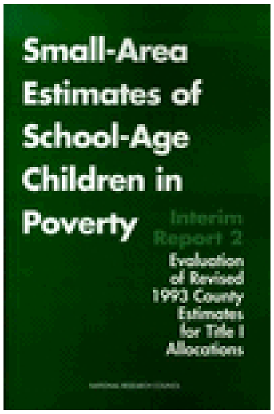


**Small-Area Estimates of School-Age Children in Poverty: Interim Report 2, Evaluation of Revised 1993 County Estimates for Title I Allocations**

Committee on National Statistics, National Research Council

ISBN: 0-309-52203-X, 186 pages, 6 x 9, (1998)

# Small-Area Estimates of School-Age Children in Poverty

## Interim Report 2: Evaluation of Revised 1993 County Estimates for Title I Allocations

Constance F. Citro, Michael L. Cohen, and Graham Kalton, *Editors*

Panel on Estimates of Poverty for Small Geographic Areas

Committee on National Statistics

Commission on Behavioral and Social Sciences and Education

National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C.    1998

The project that is the subject of this report is supported by Contract RN96131001 between the National Academy of Sciences and the U.S. Department of Education. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

Copies of this report are available for sale from the National Academy Press, 2101 Constitution Avenue, N.W., Box 285, Washington, D.C. 20055. Call 800-624-6242 or 202-334-3313 (in the Washington Metropolitan Area). This report is also available on line at htt;?//www.nap.edu.

International Standard Book Number 0-3090-6045-1

Printed in the United States of America.

## PANEL ON ESTIMATES OF POVERTY FOR SMALL GEOGRAPHIC AREAS

GRAHAM KALTON (*Chair*), Westat, Rockville, Maryland
DAVID M. BETSON, Department of Economics, University of Notre Dame
NANCY E. DUNTON, Midwest Research Institute, Kansas City, Missouri
WAYNE A. FULLER, Department of Statistics, Iowa State University
THOMAS B. JABINE, Consultant, Washington, D.C.
SYLVIA T. JOHNSON, School of Education, Howard University
THOMAS A. LOUIS, School of Public Health, University of Minnesota
SALLY C. MORTON, RAND, Santa Monica, California
JEFFREY S. PASSEL, Urban Institute, Washington, D.C.
J.N.K. RAO, Department of Mathematics and Statistics, Carleton University
ALLEN L. SCHIRM, Mathematica Policy Research, Inc., Washington, D.C.
PAUL R. VOSS, Department of Rural Sociology, University of Wisconsin
JAMES H. WYCKOFF, Graduate School of Public Affairs, State University of New York, Albany
ALAN M. ZASLAVSKY, Department of Health Care Policy, Harvard Medical School

CONSTANCE F. CITRO, *Study Director*
MICHAEL L. COHEN, *Senior Staff Officer*
KIRSTEN K. WEST, *Research Associate*
MEYER ZITTER, *Consultant*
CANDICE S. EVANS, *Senior Project Assistant*

## COMMITTEE ON NATIONAL STATISTICS
## 1997-1998

# Contents

# Tables and Figures

TABLES

FIGURES

# Acknowledgments

The Panel on Estimates of Poverty for Small Geographic Areas wishes to thank the many people who contributed to the preparation of this, the panel's second interim report.

We wish to thank particularly the staff of the Census Bureau who worked with us in June-October 1997 to evaluate the Census Bureau's original model and alternative models for producing county estimates of poor school-age children in 1993: William Bell, John Coder, Robert Fay, Robin Fisher, Matt Kramer, Mark Otto, Paul Siegel, and George Train. We also thank the Census Bureau staff who facilitated the arrangements for the work: Cynthia Clark, Nancy Gordon, Charles Nelson, and Daniel Weinberg. Members of the Census Bureau's Population Division, including Bashir Ahmed, Michael Batutis, Sam Davis, John Long, Barbara Martin, Esther Miller, Douglas Sater, Larry Sink, Signe Wetrogan, and Janet Wysocki, provided informative briefings and evaluation materials on the county population estimates of school-age children.

Daniel Kasprzyk of the National Center for Education Statistics, who serves as the Department of Education project officer for the study, was helpful as always in facilitating this phase of its work. The panel also appreciates the continued help of other Department of Education staff, in particular, Sandy Brown, Kay Rigling, William Sonnenberg, and Stephanie Stullich, in educating us about the Title I allocation process.

I thank my panel colleagues for their continued commitment to the work of the panel and to its second interim report, which was prepared, like the first, under a very demanding time schedule. We were assisted by a very able staff. Constance Citro, who became the panel's study director in spring 1997, had overall responsibility for the project. Without her excellent work in preparing and revising drafts of the report, we would not have been able to complete this report on time. Michael Cohen made important contributions to the evaluation of alternative models and to many sections of the report. Meyer Zitter, with the help of Kirsten West, ably assisted the panel's working group on the county population estimates and provided other materials for the panel. Candice Evans provided administrative support for our study and capably and cheerfully carried the brunt of the work to produce the report. Eugenia Grohman, associate director for reports of the Commission on Behavioral and Social Sciences and Education, contributed substantially to the readability of the report through her fine technical editing. To all we are grateful.

Graham Kalton, *Chair*
Panel on Estimates of Poverty for
Small Geographic Areas

# Small-Area Estimates of School-Age Children in Poverty

# Executive Summary

The U.S. Department of Education uses estimates of school-age children in poverty to allocate federal funds under Title I of the Elementary and Secondary Education Act for education programs to aid disadvantaged children. Until recently, the practice has been to base that allocation on the numbers and proportions of school-age children in poverty by county from the most recent decennial census. In 1994 Congress authorized the Bureau of the Census to provide updated estimates of poor school-age children every 2 years, first for counties and subsequently for school districts. Congress also authorized a study of the Census Bureau's program for producing these small-area poverty estimates. That study is being carried out by the Committee on National Statistics' Panel on Estimates of Poverty for Small Geographic Areas, which is charged to advise the Secretaries of Education and Commerce on the appropriateness and reliability of the Census Bureau's estimates for use in Title I allocations. This is the panel's second report.

The Census Bureau's procedure for producing updated county estimates of poor school-age children uses a statistical model that combines data from several sources, including the March Current Population Survey (CPS), income tax return records, food stamp program records, and county population estimates. The data sources used in the model are generally available only 1-2 years after the period to which they refer. For this reason, the Census Bureau decided that the most recent estimates it could produce by the end of 1996 (for the Title I allocations in spring 1997) were for poor school-age children in 1993: all the data for those estimates would not be available until the end of 1995, and model development work would require considerable time after that. The Census Bureau followed that plan and in January 1997 provided to the panel county estimates of the number of school-age children in 1994 who were living in and related to a family in poverty in 1993, which were intended to be used for Title I allocations for the 1997-1998 and 1998-1999 school years.

In its first interim report (National Research Council, 1997), the panel strongly supported a model-based approach for developing county estimates of poor school-age children. In selecting a model, however, the panel noted that it is important to conduct a thorough evaluation to assess the reasonableness of the model's assumptions, to examine the model's predictions to see that they contain no identifiable systematic errors, and to compare alternative models. Such an evaluation is a critical component of a model-based approach.

The panel concluded that the Census Bureau's original model was a substantial step toward the provision of more up-to-date county estimates of poor school-age children. However, there had not been time to complete a full evaluation of the model prior to release of the original 1993 estimates to the panel in January 1997. Therefore, the panel did not recommend sole use of those estimates to allocate funds under Title I. Instead, the panel recommended to the Secretaries of Education and Commerce that an average of the 1990 census estimates (which pertain to poverty status in 1989) and the Census Bureau's original model-based estimates of poor school-age children in 1993 be used to allocate Title I funds for the 1997-1998 school year.

In April 1997 the Department of Education allocated Title I funds for the 1997-1998 school year on the basis of the panel's recommended averaging procedure. The department subsequently requested that the panel and the Census Bureau carry out an in-depth evaluation of the Census Bureau's model and reasonable alternative models. It further requested that, on the basis of the evaluation findings, the Census Bureau develop a revised set of county estimates of the number of poor school-age children for 1993 and that the panel evaluate the appropriateness and reliability of the revised estimates for use in Title I allocations for the 1998-1999 school year. (There was neither time, nor a legislative requirement, for the Census Bureau to produce estimates for later than 1993.)

Between June and October 1997 the Census Bureau carried out extensive evaluations of its model and alternative models. On the basis of those evaluations, it revised the county model and prepared a revised set of 1993 county estimates of poor school-age children, which were provided to the panel in October 1997. The panel commends the Census Bureau for its work in developing a model-based approach for updated county-level estimates of school-age children in poverty. It also commends the Census Bureau's efforts, carried out in a short time period, to fully evaluate the original county model and alternatives to it.

The panel has undertaken a full assessment of the Census Bureau's work and the evaluation results. On the basis of its review, the panel makes the following recommendation for use of the revised 1993 estimates:

**The panel recommends to the Secretaries of Education and Commerce that the Census Bureau's revised 1993 county estimates of poor school-age children be used in the Title I allocations for the 1998-1999 school year.**

The revised estimates should not be averaged with estimates from the 1990 census, as was done for the allocations for the 1997-1998 school year.

The panel concluded that the Census Bureau's revised county model provides estimates of poor school-age children in 1993 that are demonstrably superior to estimates from the outdated 1990 census. The census estimates do not reflect the major changes in the distribution of poverty that occurred between 1989 and 1993.

The panel also concluded that the Census Bureau's revised county model performs as well as or better than reasonable alternative models that are practicable for implementation at this time. Performance was measured principally by examining the assumptions of the underlying regression models and comparing the predictions from the Census Bureau's model and alternative models for 1989 with 1990 census estimates of the numbers and proportions of poor school-age children for categories of counties. The changes that were made in the revised county model (principally, modifying one of the predictor variables in the regression equation) improved its performance relative to the original model.

The panel notes that the revised county model, like other models, has both strengths and weaknesses. Some level of error in model-based estimates (or in any estimates obtained from a census, from a survey, or indirectly from a model) is inevitable and is not a reason for rejecting such estimates. Yet the county model can very likely be improved with continuing research and development. It may also be possible to reduce the time lag of the estimates. In addition, the model will need to change to accommodate changes in the available data that occur in future years. Hence, the panel recommends that the Census Bureau continue research and development for further improving model-based county estimates of poor school-age children.

At the same time, the Bureau is required by law to develop updated estimates for school districts: this is a challenging task, given such factors as the small size of many school districts, the different ways in which districts are defined, changes in district boundaries over time, and the scarcity of relevant data for estimation. For developing updated estimates of poor school-age children for counties and school districts, as well as other small-area estimates of income and poverty, the Census Bureau will need to provide adequate staff and other resources to support a small-area estimation program on a continuing basis.

# 1

# Introduction

Small-area estimates of poverty for children aged 5-17 in families are used by the U.S. Department of Education to allocate funds under Title I of the Elementary and Secondary Education Act, which supports compensatory education programs to meet the needs of educationally disadvantaged children. At present, the department allocates Title I funds—over $7 billion for the 1997-1998 school year—to counties, and the states then distribute these funds among school districts within each county (see Moskowitz et al., 1993).

The county allocations are based on estimates of eligible children: predominantly, children aged 5-17 in families with incomes below the poverty level,[1] but also children in foster homes, children in families above the poverty level that receive Aid to Families with Dependent Children (AFDC),[2] and children in local institutions for neglected and delinquent children. The allocations depend primarily on the number of eligible children but also on the proportion of school-age children who are eligible. The allocations also take into consideration the state's average per-pupil expenditure, and the formula includes a hold-harmless provision to cushion the impact of decreases in allocations (for details of the allocation process, see National Research Council, 1997:App. A).

[1]The poverty status of individuals is determined by comparing the before-tax money income of their family to the appropriate poverty threshold. The poverty thresholds vary by family size and are updated by the change in the Consumer Price Index each year. See National Research Council (1995) for an evaluation of the current official poverty measure and proposed alternative measure; the issue of how poverty should be defined is not considered in this report.

[2]The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 abolished AFDC and replaced it with Temporary Assistance to Needy Families (TANF).

The poverty estimates for the Title I program used by the Department of Education to allocate funds to counties are provided by the Bureau of the Census. The practice until recently had been to use poverty estimates for a decade or more based on the most recent decennial census for which data were available. Since the proportions and numbers of children in poverty change over time, however, Congress in 1994 authorized the Bureau of the Census to provide updated estimates of poverty every 2 years for counties and, subsequently, for school districts for use in Title I allocations. Having the most up-to-date estimates is important so that resources can be directed towards areas that are most in need.[3]

Congress also authorized a study—through the Department of Education—by a panel of the National Research Council's Committee on National Statistics to review the Census Bureau's program for small-area poverty estimates. The statute requires that the Department of Education use the Census Bureau's updated estimates for the allocations unless the Secretaries of Education and Commerce determine that they are "inappropriate or unreliable" on the basis of the panel's study ("Improving America's Schools Act of 1994," P.L. 103-382, and 1996 continuing resolution).

The Panel on Estimates of Poverty for Small Geographic Areas was set up to carry out the authorized study. The panel is charged with a broad review of the Census Bureau's postcensal poverty estimates for small geographic areas and their utility for Title I allocations. The panel began its work in June 1996 and is scheduled to work through 1998, producing a final report at that time and such interim reports as are needed.

In January 1997 the Census Bureau provided to the panel the first set of updated estimates for counties of the numbers of children aged 5-17 in 1994 from families with incomes below the poverty level in 1993.[4] The original 1993 county estimates were developed from a statistical model that used administrative data from Internal Revenue Service and food stamp program records for 1993, estimates of poor school-age children in 1989 from the 1990 census, and 1994 population estimates to predict poverty for school-age children in 1993 as measured in the March Income Supplement to the Current Population Survey (CPS). To increase the reliability of the predictions, the model used a weighted

[3]See National Research Council (1997:Ch. 2; App. B) for data on the significant changes that occurred in the numbers and proportions of poor school-age children between the 1980 and 1990 censuses and following the 1990 census.

[4]More precisely, the Census Bureau's estimates pertain to related children aged 5-17 in poor families, termed "poor school-age children" in this report. Related children include all family members in a household who are under 18 years of age and related to the householder by birth, marriage, or adoption, except the spouse of the householder. Foster children are not included since they are not related to the householder, who is the person in whose name the house is owned or rented (see Bureau of the Census, 1993). The January estimates were released for public use on March 27, 1997, after a few corrections for erroneous input data were made.

average of 3 years of data from the March 1993, 1994, and 1995 CPS, covering income in 1992, 1993, and 1994.

The data used in the model are obtained from several sources, and most data are not available until 2 years after the period to which they refer. When the developmental work began in 1994, the Census Bureau decided that it could not expect to produce estimates by the end of 1996 for a later year than 1993, given the time required for acquiring, processing, and using the data in a new statistical model.

As required by the legislation, the panel's first interim report assessed the reliability and utility of the original 1993 estimates for use in Title I allocations for the 1997-1998 and 1998-1999 school years (National Research Council, 1997). The panel concluded that a model-based approach to producing updated small-area estimates is appropriate and necessary when it is not possible to obtain estimates from a single data source, such as a sample survey, that are sufficiently reliable for the intended use. The panel stated that model-based estimates of poor school-age children can be produced that are clearly preferable to estimates from the 1990 decennial census given the major changes that occurred in the numbers and distribution of poor children between 1989 and 1993. However, the panel concluded that there had not been sufficient time to thoroughly evaluate the Census Bureau's original model-based estimates for 1993. As an interim solution for Title I allocations for the 1997-1998 school year, the panel recommended that the model-based estimates be averaged with 1990 census estimates (see National Research Council, 1997:38). The panel suggested additional evaluations for the Census Bureau's county model and alternative models, many of which the Census Bureau had begun but had not had time to complete.

