SAHIE 2000 Age Model Methodology

Overview

Our age model creates county and state estimates of people with and without health insurance coverage by age. We have a county-level model and state-level estimates are aggregated from the county estimates. The estimates are adjusted so that, before rounding, county numbers sum to their states and similarly the states sum to the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) national estimates. We have developed experimental county and state estimates of people with and without health insurance coverage by age in 2000 for:

- total population;
- children under age 18; and
- measures of uncertainty of the estimates.

Estimation Details

Several features of the county estimates should be noted:

- We estimate the number of people with health insurance coverage. The method is a mixed effects linear regression, where the log proportion insured is modeled as a linear combination of several predictors, mostly administrative records. The proportion insured in the dependent variable is a 3-year average of county-level observations from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS). Although we use only the approximately 1,200 counties with CPS ASEC sample cases to estimate the equation, we estimate insurance coverage for all 3,140 counties in 2000.
- Since the regression is in the log scale, one can think of the model as multiplicative; that is, we model the proportion of people with insurance as the product of a series of predictors. While we may omit reference to logs in the description, all variables in the county regression models for proportions of people with insurance are logarithmic, except for the region indicator variables, which have the value of 1 for counties in the region and 0 otherwise.
- The CPS ASEC estimates for different counties are of different reliability because of the size of the sample and the proportions insured in them. Our estimates take this factor into account.
- To use the information contained in the direct estimates for the approximately 1,200 counties in the CPS ASEC, we combine the regression predictions with these direct estimates implicitly using Bayesian techniques. The effect of these techniques is to weight the counties' contributions to the parameters' estimates according to the precision of their direct estimates.
- We control the estimates for the counties so the following conditions are met:
    - state insured and uninsured sum to the CPS ASEC population for that state;
    - numbers of insured sum to the CPS ASEC national direct estimate of insured;

    - numbers of uninsured sum to the CPS ASEC national direct estimate of uninsured;
    - county numbers of insured sum to state numbers of insured; and
    - county numbers of uninsured sum to state numbers of uninsured.
- We form the state-level estimates by aggregating the county-level estimates.
- We provide a confidence interval, which represents uncertainty from both sampling and modeling, for each estimate.

Model Details

Several features of the county estimates should be noted:

The model is multiplicative; that is, we model the proportion of people with insurance as the product of a series of predictors that are mostly rates, and we model the unknown errors. To estimate the coefficients in the model, we take logarithms of the dependent and all predictor variables, except for the region indicator variables, which have the value of 1 for counties in the region and 0 otherwise. Another advantage of a multiplicative model is that it makes it plausible to maintain that the (unobserved) errors for every county, no matter how large or small, are drawn from a normal distribution, which is how they are modeled. The regression predictions are, in effect, combined with the direct CPS ASEC sample estimates. Finally, we control the county estimates to the national CPS ASEC estimates and form the state-level estimates.

Model for the proportion of people with health insurance coverage

Dependent variable:

- log of the proportion insured in each county as measured by the 3-year average of values from the CPS ASEC.

Predictor variables:

- log of the proportion of people with family Income to Poverty Ratios (IPRs) between 200% and 300%, as estimated from tax returns;
- mean of the log IPR, as estimated from tax returns;
- variance of the log IPR, as estimated from tax returns;
- log proportions of persons under age 18 who are participants in the Medicaid program;
- log proportions of persons age 35-64 years who are participants in the Medicaid program;
- log proportion of the population who are receiving Supplemental Nutrition Assistance Program (SNAP) formerly known as the Food Stamp program;
- indicator for the West Census region;
- log of the proportion of people of Hispanic origin from demographic population estimates;
- product of the indicator variable for the South Census region and the log proportion Hispanic;
- log of the proportion of people who are American Indian or Alaska Native from demographic population estimates; and

- log proportion who are 65 or more years old from demographic population estimates.

Model for the proportion of children under age 18 with health insurance coverage

Dependent variable:

- log of the proportion insured under age 18 in each county as measured by the 3-year average of values from the CPS ASEC.

Predictor variables:

- log of the proportion of people with family Income to Poverty Ratios (IPRs) between 200% and 300%, as estimated from tax returns;
- mean of the log IPR, as estimated from tax returns;
- variance of the log IPR, as estimated from tax returns;
- log proportions of persons under age 18 who are participants in the Medicaid program;
- log proportions of persons age 35-64 who are participants in the Medicaid program;
- log proportion of the population who are receiving SNAP;
- indicator for the West Census region;
- indicator for the South Census region; and
- product of the indicator variable for the South Census region and the log proportion Hispanic.

