

DEVELOPING POSTCENSAL INCOME AND POVERTY ESTIMATES FOR ALL US COUNTIES

Paul M. Siegel, U.S. Bureau of the Census
HHES, Census Bureau, Washington, D.C. 20233

Key Words: Income, Poverty, Small Area Estimates, Postcensal Estimates

¹ Recently, the Census Bureau and a consortium of other Federal Agencies -- Food and Nutrition Service (USDA), National Center for Education Statistics (DoEd), Head Start (HHS), HUD, and the Employment and Training Administration (DoL) -- initiated a project to provide post-censal estimates of income and poverty for small areas during the 1990's. While current plans for the 2000 census include a system of "continuous measurement" which will periodically furnish Census-level measurements and estimates for small areas, in this decade, more traditional estimates are likely to be the only ones available.

Motivation for this project rests in the use of income and poverty data for small areas in the administration of Federal programs. In fiscal 1994 more than \$30 billion in Federal funds were allocated to States or local jurisdictions on the basis of 1990 Census data on income or poverty. State and local programs also make extensive use of these data, and their concerns cannot be overlooked. Absent new measurements, the same data would continue to be used throughout the decade. Both legislators and administrators recognized that these estimates were less and less adequate representations of current conditions as time passed. Some Federal Laws have been rewritten to require the use of "updated" Census estimates, when they become available and are judged of sufficient quality.

Local interest in Federal estimates should be clarified. Besides lower costs, there are local benefits to the provision of a single uniform set of consistent estimates for all states and counties. On the other hand, the benefits of national consistency need to be weighed against the

¹This paper reports the results of research undertaken by the author, an employee of the Census Bureau. The views expressed are attributable to the author, and not the Census Bureau. John Coder was listed as second author on earlier versions of this paper. His name is absent from this draft for purely administrative reasons. It has benefitted immeasurably from his collaboration and it shamelessly borrows from his background paper, "The Small Area Income and Poverty Estimates Program for the 1990's," presented at the annual meeting of the Association of Public Data Users, October, 1994.

possibility of increased accuracy in some but not all states, which might be attained by using "better" local estimates where available. I will not undertake that discussion here.

This paper reports work in progress to develop a system to make these estimates for the counties of the US. I will discuss first some considerations which shape the properties I want our estimators to have and which limit the kinds of data I consider employing. Then I will turn to a discussion of properties of the datasets available, and the problems and directions they offer. Finally, I'll be a little more specific about two kinds of estimators I'm considering.

Scope and goals of the estimation project:

The present project is more ambitious than that undertaken by the Census Bureau for the General Revenue Sharing Program during 1972-87. The revenue-sharing funds allocation formula required estimates of per capita income for all governmental units. While per capita income estimates are useful for some purposes, other applications require distributional measures, such as median household income, or poverty measures for specific subpopulations. In order to meet the diverse needs for income and poverty data, we plan to provide estimates of six key statistics:

Median household income,
Per capita income,
Number of poor persons and poverty rate for four groups --
the total population,
children age five to 17,
children under age five, and persons age 65 and over.

We will produce estimates of income and poverty for counties biennially with the initial estimates for calendar (income) year 1993 to be released in late 1996. There is great interest in extending these estimates to such sub-county areas as school districts. ² That possibility will have to be examined after the county estimates have been evaluated.

It is vitally important to provide interpretable uncertainty bounds with whatever estimates are produced. Such bounds are useful for evaluating competing estimates and estimators, and for combining alternative estimates on

²The "Improving America's Schools Act of 1994" (P.L.103-382) includes an expression of this interest.

the basis of their strength. A major thrust of this project is to develop estimates of at least the sampling error associated with the estimates we provide.