By the terms of the legislation, the estimates used for the 1997-1998 school year allocations could also be used for the 1998-1999 school year. However, the Department of Education wanted to pursue the possibility of a new allocation for 1998-1999 that would use only the 1993 estimates (i.e., not averaged with census estimates), and in May 1997 it requested the panel and the Census Bureau to further evaluate the original county model and alternative models. On the basis of the evaluation results, it asked the Census Bureau to produce revised 1993 estimates of poor school-age children by county in October 1997, and it asked the panel to assess the suitability of those estimates for use in allocating Title I funds for the 1998-1999 school year. (There was not time enough either to obtain the necessary data for the model or to conduct a full evaluation of the estimates for a later year than 1993.)

The Census Bureau completed the evaluation work and produced revised 1993 county estimates of poor school-age children, which were provided to the panel and the Department of Education in October 1997.[5] This, the panel's

[5]The revised estimates were made available on the Census Bureau's web site in January 1998: http://www.census.gov/hhes/www/saipe93.html.

second interim report, advises on the use of the revised estimates for Title I allocations for the 1998-1999 school year.

This report contains six chapters and four appendices. Chapter 2 describes the Census Bureau's procedure for obtaining revised county estimates of the numbers and proportions of poor school-age children in 1993; the procedure uses a county model, a separate state model, and county population estimates of total school-age children. Chapter 3 describes alternative county models that were evaluated, and Chapter 4 summarizes the evaluation results. Chapter 4 also comments on the state model and the population estimates. Chapter 5 provides the panel's assessment of the revised 1993 county estimates and its recommendation for 1998-1999 Title I allocations. Chapter 6 outlines research and development activities for further work on developing updated county estimates of poor school-age children. The appendices provide additional technical information on: alternative county models (Appendix A); county population estimates (Appendix B); internal evaluation of county model regression output (Appendix C); and comparisons of county model estimates with 1990 census estimates of poor school-age children for 1989 (Appendix D).

# 2

# Census Bureau Estimation Procedure

Reliance on the most recent decennial census to allocate federal funds to counties and other small areas has primarily reflected the absence of alternative data sources with comparable or superior reliability. Mindful of the need for small-area estimates that are more up to date than census estimates, the Census Bureau organized a program—Small Area Income and Poverty Estimates (SAIPE)—to study methods for producing postcensal income and poverty estimates for states and counties by using multiple data sources and innovative statistical methods. The Census Bureau launched this program in late 1993 with financial support from a consortium of five federal agencies. Congress made this work more urgent by charging the Census Bureau in late 1994 to produce updated estimates of poor school-age children for counties and school districts every 2 years to begin in 1996 with 1993 estimates for counties and in 1998 with 1995 estimates for school districts.

The program faces a challenging task. For Title I allocations, there is no single administrative or survey data source that provides all of the information required to develop reliable estimates of the number and proportion of school-age children in families in poverty by county or school district. The March Income Supplement to the CPS can provide reasonably reliable annual estimates of such population characteristics as the number and proportion of poor children at the national level and for some states. However, the CPS cannot provide estimates for the majority of counties because the sample does not include any households in them. And for almost all of the counties with households in the CPS sample (about 1,500 of a total of 3,143 counties in 1993), the estimates have a high degree of sampling variability.[1] Nonetheless, the CPS data can serve as the basis for creating usable estimates for counties through the application of statistical estimation techniques to develop "model-based" or "indirect" estimates.

Indirect or model-based estimators use data from several areas, time periods, or data sources (which could include the previous census) to "borrow strength" and improve precision. A model-based approach is useful when there is no single data source for the area and time period in question that can provide direct estimates that are sufficiently reliable for the intended purpose. Previously, the Census Bureau used this strategy to develop estimates of median family income for states (Fay et al., 1993) and, in part, to develop population estimates for states and counties (see Spencer and Lee, 1980).

This chapter describes the model-based approach as used by the Census Bureau to develop revised estimates by county of the number and proportion of school-age children in families in 1994 who were poor in 1993 (referred to as the revised 1993 estimates). The Census Bureau's estimation procedure for counties uses two regression models that predict poor school-age children—a county model (revised from the original model) and a separate state model—along with county population estimates. The steps in the procedure for the revised 1993 estimates include:

(1) Developing and applying the Census Bureau's revised county model to produce initial estimates of the number of poor school-age children. The county estimation process involves:

—obtaining data from administrative records and other sources that are available for all counties to use as predictor variables;

—specifying and estimating a regression equation that relates the predictor variables to a dependent variable, which is the estimated log number of poor school-age children from 3 years of the March CPS for counties with households in the CPS sample; and

—using the estimated regression coefficients from the equation and the predictor variables to develop estimates of poor school-age children for all counties. For counties with households in the CPS sample, the predictions from the model are then combined by a "shrinkage" procedure with the CPS estimates for those counties.

(2) Developing and applying the Census Bureau's state model to produce estimates of the number of poor school-age children by state. The state estima-

[1]For a description of the March CPS and differences between income and poverty data from the CPS and the 1990 census long-form sample, see National Research Council (1997:Ch. 2; App. B). The 1990 census sample includes households in all counties and covers 15 million households, 30 times more than the 50,000 households in the CPS; even the 1990 census estimates are highly variable for some small counties (National Research Council, 1997:Table 2-1).

tion process is similar to that for counties, although the state model differs from the county model in several respects.

(3) Adjusting the initial estimates of poor school-age children from the county model (step 1) for consistency by state with the estimates from the state model (step 2) to produce final estimates of the numbers of related children aged 5-17 in poverty by county for 1993.

(4) Producing county estimates for 1994 of the total number of children aged 5-17 from the Census Bureau's population estimates program. The Department of Education uses the estimates from step 3 and step 4 to calculate estimated proportions of poor school-age children for counties, which are also needed for the Title I allocation formulas.

Estimates for Puerto Rico, which is treated as a county equivalent in the allocation formula, are developed separately (see Chapter 5; see also National Research Council, 1997:App. F).

Steps 1-4 are summarized in the remainder of this chapter (see also Appendices A and B; Coder et al., 1996; Fisher and Siegel, 1997). The last section describes the differences between the revised 1993 estimates that were provided to the panel in October 1997, which are assessed in this report, and the original 1993 estimates that were provided to the panel in January 1997 and assessed in its first interim report (National Research Council, 1997). The changes in the estimates result principally from a change in one of the predictor variables in the county model that was found to improve its performance.[2]

## REVISED COUNTY MODEL

### County Equation

The county equation uses as predictor variables county estimates from Internal Revenue Service (IRS) records for 1993, food stamp program records for 1993, the 1990 census, and the Census Bureau's population estimates program for 1994. As the dependent or outcome variable, it uses county estimates of the number of poor school-age children averaged over 3 years of the March CPS (data from the March 1993, 1994, and 1995 CPS, covering income in 1992, 1993, and 1994). The equation takes the following form:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + u_i + e_i \ , \qquad (1)$$

[2]Subsequent chapters refer to the revised county model as the "log number (under 18) model" to distinguish it not only from the original model, but also from alternative models that were evaluated (see Chapters 3 and 4).

where:

$y_i$ = log(3-year weighted average of poor school-age children in county $i$),[3]
$x_{1i}$ = log(number of child exemptions reported by families in poverty on tax returns in county $i$),
$x_{2i}$ = log(number of people receiving food stamps in county $i$),
$x_{3i}$ = log(estimated population under age 18 in county $i$),
$x_{4i}$ = log(number of child exemptions on tax returns in county $i$),
$x_{5i}$ = log(number of poor school-age children in county $i$ in the previous census),
$u_i$ = model error for county $i$, and
$e_i$ = sampling error of the dependent variable for county $i$.

***Dependent Variable*** The Census Bureau decided to model the *number* of poor school-age children, instead of the *proportion*, because of concern that the county population estimates of school-age children that would form the basis for converting the estimated proportions to estimated numbers were of uncertain quality. Hence, it would be difficult to construct estimates of the precision of the estimated numbers of poor school-age children, which play the most important role in the Title I allocation formula.

The Census Bureau decided to estimate the number of poor school-age children *at a particular time* and not to estimate the *change* in the number since the 1990 census because it concluded that the available administrative data were likely to be measured more consistently across areas at a given time than they would be over time, given changes in tax and transfer programs. The Census Bureau decided to combine 3 years of CPS data for county estimation to improve the precision of the CPS estimates. Because only a subset of counties have households in the March CPS sample, the relationships between the predictor variables and the dependent variable in the model are estimated solely from this subset of counties. This subset includes proportionately more large counties and proportionately fewer small counties than the distribution of all counties. Because values of 0 cannot be transformed into logarithms, a number of counties whose sampled households contain no poor school-age children are excluded from the estimation. In all, 1,184 of 3,143 counties were included in the 1993 model estimation—the remainder either had no CPS sampled households with poor school-age children (304 counties) or no CPS sampled households at all (1,655 counties).

[3]The weighted average of the number of poor school-age children in each county is the product of the weighted 3-year average CPS poverty rate for related children aged 5-17 and the weighted 3-year average CPS number of related children aged 5-17; see National Research Council (1997:Ch. 3) for how the weights are derived.

***Predictor Variables*** The choice of predictor variables was governed by data availability and the assumed relationship of the variables to poverty. The number of child exemptions reported by families in poverty on tax returns and the number of food stamp recipients were included as variables that are indicative of poverty and available on a consistent basis (or reasonably consistent basis, in the case of food stamps) for all counties in the nation.[4] The 1990 census estimate of poor school-age children was used in the 1993 model on the assumption that previous poverty is likely to be indicative of subsequent poverty. The total number of child exemptions on tax returns and the population estimate of the total number of children under 18 were included in order to cover children not reported on tax returns (i.e., in nonfiling families), who are assumed to be poorer on average than other children. (The estimated regression coefficients for the county model predictor variables are given in Table 4-1.)

***Form of the Variables*** The dependent variable and all of the predictor variables are measured on a logarithmic scale. A reason to use logarithms is the wide variation in the CPS estimate and the values of the predictor variables among counties: transforming the variables to logarithms made their distributions more symmetric and the relationships between some of them and the dependent variable more linear.

***Estimation of Model and Sampling Error Variance*** The total squared error of the county estimates (the difference between the model estimates and the direct estimates from the CPS) has two sources: model error ($u$) and sampling error ($e$), which are the last two terms in the county equation.[5] Model error is the difference between the value of the dependent variable that would have been obtained had all the households in the county been included in the CPS sample and the model estimate based on the predictor variables. Sampling error is the difference between the estimate of the dependent variable from the CPS sample and the value of the dependent variable that would have been obtained had all households in the county been included in the CPS sample. Model error is assumed to be constant across counties (see below). Sampling error is not constant across counties: it is larger for counties that have fewer households included in the CPS sample.

Because a procedure to estimate the sampling error variance directly for the March CPS has not yet been developed (see Chapter 6), the variances of the

---

[4]Poverty status for families on tax returns is determined by comparing the adjusted gross income on each return to the average poverty threshold for the total number of exemptions on the form. Although there are differences between the CPS and IRS definitions of income and family composition, they are not critical for purposes of developing a predictive model.

[5]As used in statistics, "error" is the inevitable discrepancy between the truth and an estimate due to variability in measurements and the fact that model relationships are not precise.

model error and sampling error terms in the county equation are estimated in a multiple-step process that involves several assumptions. First, equation (1) is estimated for 1989, using the 1990 census estimate of poor school-age children as the dependent variable and 1989 IRS and food stamp data, 1990 census population data, and 1980 census poverty data as the predictor variables. A generalized variance function is used to estimate the sampling variance of the census estimates, which is quite small because of the large size of the census long-form sample. The total model error variance is then obtained by subtracting the sum of the estimated sampling variances from the estimated total squared error in the census equation. It is assumed that the total model error variance for the CPS equation for 1993 is the same as that for the 1990 census equation and that it has the same value for each county. The total sampling variance for the CPS equation, which is obtained by subtracting the total model error variance from the estimated total squared error, is then distributed among the counties as an inverse function of their sample size.

The resulting estimates of model error variance and sampling error variance are used to form weights for use in estimating the county model equation by weighted least squares.[6] They are also used to determine the weight to give to the model prediction and to the CPS direct estimate in developing estimates of poor school-age children for counties with sampled households in the CPS.

## Combining the County Equation and CPS Estimates

By calculating the relationships among the predictor variables and the CPS estimates of school-age children in poverty for the subset of counties that have households with poor school-age children in the March CPS sample, it is possible to obtain a good estimate of an equation for predicting the number of poor school-age children in a county, even though the CPS estimate for any specific county has a large level of uncertainty for many small counties. The prediction equation can then be used to predict the number of school-age children in poverty from the food stamp, IRS, population estimates, and previous census predictor variables for each county, whether or not the county is in the March CPS sample.

For counties that have households with poor school-age children in the March CPS sample, a weighted average of the model prediction and the estimate based on data from the sampled households (the direct estimate) is used to produce an estimate for that county using empirical Bayes ("shrinkage") procedures for combining estimates (see Fay and Herriot, 1979; Ghosh and Rao, 1994; and Platek et

[6]The weights used are the reciprocal of the sum of the estimated sampling variance of the estimate of the log number of poor school-age children in a given county plus the estimated model error variance, assumed to be constant across counties; see Appendix A (see also National Research Council, 1997:App. C)

al., 1987). The weights that are given to the model prediction and the direct estimate depend on their relative precision (see discussion above of how model error variance and sampling error variance are estimated). For a county with very few sample households in the CPS and hence a high level of sampling variability in the direct estimate, most of the weight will be given to the model prediction and little to the direct estimate. For a county with a larger number of sampled households in the CPS, more weight will be given to the direct estimate and less to the model prediction. In either case, assuming that the weights have been well estimated, the combined estimate will be at least as accurate as the better of the separate estimates (from the model or the CPS).[7] For counties that lack households with poor school-age children in the CPS sample, the prediction from the model is the estimate.

## STATE MODEL

### State Equation

The state model equation takes the following form (see also Fay, 1996; Fay and Train, 1997):

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i + e_i \ , \tag{2}$$

where:

$y_i$ = proportion of poor school-age children in state $i$ from one year of the CPS,[8]

$x_{1i}$ = proportion of child exemptions reported by families in poverty on tax returns in state $i$,

$x_{2i}$ = proportion of people receiving food stamps in state $i$,

$x_{3i}$ = proportion of people under age 65 who did not file an income tax return in state $i$,[9]

---

[7]For almost all counties that have households with poor school-age children in the CPS, most of the weight is given to the model prediction; for only 2 counties is the weight for the model prediction less than 0.5 and for only 13 counties is the weight for the model prediction less than 0.75.

[8]The numerator is the estimated number of poor related children aged 5-17 from the CPS, and the denominator is the estimated total population of children aged 5-17 (whether related or not) from the CPS. (The CPS universe excludes people in institutions and in military group quarters.)

[9]This percentage is obtained by subtracting the estimated number of exemptions on income tax returns for people under age 65 from the estimated total population under age 65 derived from demographic analysis; see Appendix B.

$x_{4i}$ = residual for state $i$ from a regression of the proportion of poor school-age children from the prior decennial census on the other three predictor variables,

$u_i$ = model error for state $i$, and

$e_i$ = sampling error of the dependent variable for state $i$.