Using Counties in the CPS ASEC Sample

Our use of the CPS ASEC implicitly assumes that the counties in the survey sample are representative of those not selected. The CPS was designed so that Primary Sampling Units (PSUs) are representative of their strata, primarily for unemployment, but the degree to which the CPS ASEC sample is representative for health insurance coverage is unknown. The characteristics of some counties guarantee they are included, e.g., most counties in large metropolitan areas and counties with large populations. More generally, while all counties have a nonzero probability of being included in the sample, some have higher probabilities than others. Further, the probability of selecting a county is related to its income and poverty level which, in turn, are related to the level of health insurance coverage. In the related Small Area Income and Poverty Estimates (SAIPE) program, comparison of regression equations based on census data for counties in the CPS ASEC sample and equations based on all counties indicates remarkably similar results, providing some assurance that the CPS ASEC counties are largely representative of all counties for poverty. Unfortunately, the analogous test is unavailable for health insurance coverage, since there are no health insurance questions on the decennial census.

The survey weights used in estimation at the national level are not appropriate for county-level estimates. The CPS ASEC sample design selects some PSUs (usually a county or group of counties) to represent a set of counties in the same stratum. The sum of the weights for sample households from such a county estimates the total population of the entire set of counties it represents. Because we want each county in the CPS ASEC sample to stand for itself, we have adjusted the weights to make the direct estimate for each county approximately unbiased.

Estimation of the Model Equation

CPS ASEC sampling variances are not constant over all counties. We avoid giving observations with a great deal of uncertainty (larger variances) the same influence on the regression as observations with less uncertainty (smaller variances) by, in effect, weighting each observation by the inverse of its variance. Representing this uncertainty requires recognizing that it arises from two sources:

- uncertainty about where the estimates lie relative to the true values for each county (sampling error), and
- uncertainty about where the true county values lie with respect to the regression surface (lack of fit).

To estimate the two components of variance, we model them as having different forms. We model the sampling error variance to depend on the sample size and on the proportion insured. The lack-of-fit component, on the other hand, is modeled as constant across all counties. Then the components can be distinguished using our Bayesian estimation method.

Model-based County-level Estimates

The estimated insured rate from the modeling is the posterior mean insured rate conditioned on the CPS ASEC data. The effect of this is similar to that of the empirical Bayes method used in the SAIPE program's estimates. The final estimates for counties where there is no sample is the same as the regression estimate, while the estimates for counties with lots of sample or very high insured rates and, thus, low variance, tend to be closer to the direct estimates.

The estimated number of insured in a county is the estimated insured rate times an estimate of the CPS universe. We create an estimate of the CPS universe by adjusting estimates of the total resident population to the CPS universe by subtracting unpublished demographic estimates of the group quarters population by age and the appropriate type of group quarters from the estimate of the total resident population. The number of uninsured, then, is that estimated CPS universe minus the estimated number of insured. The reported confidence intervals are based on the posterior standard deviation of the insured rate, conditioned on the CPS ASEC data.

Controlling to the National CPS ASEC Estimate and Forming the State-level Estimates

The last steps in the production process are controlling the county estimates to the national CPS ASEC estimates and forming the state-level estimates. The number of uninsured from the model are aggregated to the state and national levels, and the ratio of the national CPS ASEC direct estimate to the aggregated national model-based estimate is formed; this ratio is the raking factor. The raking factor is multiplied with all of the county- and state-level uninsured to get the controlled numbers of uninsured. This is subtracted from the state and county CPS ASEC universe estimates, yielding the estimated numbers of insured. Finally, everything is rounded to an integer.

Standard Errors and Confidence Intervals

One goal of our small area work is to provide measures of the uncertainty surrounding the estimates. The model-based estimates shown in the tables are accompanied by their 90-percent confidence intervals constructed from estimated standard errors.

We assume that the variance at the national level and the variance of the CPS ASEC universe estimates are negligible. The posterior standard deviations of the aggregated state-level estimates need only be adjusted for correlations between the counties, which is handled by the estimation procedure, and multiplied by the raking factors. Confidence interval half-widths for estimated numbers are rounded up to preserve coverage probabilities. Note also that the widths of the confidence intervals are the same for the number of insured and uninsured. This follows from the fact that the two must add up to the national CPS ASEC estimate which has negligible variance.