It is not clear to me that users are able to decide or discuss where the line between acceptable and unacceptable uncertainty lies, but they must be given the materials. Table 1 shows "margins of error" for 1990 Census county estimates of the number of children in poverty. I invite the reader to try to describe in lay terms the estimates of poverty in the 500 counties for which the margin of error based on the 90% confidence interval is 25% or more. The fact that there are 500 such counties indicates that the Census does not set an unattainable standard. The mere existence of

confidence intervals about the Census estimates suggests some caution will be required in using them to evaluate the postcensal estimates. It also indicates that with any estimates, even estimates as precise as the census, change is not easily measured for some counties. Finally, the table shows that more than three quarters of the poor children live in the one quarter of counties where poverty was well measured by the Census, and where the postcensal estimates are likely to be quite good.

Conceptual constraints on estimation strategies:

The conceptual difference between poverty and income imposes significant restrictions on the data and models that can be employed in making these estimates.

Table 1. -- Distribution of Relative Margins of Error for Estimated Numbers of Poor Children Age 5 to 17 Years Old, by County: Data from the 1990 Decennial Census of Population and Housing.

Margin of Error	Counties		Poor Children	
	Number	Percent	Number	Percent
All Counties	3,141	100.0	7,544,737	100.0
Less than 5 percent	112	3.6	3,426,266	45.4
5 to 10 percent	630	20.1	2,283,882	30.3
10 to 15 percent	861	27.4	1,153,663	15.3
15 to 20 percent	616	19.6	410,362	5.4
20 to 25 percent	412	13.1	168,203	2.2
25 to 50 percent	468	14.9	100,013	1.3
50 to 75 percent	33	1.0	2,121	(z)
75 to 100 percent	6	(z)	206	(z)
100 percent or more	3	(z)	21	(z)

(z) Less than 0.5 percent.

Because poverty is the most complex, I will focus on it in this paper.

While one might approach the estimation of population by means of a system of demographic accounts in which estimated numbers of births and in-migrants were added to the population and deaths and out-migrants were subtracted from it, maintaining demographic accounts of the poverty population seems hopeless. The measurement of poverty requires the definition, specification, and measure-

ment of both a domestic group and its relevant income. The current approach to assessing poverty compares the combined income of the family members resident in a household to a standard of income need which depends upon family size and age composition. This means that even abstracting from the considerable problem of measuring income, income accounts do not provide poverty accounts. Estimating the joint distribution of families by size, age-composition, and income for small areas (counties) would appear to require data not currently available. The lack of such data

precludes employing an accounting approach to estimation, and inclines me toward symptomatic regression estimation techniques. That is, I am inclined toward techniques which base estimates on indicator variables which are measured for all counties at the time for which I want estimates. The problem then is characterizing the relation between indicator data from administrative sources, and poverty.

I assume a set of state-level estimates which borrow strength from data and methods unavailable or inapplicable at the county level. Thus the county estimators will serve to distribute these superior estimates of state totals, and their sum will be constrained to equal them. There appears to be a rhetorical advantage to applying uniform methods and data from the same source(s) to estimation for areas within all states. It confers a kind of equity. Endowing our estimators with this hierarchical character, that is, distributing state totals, has the additional advantage that, should superior methods or data be applicable within a single state, the county estimates for that state could be improved without affecting estimates within other states. This would seem to create the possibility for states to participate in improving the Federal estimates. Peculiarities of the design of programs or allocation formulae could negate this advantage, however. Imagine an allocation formula in which the concentration of poor persons in a single area of a particular state increased the total of funds allocated to that state over what they would be if the same number of poor persons were spread over several areas of that state. Under such a formula, states could be tempted to manipulate their estimates of the distribution of the poor within their state.

Available data:

Indicators of poverty which are available at the county level have two kinds of weaknesses. On the one hand, they reflect concepts of poverty which differ from the Census definition, and on the other hand, they cannot be retabulated for sets of geographic units which do not respect county boundaries. We have found one set of administrative records which comes close to meeting the second objection, and may not be outrageously poor at meeting the first.