## Differences from the County Equation

The Census Bureau's state model for estimates of poverty among school-age children is similar to the county model. However, it differs in a number of respects:

***Dependent Variable*** The state model uses the proportion of school-age children in poverty in each state as the dependent variable: that is, the dependent variable is a poverty ratio rather than the number of poor school-age children, as in the county model.[10] The numerator for the ratio is the CPS estimate of poor school-age children in a state (i.e., the estimate of the number of poor related children aged 5-17); the denominator is the CPS estimate of the total number of children aged 5-17 in the state. A different denominator—total CPS school-age children, rather than the slightly smaller universe of related school-age children—is used for consistency with the population estimates that are available to convert the estimated poverty ratios to estimated numbers of poor school-age children.

In addition, the dependent variable in the state model is derived from 1 year of CPS data (the March 1994 CPS for the 1993 model), rather than a 3-year average as in the county model. This decision was made because the sample sizes for states permit estimating the model with reasonable accuracy. It implicitly assumes that it is preferable when possible to have estimates that pertain directly to the income year.

***Predictor Variables*** The state model uses a somewhat different set of predictor variables than the county model. (The estimated regression coefficients for the state model predictor variables are given in Table 4-5.) The state model includes a predictor variable that is the residual from a regression of the proportion of poor school-age children from the prior decennial census on the other three predictor variables. During the development of the state model, the Census Bureau determined that there was a correlation between the residuals from estimating the model for 1979 with 1980 census data and the residuals from estimating the model for 1989 with 1990 census data. In other words, states that had

[10]The predicted variable is termed a ratio because the denominator is not exactly the same as that for the official published poverty rates.

more poverty than predicted by the cross-sectional model for 1979 also tended to have more poverty than predicted by the cross-sectional model for 1989. This result was used to improve the model predictions by including the residual from a regression for the prior census as one of the predictor variables.

***Form of the Variables*** The variables in the state model are proportions rather than numbers and are not transformed to a logarithmic scale as is done in the county model.[11] A log-based model was examined, but the Census Bureau decided not to transform the variables because, unlike the situation with the county model, the state-level distributions of the estimated proportions for the predictor variables are reasonably symmetric, and the relationships of the state-level estimated proportions with the dependent variable are approximately linear.

### Combining the State Equation and CPS Estimates

All states have sampled households in the CPS; however, the variability associated with estimates from the CPS is large for some states. As is done for the initial county estimates, the predictions from the state model and the CPS estimates are weighted according to their relative precision to produce estimates of the proportion of poor school-age children in each state. To produce estimates of the number of poor school-age children in each state, the estimates of the proportion poor are multiplied by estimates of the total number of noninstitutionalized school-age children. For the 1993 model, these estimates are derived from the Census Bureau's program of population estimates.[12] Finally, the state estimates of the number of poor school-age children are adjusted to sum to the CPS national estimate of related school-age children in poverty: this adjustment is a minor one, involving multiplying the state estimates for 1993 by 1.0091.

## RAKING THE COUNTY ESTIMATES TO STATE ESTIMATES

The final step in developing estimates of numbers of poor school-age children by county is to adjust the estimates from the county model for consistency with the estimates from the state model. The estimated logarithmic counts from the county model are first transformed to numbers (with a correction for transfor-

[11]The estimates that are transformed into logarithms in the county model are numbers, not proportions. However, evaluation determined that, if the county model were to estimate proportions, a logarithmic transformation of the dependent and predictor variables would be helpful in that case as well.

[12]The estimates of noninstitutionalized school-age children, which include some adjustments for residents of military group quarters and college dormitories, are the closest approximation available to the CPS estimates of school-age children.

mation bias).[13] The estimate for each state from the state model is then divided by the sum of the estimates for each county in that state to form a state raking factor. Each of the county estimates in a state is multiplied by the state raking factor so that the sum of the adjusted county estimates equals the state estimate. For the revised county estimates of poor school-age children in 1993, the average state raking factor was 1.065; two-thirds of the factors were between 0.975 and 1.154.

## ESTIMATING PROPORTIONS

The Census Bureau's county model predicts the *number* of school-age children in families in poverty. Estimates of the *proportion* of poor school-age children in families, which play an important but secondary role in the Title I allocation formula, are obtained by the Department of Education by dividing the estimated number of poor school-age children from the county model by an updated estimate of the total county population aged 5-17. These estimates are produced from the Census Bureau's population estimates program (see Appendix B).

## DIFFERENCES BETWEEN TWO PROCEDURES

The procedure described above to produce the revised 1993 county estimates that were provided to the panel in October 1997 differs in some respects from the procedure that was used to produce the original 1993 estimates. Specifically:

- The revised county model includes the population under 18 as a predictor variable; the original county model included the population under 21 as a predictor variable. The purpose of this variable (whether for the population under 18 or under 21) is to estimate—in conjunction with the variable measuring total child exemptions on IRS tax returns—the number of children in families that did not file a tax return. Evaluation determined that the estimation was not working well for counties with large numbers of people under age 21 in group quarters, primarily college students and military personnel. Specifically, the model was overpredicting the number of school-age children for those counties. Limiting the predictor variable to the population under 18 reduced the bias in the model

[13] Transformation bias occurs when a regression model estimates an expected value for the dependent variable that is on a different scale than that for which estimates are needed. In this instance, the county model predicts poor school-age children on the log scale; when the predictions on the log scale are exponentiated back to the original numeric scale, the result is the exponential of the expected value of the dependent variable on the log scale, which is different from the expected value of the dependent variable on the original scale. This difference is referred to as transformation bias, for which a correction is made.

predictions for counties classified by percent group quarters residents and improved the model predictions in other respects (see Chapter 4).

- Examination of the pattern of residuals (differences between the model predictions and the direct estimates) for counties with sampled households in the March CPS indicated that the original method for estimating model error variance and sampling error variance (described above) was not working as well as it should. The variability of the standardized residuals increased with the number of CPS sample cases rather than remaining constant, and this pattern was common to a variety of alternative models that were examined. The revised 1993 model includes a slight revision to the procedure for estimating the sampling error variance, which moderated but did not eliminate the anomalous pattern. Further work (see Chapter 6) will be required to further reduce the problem. However, this variability probably has limited effect on the estimates because the main effect of the sampling error variance estimation is on the weight to give to the model prediction versus the CPS direct estimate in forming estimates for counties that have sampled households with poor school-age children in the CPS. Since the direct estimates have small weights for most counties, changing the weights will not have a substantial impact.

- The original model was estimated using a method-of-moments procedure; for the revised model, it was decided to use maximum likelihood estimation. There is a small effect on the estimated regression coefficients for the predictor variables from the use of maximum likelihood instead of method of moments. Primarily, the effect is to increase the estimated sampling error variance. Hence, in comparison with the original 1993 estimates, the revised model predictions are given somewhat more weight and the CPS direct estimates are given somewhat less weight when weighted estimates are formed for counties that have sampled households with poor school-age children in the CPS. However, as just noted, relatively few counties have large weights on the direct estimates.

- The 1994 population estimates of children aged 5-17 that are used to convert the revised estimated numbers of poor school-age children to estimated proportions differ somewhat from the original 1994 population estimates that were used. These revised estimates incorporate more complete records of births and deaths. They also include a refined raking adjustment: the estimates are derived by an iterative proportional fitting procedure that rakes the 1990 census county estimates for school-age children to independently derived county total population estimates and state estimates of school-age children for 1994. The refinement was to rake separately the 1990 census estimates of school-age children in group quarters and school-age children not in group quarters.

# 3

# Alternative County Models

The Census Bureau's procedure for developing updated county estimates of poor school-age children in 1993 uses a county model, a separate state model, and county population estimates. All three components are important, and the panel considers all three in this report. However, the heart of the estimation procedure is the county model. The task of developing good estimates of poor school-age children from a county model is more challenging than the task of developing good state estimates of poor school-age children or good county estimates of total school-age children. Hence, the panel focused its evaluation efforts mainly on the county model.

In selecting a specific model for developing small-area poverty estimates that are to be used for such an important public purpose as allocating funds, it is important to compare the selected model to alternative models that may have specific advantages or that appear to be equally good. When the original county estimates of poor school-age children in 1993 were provided to the panel, the Census Bureau had not had time to undertake a thorough assessment of the performance of that model or to compare it to other models. Subsequently, the panel and the Census Bureau developed a range of alternative county models to evaluate. In a first round of evaluations, 12 models were examined. On the basis of the results of those evaluations, a second round of evaluations examined four models that appeared practicable to use to provide revised county estimates of poor school-age children in 1993. The basic features of the models that were examined are summarized below.[1]

[1]For technical information on the models included in the first round of evaluations, see Appendix A. The models specified do not exhaust the list of possibilities, but they are a reasonable range of alternatives to consider at the present time. See Chapter 6 for model formulations that could be considered as part of a longer term research program for small-area estimation.

## MODEL CHARACTERISTICS

The alternative county models that the Census Bureau and the panel examined are distinguished broadly by three characteristics: (1) treatment of information from the previous census—whether the model includes a predictor variable from the previous census in a single equation or uses a bivariate formulation that links a census equation with a CPS equation; (2) the form of the variables—whether they are numbers or proportions, transformed to logarithms or untransformed; and (3) whether the model includes intercept terms for each state (i.e., fixed state effects).

***Treatment of Information from the Previous Census*** The revised county model the Census Bureau used to produce estimates of poor school-age children in 1993 is a single-equation model in which the dependent variable is from the CPS and one of the predictor variables is the estimated number of poor school-age children from the previous census. The inclusion of the census predictor variable is based on the assumption that poverty in a prior year is indicative of poverty in a later year.

The state model makes use of information from the previous census in a different way. The state model equation, in which the dependent variable is also from the CPS, includes a predictor variable that is the estimated residual from a similar regression for the previous census. The underlying assumption is that states that had more (less) poverty than predicted for the census year will continue to have more (less) poverty for a later year than the model would predict without the residual variable. This assumption was supported by an analysis that showed the residuals to be correlated from a state model estimated with 1980 census data and a state model estimated with 1990 census data (see Chapter 2).

The possible advantage of having the county model include the estimated residual from an equation for the previous census could not be established because the necessary administrative data are not available with which to estimate a county equation from the 1980 census (for 1979). The Census Bureau developed a bivariate formulation of the county model as a way to make more complete use of information from the previous census in a manner analogous to the state model (Bell, 1997a). In the bivariate formulation, the 1993 county model jointly estimates two separate equations for March 1993-1995 CPS data and 1990 census data, respectively, in which the model errors of the two equations are allowed to be correlated (see below, "Bivariate Models").

***Form of the Variables*** In the revised county model, the dependent variable is the log number of poor school-age children, and the predictor variables are also numbers that are transformed to logarithms. The Census Bureau and the panel examined alternative county models in which the dependent variable is the proportion, or rate, of poor school-age children. For some of these rate models, the dependent and predictor variables are transformed to logarithms; for others, they are not transformed. Models for which the dependent and predictor variables are untransformed numbers were not considered because, when not transformed to logarithms, the distributions of the predictor variables at the county level have a wide range and are not symmetric; also, the predictor variables do not have linear relationships with the dependent variable. Untransformed poverty rates do not share these problems to the same extent, although it is possible to obtain predicted negative values from an untransformed formulation.

***Inclusion of Fixed State Effects*** In the revised county model, there are no predictor variables that explicitly account for regional or state effects. After the initial county estimates are produced from the model, they are raked for consistency with the estimates from the state model. Analysis of the size and variability of the raking factors (see Chapter 4) suggested that the county model may not adequately account for differences among states in the relationship of the predictor variables to the dependent variable and, consequently, that the county model may not adequately account for the variation among counties within a state.

As a way to explore this problem, the Census Bureau developed a fixed state effects model. This model includes a dummy variable for each state, which is 1 for all counties in the state and 0 otherwise. The purpose of these state indicator variables is to enable the model to more accurately capture the variation among counties within each state by accounting for differences in the level of the dependent variable by state.

## MODELS EXAMINED IN FIRST ROUND OF EVALUATIONS

Of the 12 models examined in the first round of evaluations, 6 were single-equation models, and 6 were bivariate models. Nine of the 12 models transform the values of the dependent and predictor variables into logarithms. Because logarithms cannot be taken for values of 0, these models are estimated only for the counties with sampled households in the CPS that contain at least one poor school-age child: 1,184 of 3,143 counties for the 1993 models. The other three models, which do not transform the variables (all three are rate models), use data for all counties with sampled households in the CPS: 1,488 counties for the 1993 models. A topic for future work is how to use all counties with CPS sampled households in estimating a log-based model (see Chapter 6).

### Single-Equation Models

The basic form of a single-equation county model is

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} \; . \; . \; . + \beta_5 x_{5i} + u_i + e_i \;, \qquad (1)$$

where:

$y_i$ = the dependent variable in county $i$ (number or proportion of poor school-age children from 3 years of CPS data),
$x_{1i}$ . . . $x_{5i}$ = the predictor variables in county $i$,
$u_i$ = model error for county $i$, and
$e_i$ = sampling error of the dependent variable for county $i$.

The formulation with fixed state effects adds a dummy variable for each state, which is 1 for all counties in the state and 0 otherwise. The intercept term, $\alpha$, is dropped from the models with fixed state effects to avoid overidentification. (The addition of a large number of dummy variables does not result in too few degrees of freedom because more than 1,000 counties are used to fit the regression coefficients.)

Six single-equation models were evaluated in the first round (see Table 3-1):

(1) ***Log Number Model (Under 21)*** The dependent variable is the CPS estimate of the log number of poor school-age children, derived by multiplying for each county the 3-year weighted average poverty rate for related children aged 5-17 by the 3-year weighted average of total related children aged 5-17. The predictor variables are the number of child exemptions (assumed to be under age 21) reported by families in poverty on tax returns; the number of people receiving food stamps; the estimated population under age 21; the total number of child exemptions on tax returns; and the estimated number of poor school-age children in the 1990 census. For the 1993 model, the IRS and food stamp data pertain to 1993; the population estimates data pertain to 1994. All variables are transformed to logarithms. This is the original model used by the Census Bureau to produce 1993 county estimates of poor school-age children.

(2) ***Log Number Model (Under 18)*** The dependent and predictor variables are the same as in (1), except that the estimated population under age 18 replaces the estimated population under age 21. This is the revised model used by the Census Bureau to produce 1993 county estimates of poor school-age children (see Chapter 2). It was included in the first round of evaluations after it became apparent that the log number model (under 21) was not performing well for counties with large numbers of people under age 21 in group quarters (see Chapter 4).

(3) ***Log Number Model with Fixed State Effects*** The dependent and predictor variables are the same as in (1), with the addition of state indicator variables.

