EVALUATION OF THE CENSUS BUREAU'S SMALL-AREA POVERTY ESTIMATES

Michael L. Cohen, Committee on National Statistics
2101 Constitution Ave., Washington, DC 20418

KEY WORDS: small-area estimation, poverty estimates

The following presentation is a summary of the work of the Committee on National Statistics' Panel on Estimates of Poverty for Small Geographic Areas and the Census Bureau's SAIPE program staff. The author served on the staff to the panel, which was chaired by Graham Kalton and directed by Constance Citro.

OVERVIEW

Legislation of the Elementary and Secondary School Act requested that a National Academy of Sciences panel examine the new Census Bureau model-based estimates of the number of children living in poverty at both the county and school district levels for use in Title I allocation formulas in place of decennial census long-form estimates. The statute requires that the Department of Education use the Census Bureau's updated estimates for allocation "unless the Secretaries of Education and Commerce determine that they are inappropriate or unreliable on the basis of the panel's study. It remained for the panel to interpret the terms unreliable or inappropriate, and this paper summarizes the panel's efforts to do this.

The remainder of this paper will describe, in turn: (1) the models proposed by the Census Bureau at the state, county, and school district levels, (2) the formation of alternative models, at the suggestion of the panel, for evaluation purposes, (3) the specific evaluations, both internal and external, that the Census Bureau and the panel carried out, and (4) next steps.

Many important details are intentionally ignored here due to space considerations. Examples of some of the complications are that the year of poverty estimation does not coincide with the year of fund allocation, some complications regarding the precise population that is the target population being estimated, that there are two fund allocation formulas involved, many complications concerning school districts as a unit of geography and analysis, and various important details concerning the Current Population Survey (CPS) sample design. For those interested, the recent National Academy report *Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond* National Research Council (2000) is now available.

As mentioned above, the work reported on was jointly that of the Census Bureau, specifically Paul Siegel, John Coder, Bill Bell, Robin Fisher, Bob Fay, Mark Otto, and Matt Kramer along with the Panel on Estimates of Poverty for Small Geographic Areas which provided detailed suggestions and requests. The panel members were Graham Kalton (chair) (Westat), David Betson (Notre Dame), Nancy Dunton (Midwest Research Institute), Wayne Fuller (Iowa State), Tom Jabine (consultant), Sylvia Johnson (Howard), Tom Louis (RAND), Sally Morton (RAND), Jeff Passel (Urban Institute), J.N.K. Rao (Carleton), Allen Schirm (Mathematica Policy Research), Paul Voss (University of Wisconsin), Jim Wyckoff (SUNY Albany), and Alan Zaslavsky (Harvard).

THE THREE MODELS TO EVALUATE

The Census Bureau developed separate models at the state, county, and school district levels. The state-level model has the following form:

$$y_i = \alpha + \sum_{k=1}^{4} \beta_k x_{ki} + u_i + e_i$$

where:
y_i is the <u>proportion</u> poor aged 5-17 in state i from CPS
x's are proportions from the census and adm. records
u_i is model error for state i
e_i is sampling error for y_i.

Note that: (1) these regression predictions are smoothed to the direct estimates y_i (random effects model), and (2) the estimates are then controlled to CPS national estimates.

The county-level model has the following form:

$$y_i = \alpha + \sum_{k=1}^{5} \beta_k x_{ki} + u_i + e_i$$

where:
y_i is a <u>log</u> 3-year average of the <u>number</u> poor aged 5-17 in county i from the CPS
x's are log numbers from the census and adm. records
u_i is model error for county i assumed homogeneous
e_i is sampling error for y_i, (sometimes) assumed proportional to the inverse of the CPS sample size.

Note that: (1) the (log) number of poor is modeled rather than the percentage poor to avoid estimating the variance of the denominator, (2) the log scale is used to promote symmetry of the distributions of the variables and to promote linearity of the relationships between the dependent and independent variables, (3) there is an analogous model fitting census estimates of poor to the same independent variables which is used along with an heroic assumption that census model error variance is equal to CPS model error variance to estimate the variance of u_i and $e_{i.}$, (4) there is a correction for bias in transforming from the log scale back to the original scale, (5) as in the state-level model, the regression predictions are smoothed to the direct estimates, and (6) the county-level estimates are then controlled to the state-level estimates.

