using those posterior means in (4) again to get the final set of posterior means.

The variances in the model are somewhat sensitive to the variance models chosen. In the absence of other estimates of sampling variance, we rely on the functional forms of the sampling and random effects variances to identify the components. There are clearly problems with this. First, it is easy to choose functional forms which have poorly identified parameters. That does not seem to be a problem here. Examination of scatterplots describing the joint posterior distribution of the variance parameters indicates that they are in fact well identified.

Another problem is more fundamental. If the true functional forms of the model and sampling error variances differ from our choices, as they must, then the allocation of variance to the two terms would have a systematic bias. Consider expanding the true functional forms into components, for example using a polynomial expansion. Then the constant part of either term would be assigned to the random effect component and much of the remainder would be assigned to the sampling error term, regardless of the true allocation: all the model does is estimate terms in an expansion. It is therefore important that we use some other estimate of variance, ideally in the estimation, but, failing that, in a validation step.

We have done some work along those lines. First, we have staff working to make replication-based estimates of the county-level variances similar to those in Rottach (2004), which were used by Fisher and Turner (2004). The time-line on that is not known. Second, we are using the generalized variance function (GVF) parameters available for the CPS ASEC to do a sort of validation of the variances. Unfortunately, it is not clear these measures have good properties in this application, since the GVFs were not designed to be used at these low geographic levels. This work continues.

Finally on this subject, consider the random effects variance. We have done little work on alternative random effect variance models. Rather than letting v_u be constant, it may be more appropriate to derive it from assumptions about a probability distribution in the population. Fortunately, the evidence does suggest that the random effect variance is small compared with the sampling error variance, so the posterior means and variances are not very sensitive to this assumption.

5 Comparisons To Other Data

Unlike the SAIPE program's estimates, there is no question on the decennial census or American Community Survey (ACS) for health insurance coverage to which we can compare the model-based estimates. There are, however, estimates produced by some states which are, at least in principle, available for comparison studies. This process is complicated by the varying statistical properties of the states' estimates.

Several states have their own surveys. Several of these also use models to improve the precision of their estimates. In some cases, they apply the models to only a subset of their counties and use direct estimates for the rest. Each state must therefore be considered separately, making proper consideration of the estimates' properties. The situation is further complicated by the fact that access to states' data is not always straightforward, and negotiation for that access may be time-consuming. Fortunately, the SAHIE program has an agreement with the University of Minnesota's State Health Access Data Assistance

Center (SHADAC), to help with the comparison.

We still have the task, however, of developing a method for the comparison of the SAHIE program's and states' estimates. Perhaps the simplest cases are those where the states just report direct survey estimates. Then correlations can be calculated and adjusted according to sampling errors as they are available. The cases where the states report model-based estimates for all counties is also relatively simple. Then correlations may be directly interpretable.

6 Imputation

Not every predictor is available for every county. In cases where a county lacks predictors, we impute the X-values simply by calculating the ordinary least squares prediction for the missing predictors given the others. The number of counties with missing predictors is relatively small (there are typically 20 or 30 counties with missing predictors, and they usually have no sample in the CPS ASEC), so the effect on the model is neglected.

The fact that the CPS ASEC is not available for approximately two-thirds of U.S. counties is handled automatically by setting the corresponding contributions to the loglikelihood to zero. The effect of this is equivalent to allowing $v_{\epsilon,i}$ to tend to infinity, so the full conditional distribution of μ_i does not depend on the data for county i.

7 State-Level Estimates, Raking, Rounding, and Confidence Intervals

We have taken as a requirement that the state-level estimates, both for insured and for uninsured, sum approximately to the national estimate and that the county-level estimates sum approximately to the corresponding state estimates. We also round the estimates so that integer quantities, such as the number of insured people, are represented as integers, and rates are presented as percentages with three significant digits. No attempt was made to constrain the rounding to preserve the sums, thus the approximation.

The CPS ASEC universe is a subset of the resident population, notably excluding institutionalized people and military personnel living in barracks. This quantity is the denominator of the rates we report, and is the quantity we would like to see when we take the sum of the insured and the uninsured. Demographic estimates which approximate the CPS ASEC universe at the county level are constructed for this study from relevant population estimates. We will refer to this demographic estimate for county i as $DASECPOP_i$.

The model-based estimates of the number of insured for states are the posterior mean sum of the product of the county estimated insured rates and the corresponding approximate CPS ASEC universes; this is, for state j,

$$\hat{ins}_j = \sum_{i \in j} p_i \; DASECPOP_i,$$

where the summation is over the counties in state j. The posterior variance is also calculated. This addition is convenient in the context of the Metropolis algorithm (Metropolis, *et al.*, 1953).

These state totals are adjusted so the sums equal the national CPS ASEC estimates of insured.

$$\hat{ins}_{R,j} = \hat{ins}_j \frac{ins_N}{\sum_j \hat{ins}_j}.$$

The number of insured for state j is just the difference between the CPS ASEC universe for that state and $\hat{ins}_{R,j}$.