(4) ***Log Rate Model (Under 21)*** The dependent variable is the CPS estimate of the log proportion poor, or log poverty rate, for school-age children: more

TABLE 3-1 Single-Equation County Models: Dependent Variable and Predictor Variables

| Model | Dependent Variable, $y_i$ | Predictor Variables, $x_{1i} \ldots x_{5i}$ | Form of the Predictor Variables |
|---|---|---|---|
| (1) Log Number (Under 21) | Log 3-year weighted average number of poor school-age children | (1) Number of child exemptions reported by families in poverty on tax returns<br>(2) Number of people receiving food stamps<br>(3) Population under 21<br>(4) Number of child exemptions on tax returns<br>(5) Number of poor school-age children in 1990 census | Transformed to logarithms |
| (2) Log Number (Under 18) | Log 3-year weighted average number of poor school-age children | (1) Number of child exemptions reported by families in poverty on tax returns<br>(2) Number of people receiving food stamps<br>(3) Population under 18<br>(4) Number of child exemptions on tax returns<br>(5) Number of poor school-age children in 1990 census | Transformed to logarithms |
| (3) Log Number (Under 21) with Fixed State Effects | Same as Log Number Under 21 (1) | Same as Log Number Under 21 (1) with the addition of state indicator variables | Transformed to logarithms |

| | | | |
|---|---|---|---|
| (4) Log Rate | Log poverty ratio for school-age children (3-year sum of poor related children 5-17 divided by 3-year sum of total CPS children 5-17) | (1) Ratio of number of child exemptions reported by families in poverty on tax returns to total number of child exemptions on tax returns<br>(2) Ratio of number of people receiving food stamps to total population<br>(3) Ratio of total number of child exemptions on tax returns to total population under 21<br>(4) Ratio of number of poor related children aged 5-17 to total number of related children aged 5-17 from the 1990 census | Transformed to logarithms |
| (5) Rate | Poverty ratio for school-age children (same as Log Rate (4), except untransformed) | Same as Log Rate (4), except untransformed | Untransformed |
| (6) Hybrid Rate-Number | Same as Log Rate (4) | Same as Log Number Under 21 (1) | Transformed to logarithms |

NOTE: The models are estimated for 1993 from 3 years of CPS data (March 1993, 1994, and 1995, covering income in 1992, 1993, and 1994).

precisely, a poverty ratio—similar to the state model—in which for each county the numerator is the sum over 3 years of the estimated number of poor related children aged 5-17 and the denominator is the sum over 3 years of the estimated total number of CPS children aged 5-17. The predictor variables are also ratios: the ratio of the number of child exemptions reported by families in poverty on tax returns to the total number of child exemptions on tax returns; the ratio of the number of people receiving food stamps to the total population (all ages); the ratio of the total number of child exemptions on tax returns to the total population under age 21;[2] and the ratio of the estimated number of poor related children aged 5-17 to the estimated total number of related children aged 5-17 from the 1990 census. All variables are transformed to logarithms.

(5) ***Rate Model*** The dependent variable and predictor variables are the same as in (4), but all variables are ratios, untransformed.

(6) ***Hybrid Log Rate-Number Model*** The dependent variable is the CPS estimate of the poverty ratio for poor school-age children as in (4); the predictor variables are the same as in (1), that is, they represent numbers, not ratios; and all variables are transformed to logarithms.

Each single-equation model was estimated for 1993, by averaging 3 years of CPS data (March 1993, 1994, and 1995, covering income years 1992, 1993, and 1994) to form the dependent variable. Each model was also estimated for 1989: for the dependent variable, by averaging 3 years of CPS data (March 1989, 1990, and 1991, covering income years 1988, 1989, and 1990); for the predictor variables, by using appropriate data from IRS and food stamp records for 1989, 1990 population estimates of school-age children, and 1980 census estimates of poor school-age children. The 1989 models were estimated to permit comparisons with 1990 census estimates of poor school-age children in 1989 for evaluation purposes (see Chapter 4). Finally, each single-equation model was also estimated for 1989 by using 1990 census data rather than CPS data to form the dependent variable. The census equation was needed to determine how to distribute the total squared error of the CPS equation (1993 or 1989) into model error variance and sampling error variance (see Appendix A).

---

[2]In 292 counties, the ratio of total child exemptions on tax returns to the total population under age 21—the tax filer population ratio—is greater than 1, which means that the nonfiler ratio (1 minus the filer ratio) is negative. Because negative values cannot be transformed into logarithms, the log rate equation includes the filer ratio and not the nonfiler ratio. There are several reasons that filer ratios may be greater than 1: addresses on tax returns are not always the county of residence as defined for population estimates; tax filers may report exemptions for children who do not reside with them; and some child exemptions are for children aged 21 or older.

## Bivariate Models

The bivariate formulation of the county model for 1993 estimates of poor school-age children involves the joint estimation of two equations: one for 1993, in which the dependent variable is formed by averaging 3 years of CPS data, and one for 1989, in which the dependent variable is formed by using 1990 census data. The bivariate formulation allows for a correlation between the model errors in the two equations—$u_{CPSi}$ and $u_{CENi}$ in equations (2) and (3) below (see also Appendix A). It is through this mechanism that data from the previous census are incorporated in predicting the number of poor school-age children in 1993. Hence, the bivariate models do not include 1990 census estimates of poor school-age children as a predictor variable in the 1993 equation.

The basic form of the CPS equation in the bivariate formulation is

$$y_{CPSi} = \alpha + \beta_1 x_{CPS1i} + \beta_2 x_{CPS2i} \ . \ . \ . + \beta_4 x_{CPS4i} + u_{CPSi} + e_{CPSi} \ , \qquad (2)$$

where:

$y_{CPSi}$ = the dependent variable in county $i$ (number or proportion of poor school-age children from 3 years of CPS data),
$x_{CPS1i}$ . . . $x_{CPS4i}$ = the predictor variables in county $i$,
$u_{CPSi}$ = model error for county $i$, and
$e_{CPSi}$ = sampling error of $y_{CPSi}$ for county $i$.

The basic form of the census equation in the bivariate formulation is

$$y_{CENi} = \alpha^* + \beta_1^* x_{CEN1i} + \beta_2^* x_{CEN2i} \ . \ . \ . + \beta_4^* x_{CEN4i} + u_{CENi} + e_{CENi} \, , \qquad (3)$$

where:

$y_{CENi}$ = the dependent variable in county $i$ (number or proportion of poor school-age children from the 1990 census),
$x_{CEN1i}$ . . . $x_{CEN4i}$ = the predictor variables in county $i$,
$u_{CENi}$ = model error for county $i$, and
$e_{CENi}$ = sampling error of $y_{CENi}$ for county $i$.

The formulation with fixed state effects adds a dummy variable for each state, which is 1 for all counties in the state and 0 otherwise.

Six bivariate models were evaluated in the first round (see Table 3-2):

(7) ***Bivariate Log Number Model*** In the CPS equation for this bivariate model, the dependent variable is the same as in model (1), the single-equation log number model (under 21). The predictor variables are the same as in (1), except

TABLE 3-2 Bivariate County Models: Dependent Variable, Predictor Variables, and Form of the Predictor Variables for the CPS Equation for 1993

| Model | Dependent Variable, $y_{CPSi}$ | Predictor Variables, $x_{CPS1i} \ldots x_{CPS4i}$ | Form of the Predictor Variables |
|---|---|---|---|
| (7) Log Number (Under 21) | Log 3-year weighted average number of poor school-age children (same as single-equation Log Number Under 21 (1)) | (1) Number of child exemptions reported by families in poverty on tax returns<br>(2) Number of people receiving food stamps<br>(3) Population under 21<br>(4) Number of child exemptions on tax returns (same as single-equation Log Number Under 21, except there is no previous census variable) | Transformed to Logarithms |
| (8) Log Rate | Log poverty ratio for school-age children (3-year sum of poor related children 5-17 divided by 3-year sum of total CPS children 5-17 (same as single-equation Log Rate (4)) | (1) Ratio of number of child exemptions reported by families in poverty on tax returns to total number of child exemptions on tax returns<br>(2) Ratio of number of people receiving food stamps to total population<br>(3) Ratio of total number of child exemptions on tax returns to total population under 21 (same as single-equation Log Rate, except there is no previous census variable) | Transformed to Logarithms |

| | | | |
|---|---|---|---|
| (9) Rate | Poverty ratio for school-age children (same as Bivariate Log Rate (8), except untransformed) | Same as Bivariate Log Rate, except untransformed | Untransformed |
| (10) Log Number with Fixed State Effects | Same as Bivariate Log Number Under 21 (7) | Same as Bivariate Log Number Under 21 with the addition of state indicator variables | Transformed to Logarithms |
| (11) Log Rate with Fixed State Effects | Same as Bivariate Log Rate (8) | Same as Bivariate Log Rate with the addition of state indicator variables | Transformed to Logarithms |
| (12) Rate with Fixed State Effects | Poverty ratio for school-age children (same as Bivariate Log Rate (8), except untransformed) | Same as Bivariate Log Rate with the addition of state indicator variables, except untransformed | Untransformed |

NOTES: The models are estimated for 1993 from a CPS equation for 1993 and a 1990 census equation for 1989. The census equation for 1989 for each bivariate model is of the same form as the corresponding CPS equation for 1993. The 1989 equations use the number of poor school-age children or the poverty ratio for school-age children from the 1990 census as the dependent variable; the predictor variables are from IRS and food stamp records for 1989 and population estimates from the 1990 census.

that the 1990 census estimated number of poor school-age children is dropped from the equation. In the census equation for this bivariate model, the dependent variable is the 1990 census estimate of the number of poor school-age children in 1989; the predictor variables are the same as in the CPS equation, except that the IRS and food stamp data pertain to 1989 instead of 1993, and the population data are from the 1990 census rather than from the population estimates program. All variables are transformed to logarithms.

(8) ***Bivariate Log Rate Model*** In the CPS equation, the dependent variable is the same as in model (4), the single-equation log rate model (under 21). The predictor variables are the same as in (4), except that the 1990 census estimated poverty rate for school-age children is dropped from the equation. In the 1990 census equation, the dependent variable is the estimated log poverty ratio for school-age children from the census; the predictor variables are the same as in the CPS equation, except that the IRS and food stamp data pertain to 1989 instead of 1993 and the population data are from the 1990 census rather than from the population estimates program. All variables are ratios, transformed to logarithms.

(9) ***Bivariate Rate Model*** The dependent and predictor variables in the CPS and census equations are the same as in (8), but all variables are ratios, untransformed.

(10) ***Bivariate Log Number Model with Fixed State Effects*** The dependent and predictor variables in the CPS and census equations are the same as in (7), with the addition of state indicator variables in each equation. All variables are transformed to logarithms.

(11) ***Bivariate Log Rate Model with Fixed State Effects*** The dependent and predictor variables in the CPS and census equations are the same as in (8), with the addition of state indicator variables in each equation. All variables are ratios, transformed to logarithms.

(12) ***Bivariate Rate Model with Fixed State Effects*** The dependent and predictor variables in the CPS and census equations are the same as in (9), with the addition of state indicator variables in each equation. All variables are ratios, untransformed.

## MODELS EXAMINED IN SECOND ROUND OF EVALUATIONS

The first round of evaluations included an internal evaluation, in which the regression output for all 12 models was examined to assess the validity of the underlying assumptions (see Appendix C). It also included an external evaluation, in which estimates of poor school-age children in 1989 from the six single-equation models were compared with 1990 census estimates (see Appendix D). The results of these evaluations led the Census Bureau and the panel to drop several models from further consideration at this time.

The untransformed rate model (5) and the hybrid log rate-number model (6) were dropped from consideration because they performed somewhat worse, on balance, than the other models on both the internal and external evaluations. For example, in the comparisons of model estimates of poor school-age children in 1989 with 1990 census estimates, models (5) and (6) exhibited the largest overall absolute differences of their estimates from the census (see Table D-3). Also, the standardized residuals (differences between the model predictions and the reported values for each observation) from the regression equations for models (5) and (6) were not distributed normally.

The bivariate formulation (models 7-12) is promising in that it makes fuller use of the information from the previous census than the single-equation formulation. However, there is less experience with bivariate modeling than with modeling that uses a single equation for the kinds of estimates that are needed. More important, because the IRS and food stamp predictor variables at the county level are not available for 1979, it is not possible to evaluate bivariate models by comparison with estimates from the 1990 census. (Such a model would require joint estimation of a 1989 equation in which CPS data form the dependent variable and a 1979 equation in which 1980 census data form the dependent variable.) Hence, the bivariate formulation was not pursued for use at this time. However, further development of bivariate and multivariate models, which might include CPS equations for more than 1 year, as well as a census equation, is worth pursuing for the longer run (see Chapter 6).

Evaluation results indicated that the county model would likely benefit from taking account of state effects in some way. The addition of state indicator variables to either a single-equation or bivariate model (3, 10-12) was promising in some respects, but a fixed state effects approach did not seem clearly superior to other models that were examined. There was no time to investigate other approaches to account for state effects, although the panel believes that the county model could be improved in this regard in the near term with more research (see Chapter 6).

At the conclusion of the first round of evaluations, the Census Bureau and the panel focused on four models that were considered serious candidates to produce revised 1993 county estimates of poor school-age children. These four candidate models were then evaluated on several criteria. All four models are of the single-equation form with variables transformed to logarithms and without fixed state effects:

(a) Log number model (under 21), model (1) above, used by the Census Bureau to produce the original 1993 county estimates of poor school-age children.

(b) Log number model (under 18), model (2) above. This model is the same as model (a) except that the population under age 18 replaces the population under age 21 as a predictor variable.

(c) Log rate model (under 21), model (4) above. The rate formulation is used in the Census Bureau's state model, and the panel believed that, in log form, it could improve the county model.

(d) Log rate model (under 18). This model is the same as model (c) except that the ratio of total child exemptions on tax returns to the total population under 18 replaces the ratio of total child exemptions on tax returns to the total population under age 21 as a predictor variable. The panel wanted to determine if this modification would improve the log rate model, since a similar modification improved the log number model. However, for reasons that are not clear, this modification to the log rate model worsened rather than improved its performance in several respects (see Chapter 4).

The model that the Census Bureau used to prepare the revised 1993 county estimates of poor school-age children is (b)—log number model (under 18), estimated with maximum likelihood. Chapter 4 describes the evaluations that were conducted of the four candidate models (a-d) and highlights key results. Appendix C analyzes the regression output for the 12 models that were included in the first round of evaluations and model (d). Appendix D provides 1990 census evaluation results for the six single-equation models that were included in the first round of evaluations and the four candidate models that were evaluated in the second round.

# 4

# Evaluations

The development of model-based estimates for small areas is a major research and development effort for which extensive evaluation is required. For updated estimates of poor school-age children for counties, a thorough assessment of all aspects of the estimation procedure is necessary so that policy makers can have confidence in using the estimates for allocating federal Title I education funds to counties. That assessment includes both an evaluation of a given model and comparisons with alternative models. Because there are no absolute criteria for what are acceptable evaluation results, a way to determine if the performance of a model can be improved is to examine alternative models. Such comparisons may indicate changes that would be helpful for a model; they may also suggest that an alternative model is preferable.

The Census Bureau's county estimates of poor school-age children are produced by using a county regression model, a state regression model, and county population estimates developed with demographic analysis techniques (see Chapter 2). A comprehensive evaluation for each of these components of the estimation procedure should include "internal" and "external" evaluations.

An internal evaluation is primarily an investigation of the validity of the underlying assumptions and features of a model. For a regression model, an internal validation is typically based on an examination of the residuals from the regression—the differences between the predicted and reported values of the dependent variable for each observation. In an external evaluation, the estimates from a model are compared with target or "true" values that were not used to develop the model. Ideally, internal evaluation of regression model output should precede external evaluation. If the assumptions required by a regression model are not supported to a reasonable extent, then even a positive external evaluation would not justify the choice of the model. Changes made to a model to address concerns raised by an internal evaluation would likely improve its performance in an external evaluation. Both internal and external evaluations should be carried out for alternative models.