The school district - level model has the following form (which is very simple due to data limitations):

current sch. dist. estim. of # poor 5-17	=	sch. dist. share of county from census	x	current county-level estimate from model

ALTERNATIVE MODELS

The panel suggested the development of alternative models at each of the three levels for purposes of comparison, to serve as benchmarks, and to suggest possible changes in the models proposed for use by the Census Bureau possibly by incorporating features that seemed to provide advantages.

In the state-level model few alternatives were identified, though Fay did examine use of the log transform, a multivariate approach combining the inference for other age groups, and a few other alternative approaches.

In the county-level model, 13 or more alternatives were examined. This included the following choices: (1) log vs. no log, (2) # poor vs. proportion poor, (3) whether or not to include fixed state effects, (4) a single equation formulation, as described above, or a bivariate formulation (see below) due to Bill Bell (which had limited evaluation opportunities), and (5) choice of an independent variable of an estimate of the population under age 21 or under age 18.

Bill Bell's bivariate model can be described as follows. In stead of the current county-level model, written in matrix form as:

$$y_{it} = (X'_{it}\beta + \gamma Cen90_i + w_{it}) + e_{it}$$

$$Cen90_i = (X'_{it}\eta + \hat{\gamma} Cen80_i + \tilde{z}_i) + \varepsilon_i$$

(recall that the census model is used to estimate variances) Bell, by expressing this as a two-equation system with correlated model errors, arrived at the following representation:

$$y_{it} = (X'_{it}\beta + \gamma_i(Cen - X'_{i,89}\eta) + \tilde{w}_{it} + e_{it}$$

$$Cen_i = (X'_{i,89}\eta + z_i) + \varepsilon_i$$

The major differences between these two formulations are:

(1) γ_i can vary by county

(2) the census residual is the covariate, not the census.

In addition to these relatively sophisticated alternatives, some additional "simplistic" county-level alternatives were also used for evaluation, and they are described based on the assumptions that they make use of:

(Model 1) constant shares, from the previous census, of the current national poverty count

(Model 2) constant shares, from the previous census, of the current state poverty counts

(Model 3) constant state poverty rate from the previous census.

The hope was that these alternatives would provide an assessment of what the county-level model offered over simple corrections to use of the previous census estimates, which was the situation before development of the model-based estimates.

Finally the panel suggested use of three alternatives to the school-district level model:

(Model 1) the product of 1980 school district shares as part of a county and the 1990 census county estimates. (This was utopian in the 1990 evaluations since the 1990 census county-level estimates would not be available;

(Model 2) the product of 1980 school district shares as part of the state and the 1990 census state estimates, (which are also utopian); and

(Model 3) the product of 1980 school district shares of the nation and the 1990 census national estimate. This last alternative essentially is using the school district estimates from the most recent census.

Unfortunately, these alternatives were not very different, limiting the evaluation information supplied.

(4) A fourth alternative, which was not as much an alternative model but an alternative data source, were estimates based on participation in free and reduced school lunch programs. These are programs that provide free and reduced price meals to needy children. The information is unevenly collected and eligibility in these programs is not the same as being in poverty. This alternative was only examined for New York and Indiana.

EVALUATIONS

There is a lot of combinatorial complexity that must be confronted in representing the work that was done by the SAIPE staff in conjunction with the panel. There are three levels of models (state, county, and school district-level), and aside from school districts, there are multiple years of data used (three years of data 1989, 1993, and 1995 for the county-level model, and more for the state-level model), there are external validation and internal validation methods, both involving more than one loss function, and there are several alternative models. So, it was necessary to pick and choose a bit in this presentation. Also, it was necessary to ignore some additional important topics in this presentation, namely:

(1) in addition to model evaluations, there were data evaluations, which is a vitally important step in a model evaluation. The Census Bureau has carried out considerable work on the quality of the IRS data, food stamp data, and long form data it uses, making improvements in each case;

(2) another input was that of intercensal population estimates, which are used in two or three ways in the poverty estimates program, both as a covariate, but also in coverting from rates to levels and levels to rates. The Census Bureau has an ongoing program for evaluation of its population estimates.