County-level numbers of uninsured from the model are $DASECPOP_i(1 - \hat{p}_i)$. These numbers are adjusted to the state numbers as above, but here the adjustment is performed for both the insured and the uninsured.

The adjustment factors are treated here as fixed, known quantities. This is justifiable if we assume that the national CPS ASEC direct estimate of the insured has a low standard deviation compared with the posterior standard deviation of the county-level model-based estimates.

As a final note in this section, the confidence interval half-widths reported in the release are just the posterior standard deviations times the standard normal-theory based 1.645 for 90% confidence intervals.. This leads to a small number of counties with negative lower confidence bounds below zero.

8 Current Models

Here we present the current models for total insured and insured children. Note that, when interpreting the PPPV's, values close to 0 or 1.0 represent evidence that the model fits badly in some respect.

8.1 Total Insured

Response Log Insured Rate

Predictors Log IPR(2.0,3.0), West Indicator, South-Hispanic Interaction, Mean Log IPR, Variance Log IPR, Log Food Stamp rate, Log AIAN Rate, Log Aged Rate, Log Child Medicaid Rate, Log Hispanic Rate, Log Adult Medicaid Rate 2

Variance Model $v_{\epsilon,i} = v_\epsilon(\frac{(1-\tilde{p}_i)}{\tilde{p}_i})^\gamma \frac{1}{k_i^\alpha}$

Fit Measures
- $R^2 = 0.49$
- $R_a^2 = 0.80$
- PPPV for T_3: 0.54

Posterior *cv* for UI 0.15

Plot Notes

This model has an has an intuitively-appealing set of predictors and which seems to fit well. Predictors relating to the number of firms, and employment, normalized to the population, seem not to add much, and are not included. Recall analogous variables are available from the QCEW data set, and we expect those to work better.

8.2 Insured Children

Response Log Insured Rate

Predictors Log IPR(2.0,3.0), South Indicator, West Indicator, South-Hispanic Interaction, Mean Log IPR, Variance Log IPR, Log Food Stamp Rate, Log Child Medicaid Rate, Log Adult Medicaid Rate 2

Variance Model $v_{\epsilon,i} = v_\epsilon \frac{(1-\tilde{p}_i)^\gamma}{\tilde{p}_i^\gamma} \frac{1}{k_i^\alpha}$

Fit Measures

- $R^2 = 0.35$
- $R_a^2 = 0.83$
- PPPV for T_3: 0.46

Posterior *cv* for UI 0.18

The model for insured children also has an intuitive set of predictors and seems to fit well. Plots of ppp-values and standardized Bayesian residuals do not show large failures in the model for the means or variances, though higher moments do not seem to be described well.

8.3 SDC and FSCPE Suggestions

In late 2004, the SAHIE program staff presented the results of its research to members of the State Data Centers and the Federal and State Cooperative Program for Population Estimates. There was considerable positive response and several suggestions about improving the estimates. Those suggestions and our current findings regarding their utility follow.

- American Indian and Alaska Native. This variable has turned out to be significant in a statistical sense. The contribution is small, but the improvements in the AIC, for example, indicate its use.
- Number of Firms. The variable we have is from the CBP data. This variable turned out to be a marginal predictor, and, as it stands, not significant in the presence of other predictors. This variable has some potentially serious shortcomings, including the absence of government organizations. Further research will involve the use of other data sources which include government organizations.

- Employment. The paragraph above apples to employmet as well; the variable comes from the CBP and excludes some important groups such as many government employees. In this form, it was not a significant contributor to the predictive power of the model, and has not been included. Further research will include the use of other data sets such as the QCEW, which includes government employees.

- People Aged 65 or More Years. The nearly universal coverage of people in this age group makes it seem that it may be a good predictor of insured. It is significant in the model for total insured, but not for insured children, which is an intuitive result. We expect to include it in the model for total insured.

  Another approach to this group is to model it separately. This is under study, but there may not be much benefit, since the variance, under models similar to those in other age groups, may be higher than we'd like. This follows from the very low fraction uninsured. Further, the coverage in this group does not seem to vary that much, so the contribution to the posterior precision of the estimates is low, compared with just including the proportion of the population over 65 years old.

8.4 Substantive Interpretation of Parameters

Part of the model evaluation might be to examine the estimated parameters to see how they match our expectations. It is important to remember that the primary goal of the project is not to estimate coefficients, but to estimate the insurance coverage, which we do by estimating the conditional distribution of insurance given the predictors. Having said that, we would certainly like to see coefficient estimates we can understand so we can compare our results to those in the literature as a way to get some level of external validation. That activity has only been carried out in a basic way. The relationship between insurance coverage and the predictors is complicated enough that the interpretation is difficult. The coefficient for food stamp participation is positive, for example. One might expect that that parameter would be negative, as an indication of poverty. There are other measures of income in the model, however, and, perhaps, food stamp participation might be regarded as a measure of participation in aid programs in general. One might also ask whether the coefficients for Medicaid should be positive or negative; Medicaid itself is a form of insurance coverage, yet it is highly correlated with lower incomes, which in turn is negatively correlated with insurance coverage.

