

## A SMALL-AREA VIEW OF POVERTY FROM THE PANEL

Alan M. Zaslavsky, Harvard University  
Department of Health Care Policy, 180 Longwood Ave., Boston, MA 02115-5899

**Keywords:** small-area estimation, empirical Bayes, Current Population Survey

The three papers presented today represent the current state of a vastly complex and challenging project. I address this project as a member of the Panel on Estimates of Poverty for Small Geographic Areas of the National Academy of Sciences (NAS). Although I will summarize the panel's preliminary conclusions, for an official statement please refer to the interim report (Citro, Cohen, Kalton and West 1997), available from National Academy Press and at the NAS Web site (<http://www.nap.edu/readingroom/> — select **Behavioral and Social Sciences** and then **Small area estimates of poverty**). Also, as this paper appears our next report should also be out, with updated recommendations regarding estimates to be released by the Census Bureau in October, 1997. This paper represents my own views and not necessarily those of the panel or the NAS.

### 1 View from the panel

**History of the project** The current small-area estimates program stems from 1994 amendments to the Elementary and Secondary Education Act. Until this year, ESEA Title I funds for school districts with poor children have been distributed to states based on poverty rates and numbers of children in poverty calculated from the most recent census. Consequently, at the end of the decennial cycle, allocations could be based on incomes from as much as twelve years before.

The 1994 amendments called for development by the Census Bureau of updated estimates for states and counties which could be fed into the complex allocation formula to yield state and county funding levels, to be updated every other year throughout the intercensal period. The legislation also called for establishment of a NAS panel to evaluate the acceptability of the estimates as a basis for fund allocation. The panel would make a recommendation to the Secretaries of Commerce and Education, who would then decide whether to use the updated estimates or continue to allocate funds on the basis of 1990 census estimates.

The Census Bureau's research program was well underway by 1995, and the panel was formed in June, 1996. Estimates were developed beginning in September, 1996 and first released in January, 1997, with advance release of the Panel's report in

February. In March, 1997, the Secretaries adopted the panel's recommendation to base allocations on a compromise between the old (1990 census) and updated poverty estimates.

The panel has had an unusual role in this process. Most NAS panels are charged to study a policy-relevant issue in general terms, but in this case the panel's recommendation was recognized by statute. In my opinion, the collaboration between the Census Bureau and the panel has been very successful, with intense interchange on technical and policy issues of the program. Furthermore, despite the implications of the work for funds distribution, there has not been political interference from either the legislative or executive branch, except for the sense of urgency expressed by all parties for moving forward to the new process.

**Interim findings:** The panel's findings included a number of points in favor of the revised estimates. First, there is strong evidence that there have been substantial changes in the national distribution of poverty between 1989 (the reference year of income data from the 1990 census) and 1993 (the reference year of the 1994 Current Population Survey). Second, model-based procedures are essential to the production of estimates that are more accurate for 1993 than carrying forward the census-based estimates. None of the survey data sources available (in particular, the Current Population Survey) are capable of supporting direct estimates of poverty with adequate precision for any but the largest states, much less for counties. Third, the panel generally evaluated the methods used for modeling poverty as being in line with the current state of the art for small-area estimation, although some important issues about the methodology are still to be resolved, especially for the county-level model.

Finally, the panel found that the Census Bureau's model-based estimates for 1993 probably improve on the 1989 estimates (based on the 1990 census) as estimates of poverty distribution in 1993. Neither of the two forms of evidence for this belief gives an entirely conclusive answer. One uses model-based estimates of error, although these are to some extent dependent on the accuracy of the model specification and indirect estimates of CPS sampling variance. The other evaluation compares the model and the carried-forward (1980) census, treating the census in a later year (1990) as the "gold standard." Aggregate measures of error were much smaller for

the 1990 CPS-based estimates than for estimates based on various naïve models, i.e. simple updates of the 1980 census. Both the state and county components of the model contributed to accuracy. Unfortunately, 1990 is the only year for which this comparison could be implemented, so it is difficult to generalize to other years for refined comparisons among closely competitive models.

Despite these generally positive conclusions, the panel also had several concerns about moving to a set of fully model-based estimates. These included the following:

1. There appeared to be some differences between the census and CPS measures of poverty, although it was difficult to tell whether they were systematically related to features of the counties. This is unsurprising, because the CPS income questions are much more detailed than those on the census long form. Nonetheless, we were cautious about implicitly moving to a new poverty definition, even one that might prove to be superior. (Another NAS report (Citro and Michael 1995) deals at length with poverty measures.)
2. We were also concerned about possible biases due to the model specification, including biases related to size which might occur as artifacts of the nonlinearities of the county model specification.
3. We were dissatisfied with the exclusion of sampling zeros from the data set.
4. The discrepancies between predictions for states from the state model, and those obtained by aggregating predictions from the county model (before raking) up to states, were substantial. In other words, the raking factors for adjusting the county model to the state model estimates were quite variable. This does not necessarily indicate that the combined state-county procedure gives inaccurate results but suggests some lack of fit in the county model.
5. The lack of direct variance estimates for CPS estimates in small areas made it necessary to estimate CPS variances in an indirect and highly model-dependent manner.
6. The quality of the postcensal estimates, required for calculating the rates used in the allocation formulae, was unknown.
7. Because of unavoidable delays in the production of data required for the models, allocations for the school year beginning in 1997 would be based on estimates for the 1993 income year.