The general game plan of the panel was:

(A) to measure the overall performance of these various estimates, including the alternatives,

(B) to identify sources of persistent bias, both for model improvement, and also for model evaluation, since a model with evidence of substantial, persistent bias will probably not distribute funds as equitably over time as one with less substantial evidence of bias.

County-level model evaluation

External validation.

An external validation is an examination of the closeness of estimates to comparison values (truth substitutes). There are substantial benefits from replications in external validation, and replicated evidence of bias is especially important to collect.

(1) Use of Census for comparison values

The 1990 Census provided important comparison values to evaluate poverty estimates. Unfortunately, the following problems are associated with their use: (1) at this point there is only one replication, (2) these estimates are subject to appreciable sampling error for small areas, and (3) the census also has measurement error when the CPS measurement is considered to be the gold standard for poverty. This last problem can be partially addressed by controlling the census poverty counts to national CPS poverty counts before using them for evaluation; however there are other possible differences between CPS and census poverty measurement besides overall level.

In terms of overall measures of loss, the panel suggested use of the following two loss functions:

$$\sum \left| Y_{\mathrm{mod},i} - Y_{cen,i} \right| / n$$

$$\sum \left| Y_{\mathrm{mod},i} - Y_{cen,i} \right| / (nY_{cen,i})$$

These two loss functions were computed for several alternative models using 1980 census data and 1989 CPS data, with estimates from the 1990 census serving as surrogates for the truth. The result is the following table for four leading models and three "unsophisticated" models:

Overall Measures

	1989 estimates vs 1990 Census		
Models	Loss Func 1	Loss Func 2	% agree (15)
log #(<21)	272	15.4	87.1
log#(<18)	268	16.4	87.6
log rate(<21)	275	17.5	87.4
log rate(<18)	283	18.8	87.0
const. shares	570	30.1	
const. shares in state	380	27.1	
const. rates	381	26.2	

"% agree (15)" is another loss function pertaining to concentration grants within Title I allocations, which measures how often the designation that an area is above or below 15% poverty agrees with the comparison estimates, 15% being the eligibility threshold. This table shows that in an overall sense, the log#(<18) model, which is that proposed for use by the Census Bureau, is fully competitive with the alternative models.

In addition to overall measures, the census data was also used to search for sources of bias. For this purpose the following variables were used:
Census Division, metro status, 1990 population, 1980-90 change in population, percent Hispanic, percent black, % persistent rural poverty, percent group quarters, whether a county is in or out of the CPS sample design, and change in poverty rate. Two measures were used for this purpose, one of which is the following:

$$\sum_{i \in cat(j)} \left(Y_{\mathrm{mod},i} - Y_{cen,i} \right) / \sum_{i \in cat(j)} Y_{cen,i}$$

To demonstrate the analysis, we examine the output for 1990 population size and for percent Hispanic. For these two variables, we have the following categories, and associated measures for the models indicated:

Bias Measures
Alternative model: Log#(<21)

1990 population size	loss function
< 7,500	-9.0
7,500 - 14,999	-4.4
15,000 - 34,999	-5.1
35,000- 49,999	-4.2
50,000 - 99,999	-3.5
100,000 - 249,999	-1.8
250,000 +	3.3

Log # (<18)

% Hispanic	loss function
0.0 - 0.9	-3.3
1.0 - 4.9	0.1
5.0 - 9.9	-0.6
10.0 - 24.9	1.8
25.0 - 98.0	4.6

These monotonic patterns in the bias measure across categories indicates that, possibly ignoring measurement error in the census, there is evidence that the indicated model has a bias with respect to these variables. In other words, areas that are low or high for these variables have estimates that are low or high relative to other areas, and as a result some areas will receive more or less funding than was intended.