When the original 1993 county estimates of poor school-age children were provided to the panel, the Census Bureau had not had time to complete a full evaluation of them. Subsequently, the panel developed a set of evaluation criteria, and the panel and the Census Bureau conducted a series of internal and external evaluations. The focus of the evaluation effort was on alternative county models, particularly the assumptions underlying the regression equations and how the estimates of poor school-age children in 1989 from each model compared with 1990 census estimates. The state model and the county population estimates were examined as well, both directly and as they contribute to the county estimates of poor school-age children. The evaluations, which are described in this chapter, include:

(1) internal evaluation of the regression output for alternative county models estimated for 1993 and 1989;

(2) comparison of estimates of poor school-age children for 1989 from alternative county models with 1990 census estimates, a form of external evaluation;

(3) consideration of differences between the CPS and census measurement of income and poverty as a factor that could explain differences between model-based estimates and census estimates for 1989;

(4) examination of the original 1993 county estimates to identify possibly anomalous estimates that were then reviewed with knowledgeable local people, another form of external evaluation;

(5) evaluation of the state model, including examination of regression output, external evaluation in comparison with 1990 census estimates, and consideration of the state raking factors by which county model estimates are adjusted to make them consistent with the state model estimates; and

(6) evaluation of county population estimates for children aged 5-17 (see also Appendix B).

The internal evaluation of regression output and the comparison of model-based estimates of poor school-age children for 1989 with 1990 census estimates—evaluations (1) and (2) above—were carried out for the four single-equation county models that were considered serious candidates to produce revised 1993 county estimates of poor school-age children (see Chapter 3 and Appendices C and D):

(a) log number model (under 21), the original model that the Census Bureau used to produce the original 1993 county estimates of poor school-age children;

(b) log number model (under 18), the revised model that the Census Bureau used to produce the revised 1993 county estimates of poor school-age children;

(c) log rate model (under 21); and

(d) log rate model (under 18).

In addition, the 1990 census comparisons (2) were performed for some other estimation procedures that rely much more heavily than do the four candidate models on estimates from the 1980 census (see below, "Comparisons with 1990 Census County Estimates").

The internal evaluation of regression output (1) and the comparison of estimates of poor school-age children for 1989 with 1990 census estimates (2) examined residuals and model differences from the census, respectively, for categories of counties. The following characteristics were used for categorizing counties: census division; metropolitan status of county; population size in 1990; population growth from 1980 to 1990; percent poor school-age children in 1980; percent Hispanic population in 1990; percent black population in 1990; for rural counties, persistent poverty from 1960 to 1990; for rural counties, economic type; percent group quarters residents in 1990; number of households in the CPS sample (or whether the county had sampled households); and (for 1990 census comparisons only) percent change in the proportion of poor school-age children from 1980 to 1990 (see details in Table 4-3, below).

## INTERNAL EVALUATION: COUNTY MODEL REGRESSION OUTPUT

The first test of a regression model is that it perform well when evaluated internally, that is, for the set of observations for which it is estimated. The panel and the Census Bureau examined the underlying assumptions of the four candidate models through evaluation of the regression model output for 1989 and 1993.[1] Although such an evaluation is not likely to provide conclusive evidence with which to rank the performance of alternative models, particularly when they use different transformations of the dependent variable, examination of the regression output is helpful to determine which models perform reasonably well.

[1]The evaluation of the county regression output pertains to the regression models themselves, that is, before the predictions are combined with the direct CPS estimates in a "shrinkage" procedure or raked to the estimates from the state model (see Chapter 2). For these models, the regression output comprises the model predictions for counties with at least one household with poor school-age children in the CPS sample. For the two log number models, the predictions are the log number of poor school-age children; for the two log rate models, the predictions are the log proportion of poor school-age children.

The assumptions investigated fall into two groups: assumptions concerning the functional form of the regression model and assumptions concerning the error distribution. Because properties of the error distribution affect the ability to fit a model, studies of these two types of assumptions are not entirely separable.[2]

The assumptions examined in the first group are linearity of the relationship between the dependent variable and the predictor variables; constancy of the assumed linear relationship over different time periods; and whether any of the included predictor variables are *not* needed in the model and, conversely, whether other potential predictor variables *are* needed in the model. The assumptions examined in the second group are normality (primarily symmetry and moderate tail length) of the distribution of the standardized residuals;[3] whether the standardized residuals have homogeneous variances, that is, whether the variability of the standardized residuals is constant across counties and does not depend on the values of the predictor variables; and absence of outliers. Each assumption is discussed in terms of the methods used for evaluation and the results of the evaluation for the four candidate models.

## Linearity

Linearity of the relationships between the dependent variable and the predictor variables was assessed graphically, by observing whether there was evidence of curvature in the plots of standardized residuals against the predictor variables in the model. In addition, plots of standardized residuals against CPS sample size and against the predicted values from the regression model were also examined for curvature.

The only evidence of nonlinearity is for the log number (under 21) model (a) for 1989. For that year, the standardized residuals appear to have a very modest curvature when plotted against the predicted values.

## Constancy over Time

Constancy over time of the assumed linear relationship of the dependent and predictor variables was assessed through comparison of the regression coefficients on the predictor variables for 1989 and 1993. While major changes in

[2]These assumptions were also examined for the analogous 1990 census regressions. However, since the census equations only affected the weights for the weighted least squares regression and the extent of “shrinkage” in combining model estimates and direct estimates for counties with households in the CPS sample, analyses of the 1990 census regressions are not discussed here.

[3]The standardization of the residuals involved estimating the predicted standard errors of the residuals, given the predictor variables, and dividing the observed residuals by the predicted standard errors. The predicted standard error of the residual for a county is a function of the estimated model error variance and the estimated sampling error variance (see Belsley, Kuh, and Welsch, 1980).

economic conditions are expected to cause some changes in the coefficients, a relatively stable regression equation would be desirable.[4]

Table 4-1 shows the regression coefficients for the predictor variables for the four candidate models for 1989 and 1993. In the log number models (a, b) for 1989 and 1993, the coefficients for the three "poverty level" predictor variables—child exemptions reported by families in poverty on tax returns (column 1), food stamp recipients (column 2), and poor school-age children from the previous census (column 5)—are similar. There are substantial differences across the two time periods in the estimated coefficients for the other two variables—population (under age 21 or under age 18, column 3) and total number of child exemptions on tax returns (column 4). However, the sum of the these two coefficients is generally close to 0 in each model in each year. Because these two variables are highly positively correlated, the predictions from equations with a similar sum for the two coefficients will be similar.

The sum of all coefficients in each equation for models (a) and (b) ranges from 1.04 to 1.07 and is significantly greater than 1. A sum equal to 1 would mean that county population size itself has no effect on the estimated number of poor school-age children. Because the sum is greater than 1, the estimated number of poor school-age children is a larger percentage of the population in the larger counties. While this result is difficult to explain as a function of county size, it may be that size reflects the effects of variables not included in the models.

In the log rate models (c, d), the coefficients for the three "poverty rate" predictor variables—ratio of child exemptions reported by families in poverty on tax returns to total child exemptions (column 1), ratio of food stamp recipients to the total population (column 2), and ratio of poor school-age children from the previous census (column 4)—are all positive and about the same size.[5] The coefficients for the ratio of total child tax exemptions to the population (under age 21 or under age 18, column 3) are negative, and there are substantial differences across the two time periods in the estimated coefficients. The sign of the related variable (total number of child tax exemptions) is generally negative in the log number equations. As in the log number equations, the coefficients in the log rate equations for population under 21 differ from the coefficients for population under 18.

[4]Because the county model is refit for each prediction year, constancy over time is not as important as it would be if the estimated regression coefficients from the model for one year were used for predictions for subsequent years. Nonetheless, it is disturbing for the regression coefficients to exhibit large, unexplained changes over time.

[5]The coefficients are also similar to the coefficients for the corresponding variables—number of child exemptions reported by families in poverty on tax returns, number of food stamp recipients and number of poor school-age children from the previous census—in the log number equations.

TABLE 4-1 Estimates of Regression Coefficients for Four Candidate County Models for 1989 and 1993

| Model | Counties (Number) | Predictor Variables[a] 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| (a) Log Number (under 21) | | | | | | |
| 1989 | 1,028 | 0.52 | 0.30 | 0.76 | –0.81 | 0.27 |
| | | (.07) | (.05) | (.22) | (.22) | (.07) |
| 1993 | 1,184 | 0.31 | 0.30 | 0.03 | 0.03 | 0.40 |
| | | (.08) | (.07) | (.21) | (.21) | (.09) |
| (b) Log Number (under 18) | | | | | | |
| 1989 | 1,028 | 0.50 | 0.23 | 1.79 | –1.80 | 0.32 |
| | | (.06) | (.05) | (.27) | (.27) | (.07) |
| 1993 | 1,184 | 0.38 | 0.27 | 0.65 | –0.59 | 0.34 |
| | | (.08) | (.07) | (.24) | (.24) | (.09) |
| | | Predictor Variables[b] | | | | |
| (c) Log Rate (under 21) | | | | | | |
| 1989 | 1,028 | 0.32 | 0.29 | –0.73 | 0.40 | |
| | | (.07) | (.04) | (.19) | (.07) | |
| 1993 | 1,184 | 0.23 | 0.31 | –0.07 | 0.41 | |
| | | (.08) | (.06) | (.18) | (.09) | |
| (d) Log Rate (under 18) | | | | | | |
| 1989 | 1,028 | 0.29 | 0.26 | –1.13 | 0.43 | |
| | | (.07) | (.04) | (.24) | (.07) | |
| 1993 | 1,184 | 0.26 | 0.30 | –0.42 | 0.38 | |
| | | (.08) | (.06) | (.20) | (.09) | |

NOTES: All predictor variables are on the logarithmic scale for numbers and rates. Standard errors of the estimated regression coefficients are in parentheses. The four models were estimated for each year with maximum likelihood. The original 1994 population estimates were used for the 1993 models; 1990 census population estimates were used for the 1989 models.

[a]Predictor variables: (1) number of child exemptions reported by families in poverty on tax returns; (2) number of people receiving food stamps; (3) population (under age 21 or under age 18); (4) total number of child exemptions on tax returns; (5) number of poor school-age children from previous (1980 or 1990) census.

[b]Predictor variables: (1) ratio of child exemptions reported by families in poverty on tax returns to total child exemptions; (2) ratio of people receiving food stamps to total population; (3) ratio of total child exemptions on tax returns to population (under age 21 or under age 18); (4) ratio of poor school-age children from previous (1980 or 1990) census.

## Inclusion or Exclusion of Predictor Variables

The possibility that one or more predictor variables should be excluded from a model was assessed by looking for insignificant *t*-statistics for the estimated values of individual regression coefficients.[6] The need to include a predictor variable, or possibly to model some categories of counties separately, was assessed by looking for nonrandom patterns, indicative of possible model bias, in the distributions of standardized residuals displayed for the various categories of counties.[7]

The only predictor variables with nonsignificant *t*-statistics are the population under age 21 (column 3 in Table 4-1) and total child exemptions on IRS income tax returns (column 4) for the log number (under 21) model (a) in 1993, and the ratio of child tax exemptions to the population under age 21 (column 3) for the log rate (under 21) model (c) in 1993. All other regression coefficients are significantly different from 0 at the 5 percent level. Application of Akaike's information criterion (AIC) confirmed the superiority of using the population under age 18 as a predictor variable in preference to the population under age 21 in the log number model. (The test was not performed for the log rate model.)

For most ways of categorizing counties, the standardized residuals do not exhibit systematic patterns. The exceptions are that all four models in 1989 tend to overpredict poor school-age children in counties with a high percentage of Hispanic residents (i.e., the model estimates are somewhat higher than the CPS direct estimates for these counties relative to other counties) and that the log number (under 21 and under 18) models (a, b) tend to overpredict poor school-age children in counties that are in metropolitan areas but are not the central county in the area.

## Normality

The normality of the standardized residuals was evaluated through use of Q-Q plots, which match the observed distribution of the residuals with the theoretical distribution, and other displays of the distribution. All four models exhibit some skewness in their standardized residuals, with the log rate models (c, d) showing somewhat more skewness than the log number models (a, b). For none of the models does the skewness appear sufficiently marked to be a problem.

---

[6]Although the performance of a predictive regression model is best assessed in terms of the joint impact of the predictor variables, examining the individual predictor variables can suggest ways in which a model might be improved.

[7]The distributional displays examined for this and other model assumptions were box plots.

## Homogeneous Variances

The homogeneity of the variance of the standardized residuals was assessed using a variety of statistics and graphical displays (see Appendix C). Examination of them clearly demonstrates some variability in the size of the absolute standardized residuals as a function of the predicted value (number or proportion of poor school-age children) and the CPS sample size for all four models. With regard to CPS sample size, one would expect the standardized residual variance to remain constant over the distribution of CPS sample size; however, it increases with increasing CPS sample size.

The heterogeneity of the variance of the residuals should be investigated because it suggests that there may be a problem with the model specification or in the assumptions that were used to calculate the standardized residuals. However, adjusting a model to remove this type of heterogeneity is likely to have only a small effect on the estimated regression coefficients or the model estimates. The effect on estimates of poor school-age children would stem from: a shift in the weights assigned to each county in fitting the regression model, which would very likely result in only a modest change in the estimated regression coefficients; and a change in the weight given to the direct estimates, which could have an appreciable effect only on the estimates for counties with large CPS sample sizes.

## Outliers

The existence of outliers was evaluated through examination of plots of the distributions of the standardized residuals and plots of standardized residuals against the predictor variables and through analysis of patterns in the distribution of the 30 largest absolute standardized residuals for the various categories of counties. However, it is difficult to evaluate the evidence for outliers that results from a least squares model fit, which has the property that it may miss influential outliers. In addition, since the four models are so similar and make use of the identical data, it is unlikely that an observation that was a marked outlier for one model would not also be a marked outlier for the other models.

An examination of the distributions of the standardized residuals indicates that none of the four models is especially affected by outliers, although the 1993 estimates have more outliers than the 1989 estimates, and nonrural counties and metropolitan counties that are not central counties have somewhat more outliers than other categories of counties. This analysis is only a start. It would be useful, using other statistics and various graphical techniques, to identify the counties that are not well fit by robustly estimated versions of these models in order to determine any characteristics that outlier counties have in common.

### Summary

The panel concludes that the analysis of the regression output for the four candidate county models for 1989 and 1993 largely supports the assumptions of the models: there is little evidence of important problems with the assumptions. The analysis does not strongly support one model over another, although it does support use of the population under age 18 instead of the population under age 21 as a predictor variable in the log number model.

All of the models exhibit a few common problems. First, they all behave somewhat differently for larger urban counties and counties with large percentages of Hispanic residents than for other counties. The differences are not pronounced, but research should be conducted to determine possible ways to modify the models to eliminate or reduce this problem. Second, all models show evidence of some variance heterogeneity with respect to both CPS sample size and poverty rate. This problem can likely be eliminated or reduced by research currently ongoing at the Census Bureau to develop direct estimates of the county-level sampling variances (see Chapter 6).

## EXTERNAL EVALUATION: COMPARISONS WITH 1990 CENSUS COUNTY ESTIMATES

For external evaluation, the panel and the Census Bureau compared the estimated number and proportion of poor school-age children for 1989 for the four candidate models with 1990 census estimates.[8] The evaluation examined the overall difference between the estimates from a model and the census and the differences for groups of counties categorized by various characteristics.