Because of these uncertainties, the panel did not recommend unmodified use of the model-based es-

timates for 1997. Instead, it proposed averaging estimates of poverty rates from the 1990 census and the 1993 CPS-based model, and multiplying these by 1993 population estimates for counties to obtain estimated poverty counts by county. The panel did not claim to have devised a new and superior estimation method, but rather that uncertainties about the new methods were great enough that it was desirable to moderate the impact of the shift. This recommendation was accepted by the Secretaries of Commerce and Education, and became the basis for Title I funds allocations for the current school year (1997-1998).

The panel also emphasized the need for continued research to validate and improve the models. Since the report was issued, the Census Bureau has prepared, and both Census Bureau staff and the panel have studied, diagnostics for model fit and comparisons among direct CPS estimates, model-based estimates, and census-based estimates. Research has continued on evaluation of postcensal population estimates and estimation of CPS sampling variances. Alternative model specifications have also been considered, including models for rates or log rates by county, a bivariate model for census- and CPS-year poverty rates, an integrated state-county model (county model with state effects), and a generalized linear model (capable of handling sampling zeros in the CPS).

I would add that over the course of several years, enough data will accumulate to support more refined evaluations of possible systematic biases in the models and to distinguish these from accidental discrepancies that are consequences of patterns in poverty distribution that vary each year.

#### **School district estimates: the Final Frontier**

Finally, the Census Bureau and the panel have begun to confront the challenges involved in calculating estimates for school districts, which the legislation anticipates phasing in later in the decade. In all respects, this is far more challenging than the county-level estimation problem. There are many more school districts (> 16,000) than counties (about 3000), and many of them are extremely small. Even defining the districts is difficult, because they sometimes overlap, often change their boundaries, and in many cases have school-age populations that are quite different from their school enrollments. Therefore, geographic issues become crucial at this level. (The Census Bureau is conducting a major effort to update maps of school districts.)

Because of the paucity of data sources for extremely small geographic units, the methodology for county-level estimates is unlikely to translate di-

rectly and uniformly to school districts. (The new American Community Survey may be help, however.) Creating school district estimates of poverty will require another unprecedented effort.

## 2 Discussion of the papers

**Fisher and Siegel:** For county-level estimation, a normal homoscedastic linear model is assumed for underlying (logged) poverty counts, with predictors that are also (logged) counts. A simple relationship (variance inversely proportional to sample size, with a single proportionality constant) is assumed to describe CPS sampling variances. The constant in the sampling variance function is estimated indirectly, assuming constancy of model error variances across years. Counties with sampling zeros, which cannot be accommodated in the loglinear model, are omitted from the dataset.

Many specification decisions are embedded in this approach, and alternatives can be explored for many of them, e.g. nonconstant model variances, modeling of rates rather than counts, and more complex variance functions. Perhaps the loglinear model is both too simple and too hard to understand: it does not use current technologies for nonlinear modeling with random effects, and it is hard to explain the aggregate behavior of the model when the model assumptions (homoscedasticity and linearity on the logged scale) are not satisfied.

I would like to see this research move toward a generalized linear model framework, such as a quasi-logistic model for poverty rates. Random effects logistic models are routinely fitted by educational and health services researchers using commercially available software such as MLn and HLM. The design of the CPS complicates the application of such models. Nonetheless, a more appropriate regression structure with an approximation to survey design effects might give more sensible results than a simplified regression structure with elaborate estimates of variances and covariances under the design (which have not been developed yet for this application, anyway).

**Fay and Train:** This elegant, careful work was by and large accepted by the panel. The authors have also investigated a number of promising alternatives not represented in this paper, such as multivariate modeling of poverty in several age groups.

Sampling variances were related to rates by a relationship proportional to the  $p(1-p)$  relationship for binomial data. This is roughly equivalent to fitting a generalized linear model with binomial likelihood and a design effect. This is a big step in the direction of the quasi-logistic model suggested above,

and a creative method for bridging the gap between design- and model-based analyses. I believe that the full benefits of this method will be obtained when it is applied to modeling smaller domains.

**Bell and Otto:** This technically ambitious venture focuses on one of several possible directions for model expansion, namely multivariate modeling of several years of CPS data. Other directions that can be approached with similar technologies include a measurement error model for census year data (now being researched by Bell), use of additional covariates measured with error, multivariate modeling of several “outcomes” (e.g. poverty in several age groups), and more complex relationships among counties (e.g. spatial modeling).

To make these methods work well, more precise sampling variance-covariance estimates are needed. I hope that the authors will investigate estimation methods making more use of the full CPS design, including the structure of rotation groups and segments (Dempster and Hwang 1993). Rough bounds might be obtained for the amount of additional information contributed by each year of CPS data.

It might be very difficult to choose conclusively among the various possible specifications of the autocorrelation model in the time series approach, using these short noisy series. Even the assumption of stationarity is suspect, as autocorrelations are partly driven by irregular short-term economic trends. However, it may be that the consequences of the various models for prediction are not very different. For example, the random walk model is conservative in its use of past data although it is not plausible as a model for long-term dynamics.

**Conclusion:** The authors are again to be congratulated for their progress. We can expect this program to contribute both to statistical methodology in general and to the ability of government statistics to make use of sophisticated and dynamic methods.

## References

- Citro, C. F., Cohen, M. L., Kalton, G. and West, K. K., eds. (1997), *Small-Area Estimates of School-Age Children in Poverty: Interim Report I: Evaluation of 1993 County Estimates for Title I Allocations*, Washington: National Academy Press.
- Citro, C. F. and Michael, Robert T., eds. (1995), *Measuring Poverty: A New Approach*, Washington: National Academy Press.
- Dempster, A. P. and Hwang, J.-S. (1993), “Component models and Bayesian technology for estimation of state employment and unemployment rates,” *Proceedings of the Bureau of the Census Annual Research Conference*, 571-581.