Each of the candidate models exhibited similar patterns for these two or other variables. Without replications it is hard to know whether this bias is simply an idiosyncracy for the given year's data or part of a persistent pattern of bias. A bias was also discovered for counties in the West Region, but since the definition of census regions obeys state boundaries, and since county-level estimates are controlled to state-level estimates, this is a state-level model issue.

(2) Use of CPS for comparison values

This is discussed below in the section on internal validation, since the CPS is the dependent variable of the model.

(3) Use of local information for comparison values

Voss et al. (1997) investigated the face validity of county poverty estimates by comparing them to expert knowledge in local areas in Wisconsin. The result was that nothing was known that cast great suspicion on the quality of the estimates.

Internal Validation

Internal validation is an internal, component - by -component examination of a model to see if the assumptions are justified that are used in the model. For standard regression models, the statistics and graphs used are relatively standard, though for random effects models the tools are a little less well understood. Not surprisingly, regression diagnostics focuses primarily on the properties of the residuals from the regression model.

The assumptions examined were as follows:

1. Linearity
2. Constancy of regression coefficients over time
3. Normality (symmetry)
4. Homogeneous variances of standardized errors
5. Absence of outliers
6. While not an assumption, the inclusion of extraneous or the exclusion of useful independent variables.
7. In general, looking for non-random patterns in the residuals.

The residual plots supported the linearity assumption. With respect to constancy, here are the 5 regression coefficients for three models years:

	X_1	X_2	X_3	X_4	X_5
1989	.52	.29	1.55	-1.56	.26
1993	.38	.27	0.65	-0.59	.34
1995	.31	.29	0.88	-0.80	.33
Std. Dev.	(.08)	(.06)	(.28)	(.28)	(.08)

These coefficients appear to be relatively constant over time.

Q-Q plots of the standardized residuals supported the normality assumption for all models. A variety of plots, many from Carroll and Ruppert (1988), were used to diagnose heterogeneity of the error variances (in addition to the assumed heterogeneity based on different CPS sample sizes). The plots indicated a dependence of the variances on CPS sample size and on estimated proportion poor.

With respect to outliers, the fifteen largest and smallest (signed) residuals were identified and examined to see how they were distributed across the categorized variables described above. The notion was that outliers that tended to occur when certain variables were high or low might indicate the need to add some version of that variable to the model. However, no such structure was evident.

With respect to whether some included independent variables should be removed from the model, it was clear from looking at t-tests and measures related to Cp that some of the independent variables were not contributing importantly to the fit. However, given the relatively small number of independent variables, this was not considered important to act on. With respect to adding other independent variables to the model, in addition to the outliers analysis described above, the following analysis was conducted. First, box plots of standardized residuals for subsets of the data in categories determined by the above variables were examined. (For example, if there were 6 categories of data grouped by population size, 6 box plots for the associated residuals would be drawn.) The idea was that if the average residual (or possibly other moment) tended to be related to the monotonic change in category for some variable, that variable might be profitably added to the model as an independent variable (or be helpful in achieving more homogeneous variances). No interesting patterns were discovered. A little easier to interpret was the use of the following measure:

$$\sum_i \left(Y_{\text{mod},ij} - Y_{CPS,ij}\right) / \sum_i Y_{CPS,ij}$$

computed for all data within the jth category. This analysis is clearly relevant to the more general concern with model bias. While this measure is completely analogous to the measure used when the census estimates were providing comparison values, it is harder to interpret this measure for individual years since the sampling variance here is much larger. On the other hand, since there are repetitions of CPS data, and since there is no measurement error, this measure might be more successful in picking up persistent patterns of bias. Also, it is useful to point out that in some sense this analysis is an external validation in which the CPS plays a surrogate for the truth. (Here we are not looking at groupings of individual residuals but instead "residuals" for groups of counties.) The most interesting finding from this analysis for the current model, was as follows:

% Hispanic	1989	1993	1995
0.0 - 0.9	2.13	-0.75	1.26
1.0 - 4.9	4.32	1.45	9.33
5.0 - 9.9	6.38	17.24	-2.81
10.0 - 14.9	-8.29	-5.14	-4.02
25.0 - 98.0	-5.26	-3.29	-7.90

Clearly, there is evidence that the model is underpredicting poverty for areas with a large Hispanic population.