Evaluation by comparison with the 1990 census is not ideal because the census estimates are not true values. They are affected by sampling variability and population undercount; also, the census measurement of poverty differs from the CPS measurement in ways that are not fully understood (see National Research Council, 1997:Ch. 2, App. B; see also the Census Bureau's web site: http://www.census.gov/hhes/www/saipe93/inputs/cencpsdf.html). In addition, there is only one census-based validation opportunity: because of the lack of IRS and food stamp program data for counties for 1979, it is not possible to evaluate model-based estimates by comparison to the 1980 census. Reliance on a single

[8]The county estimates reflect the effects of the state model and the county population estimates as well as the county regression model, but the differences in model performance vis-à-vis the census in the evaluation are due to the particular form of the county model.

The models for which the 1990 census comparisons were performed were estimated with the method of moments. Maximum likelihood was used to estimate the log number (under 18) model (b) for the revised 1993 county estimates of poor school-age children. The differences in the estimates from the two techniques are small.

validation using the 1990 census is a problem because a model may perform better or worse in any one validation than it would on average over multiple validations. For this reason, if it were possible to compare model estimates with census or other estimates for 1993 instead of 1989, the results might turn out differently. Nonetheless, in the absence of other means of external validation, the panel and the Census Bureau relied heavily on the 1990 census comparisons to understand the performance of alternative models.

Evaluation by comparison with the 1990 census is intended to assess the accuracy of model estimates for the prediction year (i.e., 1989). The evaluation does not address the issue that model-based estimates are likely to be used for Title I allocations several years later. It would be useful to conduct research to reduce the time lag between the prediction year for model-based estimates and the year for Title I allocations to the extent possible (see Chapter 6).

The 1990 census estimates that are used in the comparisons are ratio adjusted by a constant factor to make the census national estimate of poor school-age children equal the 1989 CPS national estimate. This adjustment removes the difference of about 5 percent between the CPS and census estimates of total poor school-age children for 1989. Consequently, the differences between a model and the 1990 census in estimating poor school-age children for groups of counties can be interpreted as differences in shares. This feature is useful because the Title I allocation formula distributes funding as shares (percentages) of a fixed total dollar amount.

In addition to the four candidate models, the 1990 census comparisons were performed for four estimation procedures that rely much more heavily on 1980 census estimates. Given the substantial changes in the number and proportion of poor school-age children between the 1980 and 1990 censuses (see National Research Council, 1997:8-9), one would expect these procedures to perform less well than the candidate models in predicting poverty for school-age children in 1989.[9] In a period of less pronounced change, one or more of them might perform relatively well. The census comparisons were done for the following procedures:

(i) Stable shares procedure, in which the county estimates of poor school-age children for 1989 are the 1980 census estimates for 1979 after ratio adjustment to make the 1980 census national estimate equal the CPS national estimate for 1989. This simple procedure assumes no change over the decade in each county's share of the total number of poor school-age children nationwide: this is the same assumption that underlies previous practice for Title I allocations, in

[9]Although the interval was only 4 years instead of 10, substantial changes in the number and proportion of poor school-age children also occurred between 1989 and 1993 (see National Research Council, 1997:10-13).

which estimates from the decennial census were used in the formulas each year until the results from the next census became available.[10]

(ii) Stable shares within state procedure, in which the county estimates of poor school-age children for 1989 are the 1980 census estimates for 1979 after raking the estimates for the counties in each state to the estimates from the Census Bureau's state model for 1989. (The national raking employed in the state model also adjusts the total to equal the CPS national estimate for 1989.) This procedure assumes no change over the decade in each county's share of the total number of poor school-age children in its state.

(iii) Stable rates within state procedure (with conversion), in which the county estimates of poor school-age children for 1989 are developed by converting 1980 census estimates of the proportions of poor school-age children for 1979 to estimated numbers by use of 1990 county population estimates of total school-age children 5-17 and then raking the estimated numbers to the Census Bureau's state model estimates for 1989.

(iv) Averaging procedure, in which the county estimates of poor school-age children for 1989 are developed from an average of estimates from the 1980 census and the log number (under 21) model (a) for 1989.[11]

The rest of this section first discusses overall absolute differences from the 1990 census estimates for the four candidate models and the four procedures that rely more heavily on the 1980 census. It then discusses differences for categories of counties for the four candidate models and two of the procedures: the stable shares procedure and the averaging procedure. Differences for categories of counties for the other two procedures, which are intermediate in their reliance on 1980 census estimates, are provided in Appendix D.

---

[10]However, the estimates from the 1990 census that were previously used for Title I allocations were not adjusted to the current CPS national estimate of poor school-age children, which could affect the allocations for some counties. For example, some counties might meet the threshold test for a concentration grant if the census estimates were adjusted to the current CPS national estimate but not if the estimates were unadjusted.

[11]More precisely, the estimates are developed by averaging the proportions of poor school-age children from the 1980 census and the log number (under 21) model (a) for 1989, converting the estimates to numbers by use of 1990 county population estimates of total school-age children, and making an overall ratio adjustment to the CPS national estimate for 1989.

This procedure is analogous to the panel's recommendation for averaging 1990 census and 1993 model-based estimates for use in Title I allocations for the 1997-1998 school year. However, the panel's recommendation did not include raking the average estimates to the CPS national estimate of poor school-age children in 1993 (see National Research Council, 1997:38).

### Absolute Differences Between Model and Census County Estimates

Table 4-2 presents measures of the overall absolute difference between the model-based county estimates and the 1990 census county estimates of poor school-age children in 1989 for the four candidate models and the four procedures that rely more heavily on the 1980 census. If the 1990 census estimates are reasonably accurate, a good model will produce estimates that differ little from the census estimates, and the absolute differences will be less than for other reasonable models. Also, a good model will perform significantly better than a simple procedure that relies heavily on the previous census.

Column 1 of Table 4-2 is the average absolute difference for county estimates of the number of poor school-age children in 1989, measured as the sum for all counties of the absolute difference (ignoring the direction of the difference) between the model estimate and the 1990 census estimate for each county, divided by the total number of counties. Column 2 of Table 4-2 is the average proportional absolute difference for county estimates of the number of poor school-age children, measured as the sum for all counties of the absolute difference between the model estimate and the 1990 census estimate as a *proportion* of the census estimate for each county, divided by the total number of counties and expressed as a percentage. Column 3 is the average proportional absolute difference for county estimates of the proportion of poor school-age children.

The measure in column 1 assesses the difference between a model and the 1990 census in terms of numbers of poor children; the measures in columns 2 and 3 assess the difference in terms of percentage errors for counties. To illustrate the difference between absolute and proportional absolute differences, consider two counties, one with an estimated 10,000 poor school-age children from the census and an estimated 9,600 poor school-age children from the model and the other with an estimated 1,000 poor school-age children from the census and an estimated 1,400 poor school-age children from the model. The absolute difference in the number of poor school-age children is the same for both counties (400), but the proportional absolute difference is only 4 percent for the first county and 40 percent for the second.

From a national perspective, it can be argued that absolute differences are more important for effective Title I allocations because Title I funds are primarily distributed in proportion to the number of children in a county; therefore, the amount of funds that are misallocated depends primarily on the number of children rather than the percentages by county. For example, an error of 5 percent in the number of school-age children in poverty in a large county could correspond to tens of thousands of children and have more impact on the allocation of funds than errors of 5 percent in several smaller counties. However, from the county perspective, proportional errors are also important. Ideally, a model will perform well on both types of measures.

The panel draws several conclusions from Table 4-2:

TABLE 4-2 Comparison of Model Estimates and Other Procedures with 1990 Census County Estimates of the Number and Proportion of Poor Related Children Aged 5-17 in 1989

| Model | Average Absolute Difference: 1 Number of Poor Children 5-17[a] | Average Proportional Absolute Difference, in Percent: 2 Number of Poor Children 5-17[b] | Average Proportional Absolute Difference, in Percent: 3 Proportion of Poor Children 5-17[c] |
|---|---|---|---|
| Candidate Models | | | |
| (a) Log number (under 21) | 272 | 15.4 | 16.4 |
| (b) Log number (under 18) | 268 | 16.4 | 17.7 |
| (c) Log rate (under 21) | 275 | 17.5 | 17.1 |
| (d) Log rate (under 18) | 283 | 18.8 | 18.6 |
| Procedures that Rely More Heavily on the 1980 Census | | | |
| (i) Stable shares | 570 | 30.1 | N.A. |
| (ii) Stable shares within state | 380 | 27.1 | N.A. |
| (iii) Stable rates within state, with conversion | 381 | 26.2 | N.A. |
| (iv) Average of 1980 census and 1989 log number (under 21) model (a) | 286 | 19.0 | N.A. |

NOTES: The census estimates are controlled to the CPS national estimate for 1989. See text for definitions of models and measures; N.A.: not available.

[a]The formula where there are $n$ counties (i), is $\Sigma(|Y_{\text{model } i} - Y_{\text{census } i}|) / n$.
[b]The formula is $\Sigma\ [(|Y_{\text{model } i} - Y_{\text{census } i}|) / Y_{\text{census } i}] / n$.
[c]The formula is $\Sigma\ [(|P_{\text{model } i} - P_{\text{census } i}|) / P_{\text{census } i}] / n$.

SOURCE: Data from the Bureau of the Census.

- The performance of the four candidate models is similar, which is not surprising, given that they are variations of the same basic formulation. Thus, the range of the average absolute difference in the estimated number of poor school-age children (column 1) is from 268 children (model b) to 283 children (model d). The average county had about 2,500 poor school-age children for 1989, so that the average absolute difference ranges from 10.7 to 11.3 percent. The range of the average proportional absolute difference in the estimated number of poor school-age children (column 2) is somewhat larger, from 15.4 percent (model a) to 18.8 percent (model d).
- The log number models (a, b) have somewhat lower average absolute differences for estimates of numbers of poor school-age children than do the log rate models (c, d). This is expected because the estimates from the log rate models must be converted to numbers by use of population estimates of total school-age children, which themselves contain error (see below, "Use of Post-censal Population Estimates"). It was expected for the same reason that the log number models would have higher average absolute differences for estimates of proportions of poor school-age children than would the log rate models because population estimates must be used to convert the estimated numbers from the log number models to estimated proportions. However, model (a) shows lower and model (b) shows not appreciably higher average proportional absolute differences for estimates of poverty rates compared with the better log rate model (c)—see column 3 of Table 4-1.
- The four candidate models substantially outperform the three procedures (i-iii) that rely solely or largely on 1980 census data. For example, the *largest* average absolute difference for the four candidate models is 283 poor school-age children (11% of the average number) for the log rate (under 18) model (d), while the *smallest* average absolute difference for procedures (i-iii) is 380 poor school-age children (15% of the average number) for the procedure that assumes stable poverty shares within state (ii). The differences are even somewhat larger for the average proportional absolute difference for estimates of the number of poor school-age children: 18.8 percent for the worst candidate model, model (d), compared with 26.2 percent for the best procedure of these three, the procedure that assumes stable poverty rates within state with conversion (iii).
- The four candidate models also perform better than the procedure (iv) that averages 1980 census estimates with estimates from the log number (under 21) model (a) for 1989, although the differences are not large.

## Category Differences in Numbers of Poor School-Age Children

Table 4-3 shows the difference in the number of poor school-age children from the 1990 census for categories of counties for each of the four candidate models and two of the procedures that rely more heavily on the 1980 census—the stable shares procedure (i) and the averaging procedure (iv). The measure shown

TABLE 4-3 Comparison of Model Estimates with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989: Algebraic Difference by Category of County (in percent)

| | Model | | | | Other Procedures | | |
|---|---|---|---|---|---|---|---|
| Category | Log Number Under 21 (a) | Log Number Under 18 (b) | Log Rate Under 21 (c) | Log Rate Under 18 (d) | Stable Shares (i) | Average of Census and (a) (iv) | Number of Counties[a] |
| Census Division[b] | | | | | | | |
| New England | –2.9 | –2.9 | –2.9 | –2.9 | 35.9 | 7.8 | 67 |
| Middle Atlantic | –2.8 | –2.8 | –2.8 | –2.8 | 27.1 | 4.4 | 150 |
| East North Central | –0.2 | –0.2 | –0.2 | –0.2 | –2.8 | –5.6 | 437 |
| West North Central | 1.7 | 1.7 | 1.7 | 1.7 | –1.8 | –2.1 | 618 |
| South Atlantic | 0.5 | 0.5 | 0.5 | 0.5 | 14.8 | 8.1 | 591 |
| East South Central | –4.5 | –4.5 | –4.5 | –4.5 | 14.1 | 2.1 | 364 |
| West South Central | –2.7 | –2.7 | –2.7 | –2.7 | –18.1 | –6.3 | 470 |
| Mountain | 4.3 | 4.3 | 4.3 | 4.3 | –23.2 | –3.1 | 281 |
| Pacific | 6.5 | 6.5 | 6.5 | 6.5 | –21.3 | 0.2 | 163 |
| Metropolitan Status | | | | | | | |
| Central county of metropolitan area | 2.4 | 1.6 | –0.1 | –0.5 | –1.6 | 0.4 | 493 |
| Other metropolitan | –6.6 | –5.0 | 5.1 | 6.3 | 3.2 | 3.4 | 254 |
| Nonmetropolitan | –4.2 | –2.8 | –0.3 | 0.4 | 3.3 | –1.4 | 2394 |
| 1990 Population Size | | | | | | | |
| under 7,500 | –9.0 | –2.3 | –1.9 | 2.3 | 16.5 | 1.3 | 525 |
| 7,500-14,999 | –4.4 | 0.5 | 2.5 | 5.5 | 10.9 | 2.2 | 630 |
| 15,000-24,999 | –5.1 | –2.6 | 0.3 | 1.9 | 6.2 | –0.6 | 524 |
| 25,000-49,999 | –4.2 | –2.9 | 0.6 | 1.3 | 2.4 | –1.3 | 620 |
| 50,000-99,999 | –3.5 | –5.1 | –1.2 | –2.3 | –2.5 | –3.3 | 384 |
| 100,000-249,999 | –1.8 | –4.4 | –1.8 | –3.5 | –4.9 | –3.3 | 259 |
| 250,000 or more | 3.3 | 3.2 | 0.5 | 0.5 | –0.6 | 1.8 | 199 |