Federal Income Tax returns permit us to construct a family group whose income can be compared to the poverty standard. These data can be tabulated for any geographic units for which a correspondence to Census geography has been established. We tabulate individual tax returns assigned to counties according to the addresses from which they are filed, tallying the total number of persons on the return (filers and dependents), as well as the number of young persons and old persons in households above and below the poverty threshold as we do so. We are not matching individual tax returns to any other individual records.

One advantage of this dataset is its geographic flexibility. If estimates are desired for other systems of areas, e.g., school districts, we anticipate that this file can be tabulated to them. This presumes that the boundaries of the target areas have been mapped to census geography and that the data can be geocoded. Data available to us as county totals, like the number of food stamp recipients, are useful only in making estimates for areas which can be approximated by counties or sets of counties.

- There are lots of things wrong with tax return data:
- The tax filing unit -- the tax form -- does not provide a close fit to the family. Many "families" are represented by more than one tax return, and only some of these can be discerned in the tax data.
 - There are persons represented on tax returns who are not in the universe for which poverty is defined -- members of the Armed Forces, and institutionalized persons, for instance.
 - For at least some persons, the tax filing address is not in the county to which the persons on the form would be attributed by the Census.
 - Other families and individuals, some of whom are particularly likely to be poor, are not represented in the tax return data at all.

It is somewhat reassuring, on this score, that the difference between the number of persons in the Census and the number implied by the tax returns for 1989 is highly correlated, over all counties, with both the number of persons on tax returns in poverty and the number of persons in poverty in the Census. This suggests an important role for the new county population estimates to play in our poverty estimates.

The tax data are particularly vulnerable to change over time as a result of policy change: change the tax code and you change what the tax returns measure and how that is related to poverty. There were substantial changes in the early to mid 1980's. This should caution our use of these data.

The Census Bureau will soon launch a program of postcensal county-level estimates of **resident population**. In order to be consistent with them, the income and poverty estimates must . . .

- use these population estimates as denominators of the estimates of per capita income;
- use these population estimates as the denominators for state and local poverty rates; and
- make sure that the estimated number of persons below the poverty threshold is no larger than the estimated population.

We will, of course, use the new population estimates, and

there is a direct role for these population estimates in the estimation of persons in poverty.

Direct estimates of poverty:

The Current Population Survey (CPS) offers direct estimates of poverty and income statistics for the set of counties and minor civil divisions which comprise its in-sample primary sampling units. The annual March income supplement to the CPS provides timely estimates of poverty, but from the standpoint of estimates for counties, they suffer several severe sampling weaknesses:

- PSU sample sizes are small -- the median number of interviewed households per PSU in a given month is 47³ -- and thus the sampling error of the poverty estimates for counties is large.
- By grouping counties into PSU's the sample makes between-county variance, which we want our estimates to capture, into within-PSU variance, rendering our estimates imprecise.

The first problem can be addressed by averaging results from separate surveys. The detailed income questions upon which CPS poverty statistics depend are only asked in the March survey. About half the households in two successive March surveys are the same. Thus, the average of three successive March surveys would effectively double the sample size, at the cost of making the time reference rather broad. Alternatively, we can approximately triple the sample size for each PSU's estimate by averaging reports for the preceding November and the following July with the March estimates for a given year, since the samples interviewed in those months are different households and thus independent, within PSU. But this approach requires us to resolve a means of combining family income data from the March income survey, which has a detailed income questionnaire, with the one-question family income report available for other months. cursory comparison of the detailed reports of family income made in the March survey with the one-question reports made by the same households shows, not surprisingly, that more income (and thus less poverty) is reported by both families and primary individuals using the March instrument than using the one-question instrument. About 2% more primary families and individuals report their income below the approximate poverty threshold in the one-question inquiry than in the detailed

³Based on tabulations of 655 PSU's in the CPS of March, 1990. The New England States and Hawaii, where PSU's are defined in terms of minor civil divisions, have been excluded.