Other patterns worth mentioning are: (1) a bias in rural counties with state governments, and (2) underprediction of the number poor 5-17 in large urban counties.

In summary, regression diagnostics demonstrated that all but a few of the leading candidates for the county-level model looked relatively well-behaved with a small number of disturbing patterns.

There remain two last points worth making on evaluating the county-level models. First, so-called "raking factors", defined as the sum of county estimates within a state divided by the state estimate seemed to exhibit more variation than one would expect. For example, in 1993 the raking factors ranged from .91 to 1.31 and in 1995 they ranged from .71 to 1.14. It is not clear how useful this is as a way of indicating the need for state effects in the county-level model. This is because it is hard to develop a variance estimate of these ratios. However, the raking factors do not seem to be correlated over time, and so it is not obvious that they are indicative of the need for state fixed or random effects.

Second, tests of the constancy of γ_i showed that they were not constant, which provides evidence of the benefits of the bivariate formulation of the county-level model.

State-level model evaluation

Initially, due to the better behaved dependent variable given the larger sample sizes, and also due to the more traditional form of the model, the panel gave the state-level model's evaluation a lower priority than the county-level model. There are other differences in the two models that bear on evaluation:

(1) the state-level model is fit on a yearly basis, providing more replications,

(2) as mentioned above, fewer alternative model forms were developed, and

(3) since there are only 50 states and DC, it is hard to cluster the states into somewhat homogeneous groups as was done in some of the evaluations of the county-level models.

External validation

Bob Fay carried out some comparisons of the state-level model using the census estimates as surrogates for true values. He demonstrated a close correspondence between the CPS-based state-level estimates and the census estimates from the long form for 1990. In addition, the panel directed an analysis of the confidence intervals for the CPS direct estimates to see whether the model-based estimates were consistent with these intervals, and they were consistent.

Internal validation

Again, the internal validation was primarily standard regression diagnostics. The same assumptions were examined as before for the county-level model.

Linearity was assessed using residual plots, which did not indicate any curvature. Constancy of regression coefficients was examined using the following table:

Year	X_1	X_2	X_3	X_4
1989	.52	.71	.23	.71
1990	.46	.65	.42	1.07
1991	.46	.52	.59	.84
1992	.41	.71	.42	1.38
1993	.28	1.14	.51	1.24
1995	.57	.79	.32	1.54
1996	.37	.97	.59	1.02
Std. Dev.	(.10)	(.24)	(.14)	(.36)

The regression coefficients were considered to be consistent with a stable model of poverty change over time. (That is, this did not conflict with expert opinion as to how quickly the dynamics of poverty were likely to change over time.)

With respect to discarding independent variables currently in the model, there was little interest in dropping any variables. With respect to discovering new independent variables that should be added to the state-level regression model, residuals were grouped according to categories for percent Hispanic, percent black, percent group quarters, and percent poor. Box plots, as described above, only showed one pattern of interest, and that was that the model underpredicted poverty in the West Region, which was persistent over time. There is no proposed explanation for this at this time. The Census Bureau is considering adding a variable to the model to reduce this problem.

Examination of normality, heterogeneous variances (beyond what was modeled), and outliers showed no problems. However, a major issue that was in need of further analysis was that the model error variance, fit using maximum likelihood, was estimated to be equal to zero. While this has only a modest impact on the estimates, the notion that very large states would have estimates that gave no weight to their associated direct estimates and full weight to the estimates from the regression model seemed difficult to support.