*continued on next page*

TABLE 4-3 Continued

| Category | Model | | | | Other Procedures | | |
|---|---|---|---|---|---|---|---|
| | Log Number Under 21 (a) | Log Number Under 18 (b) | Log Rate Under 21 (c) | Log Rate Under 18 (d) | Stable Shares (i) | Average of Census and (a) (iv) | Number of Counties[a] |
| 1980 to 1990 Population Growth | | | | | | | |
| Decrease of more than 10.0% | –1.9 | 0.6 | –3.4 | –1.9 | 9.1 | –3.4 | 444 |
| Decrease of 0.1-10.0% | –0.6 | –0.5 | –1.9 | –1.8 | 7.5 | –2.7 | 972 |
| 0.0-4.9% | –2.8 | –2.8 | –3.2 | –3.1 | 11.0 | –0.2 | 547 |
| 5.0-14.9% | 0.0 | –1.0 | 0.2 | –0.6 | 6.1 | 2.1 | 620 |
| 15.0-24.9% | 7.7 | 5.8 | 5.5 | 4.6 | –12.8 | 2.4 | 260 |
| 25.0% or more | –4.0 | –1.4 | 1.7 | 3.1 | –21.2 | 1.0 | 292 |
| Percent Poor School-Age Children, 1980 | | | | | | | |
| Less than 9.4% | –4.0 | –4.5 | 0.0 | 0.2 | 2.4 | –1.1 | 516 |
| 9.4-11.6% | –0.5 | –1.0 | –1.6 | –1.8 | –9.9 | –3.6 | 524 |
| 11.7-14.1% | 3.6 | 2.3 | 1.8 | 1.0 | –4.2 | 0.2 | 530 |
| 14.2-17.2% | 0.9 | 1.2 | –1.2 | –1.4 | –5.0 | –1.8 | 523 |
| 17.3-22.3% | 1.8 | 1.7 | 0.3 | –0.1 | 10.7 | 4.2 | 519 |
| 22.4-53.0% | –2.2 | 0.8 | 1.3 | 2.8 | 12.3 | 4.1 | 523 |
| Percent Hispanic, 1990 | | | | | | | |
| 0.0-0.9% | –3.4 | –3.3 | –1.6 | –1.5 | 10.7 | 0.2 | 1770 |
| 1.0-4.9% | 0.5 | 0.1 | 0.4 | 0.1 | 0.2 | –0.4 | 847 |
| 5.0-9.9% | –1.4 | –0.6 | –1.1 | –0.8 | 6.7 | 1.7 | 193 |
| 10.0-24.9% | 2.2 | 1.8 | 0.7 | 0.5 | –5.7 | 0.1 | 181 |
| 25.0-98.0% | 3.9 | 4.6 | 2.2 | 2.7 | –16.8 | –0.4 | 150 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Percent Black, 1990 | | | | | | | |
| 0.0-0.9% | –1.2 | 0.3 | 3.9 | 4.9 | –3.7 | –0.5 | 1446 |
| 1.0-4.9% | –0.7 | –2.0 | 1.3 | 0.5 | –6.3 | –2.9 | 615 |
| 5.0-9.9% | –2.9 | –2.5 | –0.7 | –0.6 | –8.4 | –1.8 | 294 |
| 10.0-24.9% | 2.0 | 1.2 | –1.0 | –1.3 | –2.6 | 0.2 | 381 |
| 25.0-87.0% | 1.0 | 1.7 | –1.8 | –1.4 | 16.5 | 3.7 | 405 |
| Persistent Rural Poverty, 1960-1990[c] | | | | | | | |
| Rural, not poor | –4.0 | –3.7 | –1.2 | –1.0 | 0.1 | –3.4 | 1740 |
| Rural, poor | –5.0 | –2.1 | 0.7 | 2.1 | 9.8 | 1.2 | 535 |
| Not classified | 1.7 | 1.2 | 0.3 | 0.0 | –1.2 | 0.7 | 866 |
| Economic Type, Rural Counties[c] | | | | | | | |
| Farming | –5.5 | –2.5 | –1.6 | 0.7 | 13.2 | 1.1 | 556 |
| Mining | –10.7 | –5.1 | –6.3 | –3.6 | –8.9 | –10.6 | 146 |
| Manufacturing | –6.2 | –5.9 | –1.7 | –1.0 | 12.1 | –0.2 | 506 |
| Government | 2.1 | –1.3 | 6.3 | 3.2 | –0.9 | 0.0 | 243 |
| Services | –3.9 | –3.0 | –1.8 | –1.2 | –5.8 | –4.3 | 323 |
| Nonspecialized | –3.7 | –1.0 | –0.1 | 1.4 | 2.2 | –1.5 | 484 |
| Not classified | 1.7 | 1.2 | 0.3 | 0.0 | –1.2 | 0.7 | 883 |
| Percent Group Quarters Residents, 1990 | | | | | | | |
| Less than 1.0% | –6.7 | –2.7 | 2.0 | 4.7 | –1.4 | 0.3 | 545 |
| 1.0-4.9% | 0.3 | 0.7 | –0.3 | 0.1 | –0.4 | 0.1 | 2187 |
| 5.0-9.9% | 2.3 | –4.4 | 0.5 | –5.2 | 7.8 | –0.8 | 299 |
| 10.0-41.0% | 14.2 | –3.2 | 7.4 | –7.5 | 1.8 | –2.2 | 110 |

*continued on next page*

TABLE 4-3 Continued

| Category | Model: Log Number Under 21 (a) | Model: Log Number Under 18 (b) | Model: Log Rate Under 21 (c) | Model: Log Rate Under 18 (d) | Other Procedures: Stable Shares (i) | Other Procedures: Average of Census and (a) (iv) | Number of Counties[a] |
|---|---|---|---|---|---|---|---|
| Status in CPS, 1989-1991 | | | | | | | |
| In CPS sample | 1.4 | 1.0 | –0.2 | –0.5 | –0.6 | 0.5 | 1028 |
| In CPS, no poor children 5-17 | –2.6 | –1.9 | 7.3 | 7.8 | 10.0 | 5.9 | 246 |
| Not in CPS sample | –4.1 | –2.8 | –0.1 | 0.6 | 0.6 | –2.3 | 1867 |
| Change in Poverty Rate for School-Age Children, 1980-1990 | | | | | | | |
| Decrease of more than 3.0% | 7.5 | 10.4 | 16.2 | 18.1 | 51.6 | 30.0 | 536 |
| Decrease of 0.1-3.0% | 2.1 | 1.9 | 3.1 | 2.9 | 29.2 | 12.1 | 649 |
| 0.0-0.9% | –2.6 | –0.8 | –0.4 | 0.5 | 4.3 | 3.1 | 272 |
| 1.0-3.4% | 3.8 | 2.2 | 3.4 | 2.6 | –5.1 | 0.2 | 621 |
| 3.5-6.4% | –1.2 | –2.4 | –3.8 | –4.3 | –14.3 | –8.3 | 532 |
| 6.5-38.0% | –7.2 | –5.2 | –8.7 | –7.8 | –25.2 | –14.5 | 523 |

NOTES: The census estimates are controlled to the CPS national estimate for 1989. The algebraic difference by category is the sum for all counties in a category of the algebraic (signed) difference between the model estimate of poor school-age children and the 1990 census estimate for each county, divided by the sum of the census estimates for all counties in the category. See text for definitions of models.

[a]3,141 counties are assigned to a category for most characteristics; 3,135 counties are assigned to a category for 1980-1990 population growth and 1980 percent poor school-age children; 3,133 counties are assigned to a category for 1980-1990 percent change in poverty rate for school-age children.

[b]Census division states:

New England: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
Middle Atlantic: New York, New Jersey, Pennsylvania
East North Central: Ohio, Indiana, Illinois, Michigan, Wisconsin
West North Central: Missouri, Minnesota, Iowa, North Dakota, South Dakota, Nebraska, Kansas
South Atlantic: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida
East South Central: Kentucky, Tennessee, Alabama, Mississippi
West South Central: Arkansas, Louisiana, Oklahoma, Texas
Mountain: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada
Pacific: Washington, Oregon, California, Alaska, Hawaii

[c]The Economic Research Service, U.S. Department of Agriculture, classifies rural counties by 1960-1990 poverty status and economic type. Counties not classified are urban counties and rural counties for which a classification could not be made.

SOURCE: Data from the Bureau of the Census.

is the algebraic difference by category, which is the sum for all counties in a category of the algebraic (signed) difference between the model estimate of poor school-age children and the 1990 census estimate for each county, divided by the sum of the census estimates for all counties.[12] Counties are grouped into five or six categories for each of 11 characteristics—those that were considered in the assessment of the county model regression output discussed above.[13]

The measure in Table 4-3 expresses model-census differences for groups of counties in terms of numbers of poor children, similar to the overall average absolute difference in column 1 of Table 4-2. However, the category difference is expressed as an algebraic measure in which positive differences (overpredictions) within a category offset negative differences (underpredictions). The measure is intended to identify instances of potential bias in a model's predictions. For example, the model may over(under)predict, on average, the number of poor school-age children in larger counties relative to smaller counties.[14]

If the census estimates are a reasonably accurate standard for comparison, sizable category differences between model and census estimates that are not explained by differences in the CPS and census measurement of poverty or another reason would be disturbing. They would indicate that the errors in the model estimates are not random errors (which occur in any set of estimates), but occur in part because the model systematically over(under)predicts poverty in certain types of counties. Indeed, bias, in terms of over(under)prediction for different types of counties, is arguably more important than the overall absolute difference in evaluating a model that is used repeatedly because there is the risk

[12]The formula for counties ($i$) in each category ($j$) is $\Sigma_i \, (Y_{\text{model } ij} - Y_{\text{census } ij}) \;/\; \Sigma_i \, Y_{\text{census } ij}$.

[13]In addition to the algebraic difference for each category for the four candidate models and four procedures, Appendix D shows for each of them the average proportional algebraic difference—that is, the category difference expressed in terms of percentage errors for counties instead of poor children (see Tables D-1 and D-2). Differences between the two measures can help identify particular types of counties within a category for which a model performs less well than others.

[14]However, an apparent bias identified in a single validation may be a one-time discrepancy that will not occur in other years for which a model is estimated. The panel and the Census Bureau considered another type of external validation to try to identify systematic or persistent prediction biases, in which estimates from the four candidate models of poor school-age children for categories of counties for 1989 and 1993 would be compared to weighted CPS direct estimates for those categories for the two periods, using 3 years of CPS data to form the weighted estimates in each case. (This analysis is not the same as the analysis of regression output described above, in which the residual from the model estimate for each county with sampled households in the CPS is compared to the direct estimate on the log scale.) However, there was not enough time to complete the analysis.

Comparisons with weighted CPS direct estimates have the advantage that they can be performed for multiple years. They have the disadvantage that the sample size for CPS estimates, even aggregated for 3 years, is small for some categories of counties, which makes the comparisons more uncertain than the 1990 census comparisons. Also, in analyzing the CPS comparisons, one must bear in mind that the model estimates are raked to the state estimates, which are developed from a single year of the CPS.

that the bias will operate for the same areas on each occasion.[15] Although one would not want to use a model that had a large overall absolute difference from the standard of comparison, a model that performed somewhat worse in overall terms but exhibited fewer and less severe biases than another model would be preferable to it.

The panel draws several general conclusions from Table 4-3 about the performance of alternative county models in predicting numbers of poor school-age children for categories of counties:

- The performance of the four candidate models is similar. However, the log number (under 18) model (b) performs somewhat better than the log rate (under 21) model (c), which in turn performs better than the other two, the log number (under 21) model (a) and the log rate (under 18) model (d).

Performance in this instance is evaluated principally in terms of the spread among the differences for categories of counties (the spread between the largest positive and negative category differences for a characteristic). A better performing model has a narrower spread for a greater number of characteristics than other models. As an example (see Table 4-3), the spread among the category differences for counties classified by percent group quarters residents is 5.1 percentage points for model (b), 7.7 percentage points for model (c), 12.2 percentage points for model (d), and 20.9 percentage points for model (a).

Also entering into the panel's judgment is consideration of the magnitude and pattern of differences: a better performing model has smaller differences from the census and exhibits fewer obvious patterns across categories than other models. Continuing with the same example from Table 4-3, there is no pattern to the category differences for counties classified by percent group quarters residents for model (b), whereas model (a) exhibits a strong monotonic pattern in which the number of poor school-age children is overpredicted for counties with higher percentages of group quarters residents relative to counties with lower percentages. Also, the magnitude of the category differences for counties classified by percent group quarters residents is small for model (b)—no difference is larger than 5 percent in either direction. In contrast, the category differences for model (a) are as high as 14 percent for one of the categories.

- There are characteristics for which some or all models exhibit poor performance in terms of the spread between the largest and smallest category differences, the pattern of the differences across categories, or the magnitude of the differences (see below, "Category Differences for Specific Characteristics"). There are also some characteristics for which all four models perform well: percent poor school-age children in 1980; percent black population in 1990; and whether a rural county was persistently poor from 1960 to 1990.

[15] A search for potential biases is also important to identify possible approaches to model improvement.

- The four candidate models perform better on most characteristics than the four procedures that rely more heavily on the 1980 census. This is generally true, as discussed below, even for characteristics on which the candidate models perform poorly. However, the averaging procedure (iv), which averages 1980 census estimates and estimates from model (a), performs reasonably well for many characteristics. In contrast, the stable shares procedure (i), which simply ratio adjusts the 1980 census estimates to the CPS national estimate for 1989, performs substantially worse than all of the models and other procedures on almost every characteristic.

## Category Differences for Specific Characteristics

Category differences from the 1990 census estimates are discussed below for characteristics for which Table 4-3 shows that some or all four candidate models exhibit poor performance in comparison with the census in estimating the number of poor school-age children: percent change from 1980 to 1990 in the poverty rate for school-age children; percent population growth from 1980 to 1990; 1990 population size; percent Hispanic population in 1990; percent group quarters residents in 1990; and census division.

### Percent Change from 1980 to 1990 in Poverty Rate for School-Age Children

All four candidate models show a pronounced pattern of overpredicting the number of poor school-age children in counties that experienced the greatest decline in the poverty rate for school-age children from 1980 to 1990 and, conversely, underpredicting the number of poor school-age children in counties that experienced the greatest increase in the poverty rate for school-age children in that period. The category differences are smaller for the log number models (a, b) than for the log rate models (c, d): the spread between the largest positive and largest negative differences is 15-16 percentage points for models (a) and (b) and 25-26 percentage points for models (c) and (d).

One would not expect any of the candidate models to perform particularly well in predicting the number of poor school-age children for the counties at the extremes of the distribution of change in the poverty rate from 1980 to 1990. This variable is closely related to the variable that the models are trying to estimate, and the process of fitting a regression line to all of the data will generally not result in good predictions for the extreme values of the distribution. In other words, one would expect the models to perform less well for counties that experienced the largest changes (increase or decrease) in the poverty rate for school-age children.

Despite the large differences for some categories of this characteristic, however, the four candidate models perform substantially better than the procedures that rely more heavily on the 1980 census—see Table 4-3. (See also Figure 4-1, which shows the category differences for percent change in the school-age poverty rate from 1980 to 1990 for the log number (under 21) model (a), the log number (under 18) model (b), the stable shares procedure (i), and the averaging procedure (iv).) The stable shares procedure performs very poorly: because it assumes the same proportional distribution of poor school-age children in 1989 as in 1979 (from the 1980 census), by definition it will miss any change in poverty rates that occurred over time. The procedure (iv) that averages the estimates from the 1980 census and the log number model (under 21) for 1989 performs better than the stable shares procedure but not nearly as well as the four candidate models (two not shown).

### Percent Population Growth from 1980 to 1990

All four candidate models tend to overpredict the number of poor school-age children in counties that experienced larger population increases from 1980 to 1990 relative to counties that experienced smaller increases or declines in population. The exception to a generally monotonic pattern is that the four models underpredict the number of poor school-age children for counties that experienced population increases of 25 percent or more relative to counties that experienced increases of 15-25 percent. The log number (under 21) model (a) has the largest spread in category differences for this characteristic of the four candidate models—12 percentage points between the largest positive and negative differences.

The stable shares estimation procedure (i) performs very poorly on this characteristic. In contrast to the four candidate models, it overpredicts the number of poor school-age children in counties that experienced declines or smaller increases in population from 1980 to 1990 relative to counties that experienced larger population increases. The spread between the largest positive and negative category differences for the stable shares procedure is 32 percentage points. The averaging procedure (iv) exhibits small differences for population growth categories (see Figure 4-2).