March supplement⁴. More thorough investigation of the relation between these two measures is clearly required, but the problems seem soluble.

The PSU/County problem seems less tractable. The 1980's CPS sample is not designed for county estimates. It is especially weak for small counties. I may be able to get some mileage out of disaggregating PSU's into their constituent counties, and then selectively re-aggregating the small counties into pseudo-PSU's. I'm not sure how to evaluate the results of such an effort, and in any case, they will not completely evade the problem. Using the CPS in a program of county estimates entails using observations which are larger and more heterogeneous than counties.

Current Population Survey and decennial Census estimates of the number of persons below the poverty threshold do not have the same expected value. In part the differences are due to differences in the universes for which the two surveys define poverty⁵ and in part the differences

⁴These figures are percentages of unweighted unit counts shown in an unpublished Census Bureau Memorandum to Chuck Nelson, Acting Chief, Income Statistics Branch, Population Division, from Ed Welniak, "Comparison of March 1985 Current Population Survey form CPS-665 Supplement Income Data to Control Card Income Data" (May 7, 1986).

⁵The decennial Census excludes residents of college dormitories from the universe for whom poverty is defined. (There were about 2 million residents of college dormitories in the 1990 census.) The Current Population Survey probably represents most residents of college dormitories as if they were resident in their parental households. The effect of this difference in definition would likely be difficult to discern in data from the two sources and would be dispersed over the areas of residence of the parents of college students (likely increasing their measured poverty in the CPS relative to the Census). The decennial Census does not recognize groups of persons related to one another but not related to the householder of the housing unit they inhabit as a family, treating them instead as unrelated individuals. It then excludes unrelated persons under age 15 from the universe for which poverty is defined. The CPS estimates there are about 1.8 million persons in .7 million (unrelated sub-) families. About .8 million are under age 15. About half of the remaining million persons are below the poverty threshold in CPS, where their subfamily is recognized as a unit for the purposes of gauging poverty. Most of these same persons are probably classed as unrelated individuals below the poverty line in the Census.

are due to the more thorough assessment of income in the current population survey. The 1970 and 1980 Censuses showed higher numbers and proportions of persons in poverty than the CPS estimate, which might be expected if more income is reported in the CPS inquiry⁶. But when the 1990 CPS is adjusted to population controls based on the 1990 Census, the number of persons in poverty and the poverty rate are indistinguishable from the Census estimates⁷. Eller reports that the “. . . ratios of estimated state poverty rates from the [March 1990] CPS to those from the [1990] Census varied from a low of .41 (in Connecticut) to a high of 1.38 (in Hawaii). . . .” She reports the correlation between CPS and Census poverty rates as (only) .915.⁸ The bias of Census relative to CPS measurements of poverty (or vice versa) must be investigated and taken into account at some point in the estimation program. I do not want the difference between estimates for 1993 and 1990 to confound definitional differences and real change. This concern will have its greatest impact on the state-level estimates.

Postcensal sub-state income and poverty estimation strategies

The strategies we intend to pursue are determined by the available data and their qualities, and by the properties we have chosen to impose on the estimates. On the data side we have:

- Direct estimates of income and poverty from the 1990 (and 1980) Census for **all** areas.
- Direct estimates of income and poverty from the annual March income supplement to the CPS for a sample of areas. Most of these estimates are too unreliable to report individually. But they can serve well to provide observations on the dependent variables of interest in regression equations
- Administrative data offering annual measures of indicator variables for **all** areas.

We choose to require that;

- The estimates distribute independently estimated state totals.
- The estimates be particularly sensitive to change over time, since concern about the lack of timeliness of the Census data motivates their production.

The problem is to build an estimator of postcensal poverty for all counties from these materials.