School district-level model evaluation

External validation

The census was used to externally validate the school-district-level estimates. In doing so, the high level of variance of the census estimates must be kept in mind. Two measures of global performance were used:

$$1. \sum \left| Y_{\text{mod},i} - Y_{cen,i} \right| / \sum Y_{cen,i}$$

$$2. \sum \left| Y_{\text{mod},i} - Y_{cen,i} \right| / nY_{cen,i}$$

where n is the number of counties. The results for two of the three "unsophisticated" alternatives described above were:

	Measure 1	Measure 2
Current model	22.2	60.0
Census controlled to Census county-level est.	18.0	55.4
Census controlled to national estimates	28.7	71.7

This indicated that the current estimates outperform the census estimates relatively untouched, and do not do substantially worse than utopian estimates that make use of the unavailable current census county-level estimates. (The current model and the census controlled to census county-level estimates differ only in that one uses CPS-county-level estimates and one uses the census county-level estimates.)

Examination of bias was analogous to that for the county-level estimates and used the following bias measure:

$$\sum_i \left(Y_{\text{mod},ij} - Y_{cen,ij} \right) / \sum_i Y_{cen,ij}$$

within category j. This was computed for the following variables: census division, 1980 population, 1990 population, 1980-90 population change, % poor in 1980, % poor in 1990, change in percent poverty, % Hispanic, % black, and % group quarters. The results were that there was evidence of overpredicting in school districts with a low percent of minorities and with a small population, and there was evidence of a problem for the school districts in the Pacific census division. However, these indications are only suggestive since this analysis has only one replication at this time.

An additional global measure of performance was derived from the use of these estimates as input into concentration grants, one of the two specific allocations in Title I. The eligibility threshold for concentration grants is whether poverty is greater than 15%. The agreement of the current model with the census was 76%.

Finally, estimates based on free and reduced price school lunch data in New York and Indiana (work performed by Jim Wyckoff and David Betson) were compared to the current estimates using the census estimates as surrogates for the truth. In these two states, the performance of the current estimates were comparable to the estimates based on this alternative source of data (which is currently not available nationally).

Next Steps

The alternative models typically performed essentially equal to or worse than the Census Bureau's current approach at all three levels. Therefore, there is currently no need for any major modification such as switching to an alternative model.

However, there are areas in need of further work in the state and county-level models. These are:

(1) Improved modeling of sampling variances for the county model (not proportional to the inverse of the CPS sample size), which was identified through the use of regression diagnostics,

(2) Improved estimation of model and sampling error variance in the state model to avoid estimation of zero model error variance,

(3) Inclusion of random or fixed state effects in the county-level model, or further integration of these two models. This last point was a partial reaction to the raking factors, but was also based on a perception that since one data set was driving the estimates, a more integrated approach would likely be more efficient,

(4) Use of discrete variable models (e.g., Poisson regression) to eliminate the omission of counties with zero poor 5-17 due to the log transformation. (This raises issues involving generalized linear mixed models in a sampling context, which is an area of current research; some diagnostics work indicated that this improvement may not be very important),

(5) Further examination of patterns of residuals, especially in the West Region, and areas with a large percentage of Hispanic families.

For the school district-level model, the Census Bureau should:

(1) Try to further reduce the variance of the long form estimates through modeling using short form data,

(2) Look into the use of school enrollment data and free and reduced price lunch data,

(3) (Obviously) look into the possible use of data from the American Community Survey, and

(4) Look into the geocoding of food stamp and IRS data at lower levels of aggregation to provide additional information at the school district level.

The Census Bureau is in the process of carrying out a lot of very innovative work by Bill Bell, Robin Fisher, and Jana Asher to address most of these issues. They should be congratulated for their excellent work.

References

Carroll, Raymond J., and Ruppert, David (1988), *Transformation and Weighting in Regression*, Chapman and Hall, New York.

National Research Council(2000), *Small Area Income and Poverty Estimates: Priorities for 2000 and Beyond*, Panel on Estimates of Poverty for Small Geographic Areas, Constance Citro and Graham Kalton, eds., Committee on National Statistics, Washington, DC.: National Academy Press.

Voss, P.R., R. Gibson, and K. Morgen (1997), "Assessment of the Initial 1993 Estimates of Small Area Children in Poverty." Paper presented to the Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics, National Research Council. Department of Rural Sociology, University of Wisconsin-Madison.