Further research on this topic, perhaps in collaboration with an expert on health insurance coverage in general, should yield benefits both with respect to understanding the relationship between insurance coverage and the predictor variables and more directly to reducing the mean squared error of the estimates coverage rates.

Table 1 Shows the posterior means and standard deviations (SDs) for the coefficients in the child insurance coverage model. The coefficients for Mean

Table 1. Posterior Means and Standard Deviations (SDs) for the Coefficients in the Child Insurance Coverage Model

Variable	Mean	SD
Log Poor IPR(2.0,3.0)	1.18×10^{-1}	2.12×10^{-2}
South Indicator	3.05×10^{-2}	8.15×10^{-3}
West Indicator	-4.20×10^{-2}	6.24×10^{-3}
South-Hispanic Interaction	-2.88×10^{-2}	5.02×10^{-3}
Mean Log IPR	1.94×10^{-1}	1.75×10^{-2}
Variance Log IPR	-8.96×10^{-2}	1.92×10^{-2}
Log Stamp Rate	3.24×10^{-2}	6.26×10^{-3}
Log Child Medicaid Rate	-3.28×10^{-2}	7.22×10^{-3}
Log Adult Medicaid Rate 2	3.61×10^{-2}	5.26×10^{-3}

Log IPR and the Variance Log IPR, the mean and variance of the log income-to-FPT ratio, have the expected sign. As the general income level increases, so does the proportion of insured. Conversely, as the income inequality, which is measured here with the variance of the log IPR, increases, the proportion insured decreases. As the log stamp rate increases (in the presence of constant 'wealth' measures), so does the proportion insured. This suggests that food stamp participation indicates readiness of the residents of counties to participate in insurance programs, either because there is a culture in the county or because the county or its state make it easy to participate in programs in general. It appears to be easier to get insurance in the South for other variables held constant, unless there is a high proportion of Hispanics. Hispanic effect is most pronounced in the South, where it is a disadvantage. One possibility under investigation is that the Hispanics that have settled in the South are particularly unlikely for some reason to have insurance.

Table 2 shows the posterior means of the coefficients in the model for all ages. We see many of the same relationships as in the child model in Table 1. In addition we see the positive coefficient for the log number of elderly, which is as expected. There is also a negative significant coefficient on Log AIAN Rate, which was not significant (or, indeed, negative) in the child insurance model.

9 Conclusion

The U.S Census Bureau is conducting research regarding the estimation of insurance coverage for counties and has produced experimental estimates. Those estimates and models are presented in this paper. The estimates include numbers of insured and uninsured for the counties and states along with measures of uncertainty.

The model diagnostics show only modest departures from the assumptions, at least with regard to the mean and total variance estimation, though higher moments appear to be poorly described. Further work also needs to be done on

Table 2. Posterior Means and Standard Deviations (SDs) for the Coefficients in the Model for All Insured.

Variable	Mean	SD
Log Poor IPR(2.0,3.0)	8.50×10^{-2}	1.54×10^{-2}
West Indicator	-1.77×10^{-2}	5.79×10^{-3}
South-Hispanic Interaction	-6.80×10^{-3}	2.42×10^{-3}
Mean Log IPR	1.30×10^{-1}	1.82×10^-2
Variance Log IPR	-6.60×10^{-2}	1.66×10^{-2}
Log Stamp Rate	1.60×10^{-2}	4.47×10^{-3}
Log AIAN Rate	-5.08×10^{-3}	1.81×10^{-3}
Log Aged Rate	2.25×10^{-2}	6.67×10^{-3}
Log Child Medicaid Rate	-1.60×10^{-2}	6.03×10^{-3}
Log Adult Medicaid Rate 2	-1.09×10^{-2}	2.33×10^{-3}
Log Hispanic Rate	1.86×10^{-2}	4.43×10^{-3}

the estimation of variance components, since the identification of those components is based on assumptions which can not be tested with this data. While the estimation of the insurance coverage itself is consistent as the number of counties gets large and is less sensitive to misspecified variance models, the measures of uncertainty are more vulnerable and should therefore be viewed with more scepticism.

Work continues to improve these estimates, and that includes forming separate estimates of variance for the CPS ASEC, investigation into the structure of the model, and the cultivation of covariates.

10 References

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

Fisher, R., and Turner, J. (2003), "Health Insurance Estimates for Counties", Presented at the American Statistical Association Meetings in San Francisco, California in August 2003.

Fisher, R., and Turner, J. (2004), "Small Area Estimation of Health Insurance Coverage from the Current Population Survey's Social and Economic Supplent and the Survey of Income and Program Participation", Presented at the American Statistical Association Meetings in Toronto, Canada in August 2004.

Metropolis, N., et al. (1953), "Equations of State Calculations by Fast Computing Machines" Journal Chem.Phys. 21, 1087-1092.

National Academy of Sciences (2000), *Small Area Estimates of School-Age Children in Poverty*, National Academy Press, Washington, DC, 2000

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

SAS Institute, Inc. (1999) *SAS/STAT User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999