One standard approach to this problem is to estimate postcensal change in poverty from postcensal change in the indicators by assuming that the relation between change in the two series is the same in the postcensal period as it was in the preceding intercensal period. Because the period from 1980 to 1990 witnessed massive changes in the tax code, I am wary of assuming that the relation between 1980 to 1990 intercensal change in poverty and change in the tax data is the same as the postcensal relation between changes in tax and poverty data after 1990. Thus my reluctance to estimate, say, 1990 to 1993 change in poverty from tax data for the same period and the intercensal relation between tax data and poverty.

A more attractive estimator takes as the dependent variable the ratio of the direct CPS estimate of a place's share of, say, children in poverty in a state at the time of interest to its share at the Census. Regressing that share change ratio on a similar ratio of one or more auxiliary variables, e.g., the ratio of county shares of children in poverty on tax returns at the time of interest and the census, provides a "ratio regression model." Unlike standard ratio regression models, this one does not use relations from a different time period than the one of interest. Ericksen⁹ explored the properties of just such estimators, and showed that their main advantage over ratio correlation methods was in the reduction of large errors, along with their

⁶See William P. O'Hare, "Assessing post-census state poverty estimates," **Population Research and Policy Review**, 12:261-275 (1993).

⁷The 1990 Census estimate is 31.743 million persons in poverty, and the 90% confidence interval estimate from CPS estimate, based on the revised population controlled is 31.562 -- 33.268 million. Both sources estimate the poverty rate to be 13.1 percent of persons for whom poverty is defined. See U.S. Bureau of the Census, **Current Population Reports**, Series P-60-188, "Income, Poverty, and Valuation of Noncash Benefits: 1993." U.S. Government Printing Office, Washington, D.C. 1995. The revised estimates reflect the 1990 Census population controls unadjusted for estimated Census undercount.

⁸T.J. Eller, "State poverty rates and labor markets: 1980--1990," presented at the annual meetings of the American Statistical Association, Boston, 1992. Some caution is needed in interpreting Eller's results since the CPS estimates available to her employed weights based on the population as measured in the 1980 Census carried forward to 1990. Thus differences between the 1990 CPS and Census results in part reflect differences between the population estimates and the 1990 Census.

⁹Eugene P. Ericksen, "A regression method for estimating population changes of local areas," **Journal of the American Statistical Association**, 69:867-75 (1974).

obvious advantage of not depending upon relations from prior intercensal periods. County estimates could be composites of direct and regression estimates for counties in the CPS sample and the regression estimates elsewhere, though the direct estimates would probably not add much. A rhetorical advantage of this estimator is that it incorporates the Census estimates and models change from that (firm) base.

A second estimator relies entirely on cross-sectional relations. This cross-sectional estimator has the advantage that it can be used to make estimates in any year for which CPS and tax data are available. More appealing, it should be less disastrously affected by changes in the tax laws.

A particularly attractive aspect of this estimator is the extent to which it permits us to examine its properties. Estimates can be prepared for 1989 using the Census in place of the CPS and regressions over all counties. These can be compared to the Census estimates to gauge the adequacy of the indicators. A second set of estimates can be prepared for 1989 using the Census data in place of the CPS, but using only the CPS PSU's in the regression. The comparison of these estimates with the previous set affords

a clear indication of the impact of the CPS sample⁹⁹⁹. Finally, estimates for 1989 can be prepared using the March 1990 CPS estimates in the regression. These will permit assessing the impact of definitional differences and of within-PSU sampling in CPS.

This discussion has focussed on the measures of poverty. Clearly the per capita income estimates require distributing estimates of the state's total aggregate income and dividing by the independent population estimates. That seems relatively straightforward¹⁰. Approaches to estimating median family income deserve more thought than they have received to date.

¹⁰For the current BEA approach see Wallace Bailey, Linnea Hazen, and Daniel Zabronsky, "State, metropolitan area, and county income estimation," Chap IV in Federal Committee on Statistical Methodology, **Statistical Policy Working Paper 21: Indirect Estimators in Federal Programs**. Washington, D.C., Office of Management and Budget (1993)