### 1990 Population Size

The four candidate models vary in their performance for counties classified by population size. The log number (under 21) model (a) tends to overpredict the number of poor school-age children in larger size counties relative to smaller size counties. The log number (under 18) model (b) and the log rate (under 21) model (c) do not show a particular pattern to the category differences for this characteristic, and the category differences are not large. The four candidate models perform better than the stable shares model (i), which relies solely on 1980 census data. However, the model (iv) that averages 1980 census estimates with

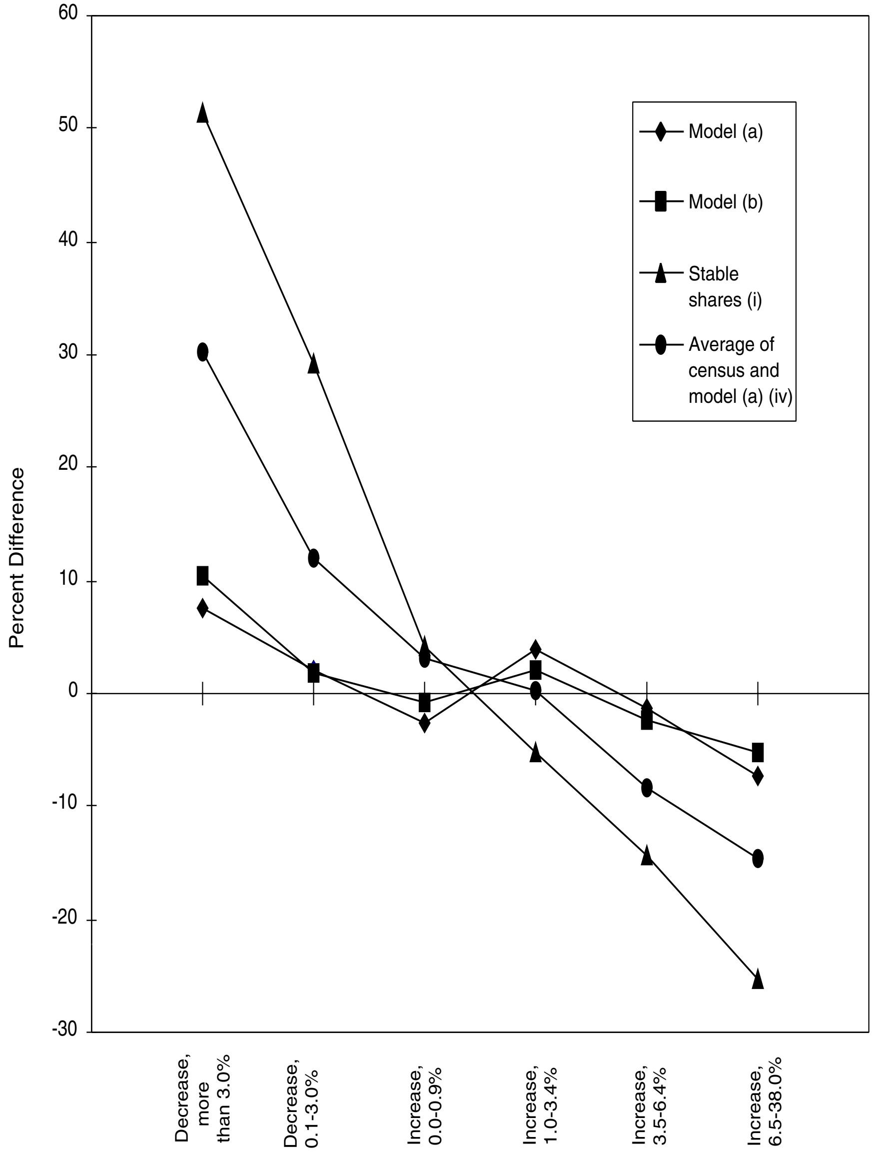


FIGURE 4-1 Change in poverty rate for school-age children, 1980-1990: Category differences from the 1990 census.

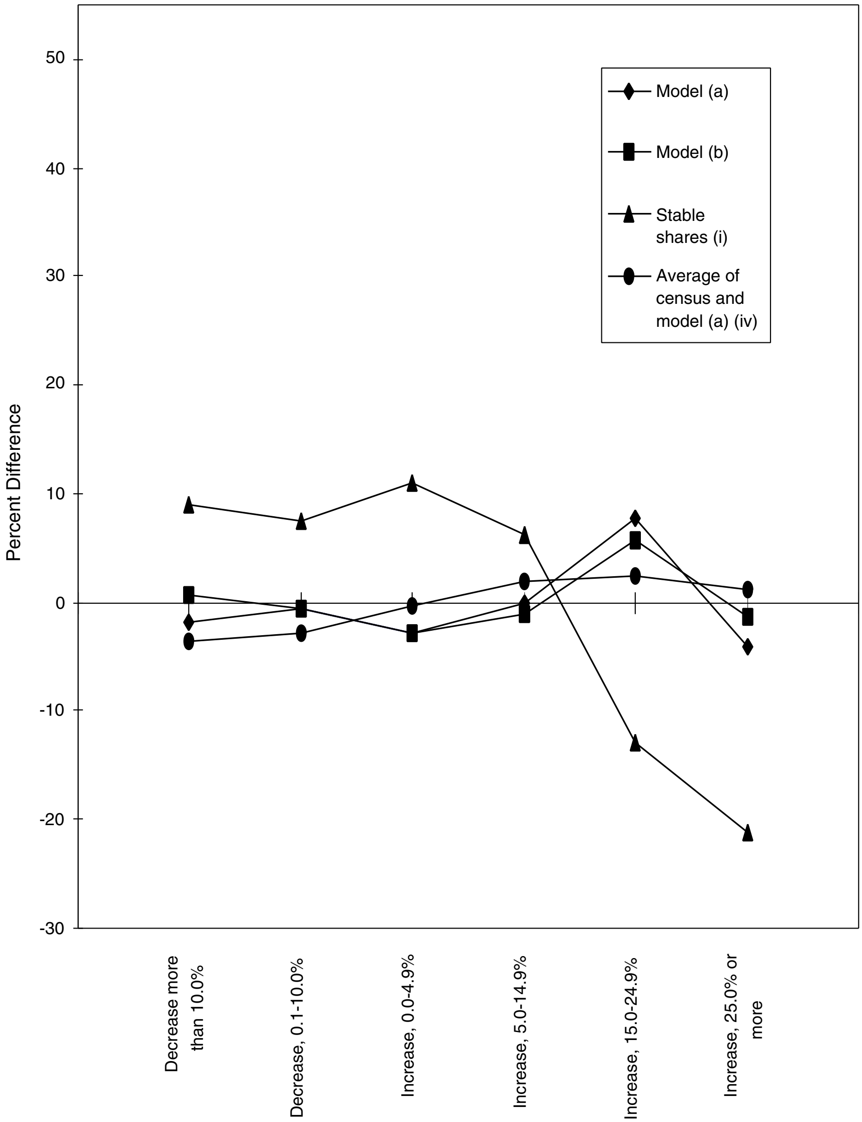


FIGURE 4-2 Population growth, 1980-1990: Category differences from the 1990 census.

estimates from the log number (under 21) model (a) for 1989 performs reasonably well in predicting numbers of poor school-age children for county population size categories (see Figure 4-3).

### Percent Hispanic Population in 1990

All four candidate models tend to overpredict the number of poor school-age children in counties with larger percentages of Hispanics relative to counties with smaller percentages, but the spread between the largest positive and negative differences is small. When the category differences are measured in proportionate terms for counties instead of in terms of numbers of poor school-age children, the models tend to *underpredict* the number of poor school-age children in counties with larger percentages of Hispanics (see Appendix D). The different patterns of the two category difference measures suggest that the models may perform differently for small counties with many Hispanics (primarily rural border counties) and large counties (cities).

The stable shares procedure (i), which relies solely on the 1980 census estimates, performs poorly on this characteristic. However, the averaging procedure (iv) performs reasonably well (see Figure 4-4).

### Percent Group Quarters Residents in 1990

The four candidate models vary in their performance for counties classified by percentage of group quarters residents. The log number (under 21) model (a) substantially overpredicts the number of poor school-age children in counties with larger proportions of group quarters residents relative to other counties. The log rate (under 21) model (c) shows a similar but less pronounced pattern of category differences. The log rate (under 18) model (d) shows the opposite pattern, in which it underpredicts the number of poor school-age children in counties with larger proportions of group quarters residents relative to other counties. In contrast, the category differences for the log number (under 18) model (b) are small and do not show a pronounced pattern across categories of this characteristic.

When the evident bias in predicting the number of poor school-age children in counties relative to their percentage of group quarters residents was discovered in the first round of evaluations of model (a), the Census Bureau developed model (b) to ameliorate the problem, with the desired result. The reasoning was as follows. In model (a), the two predictor variables—total child exemptions (assumed to be under age 21) from IRS tax records and the population estimate of the under 21 age group—are used together to estimate the number of people under age 21 in families that do not file tax returns. These families are assumed to be poorer, on average, than families that file tax returns. As can be seen from

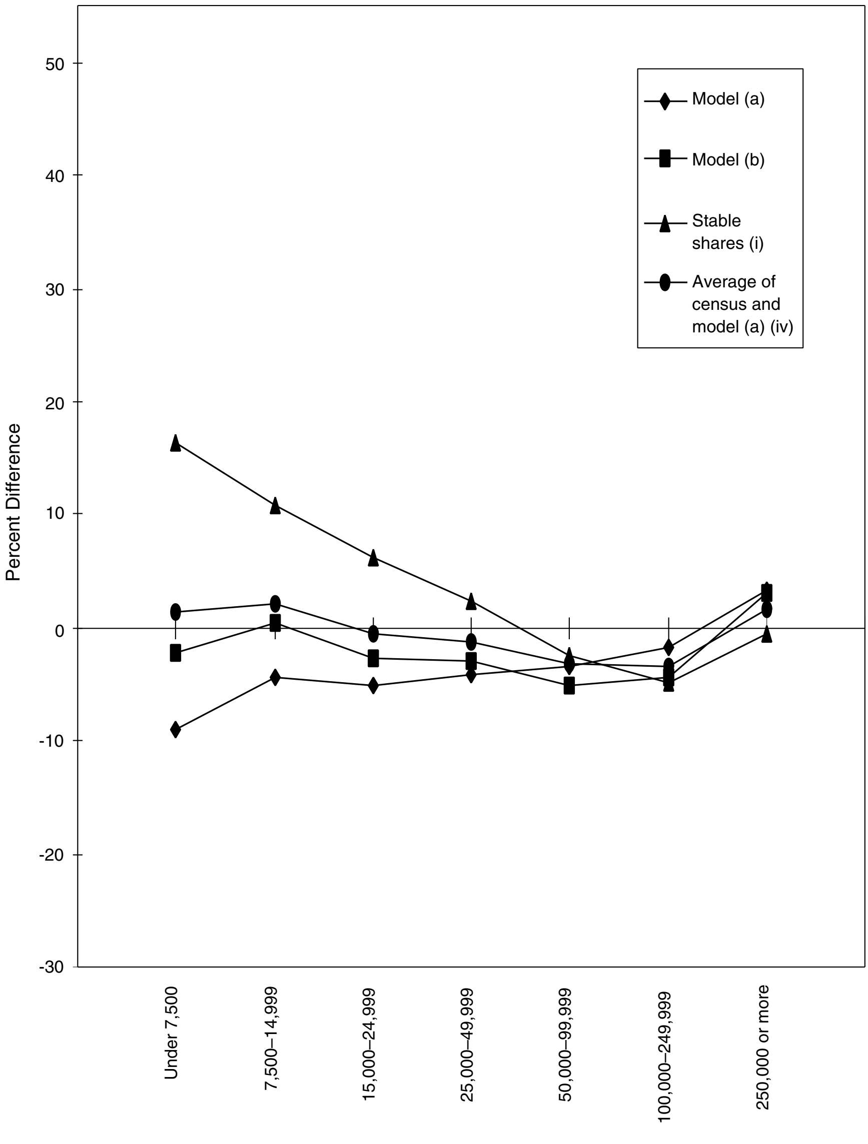


FIGURE 4-3 Population size, 1990: Category differences from the 1990 census.

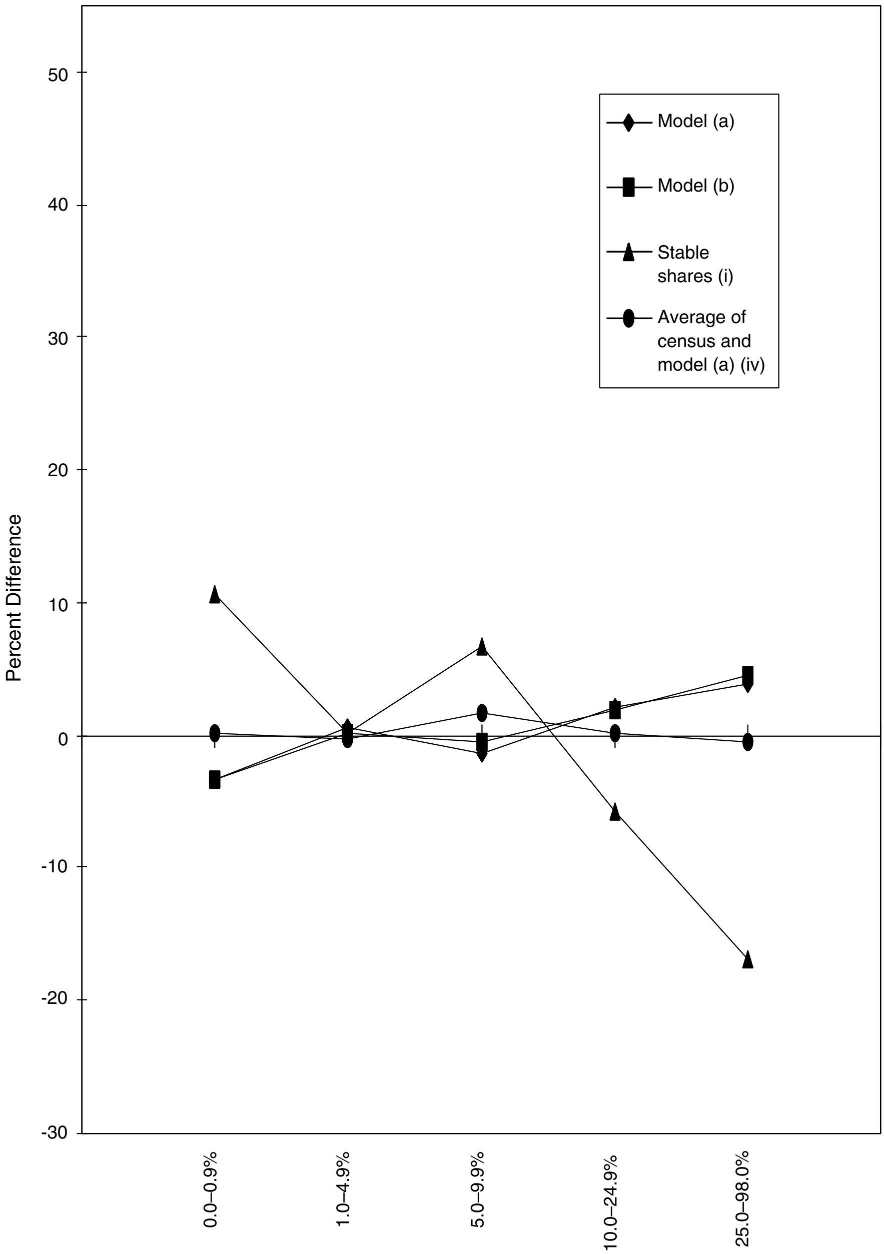


FIGURE 4-4 Percent Hispanic population, 1990: Category differences from the 1990 census.

Table 4-1, the regression coefficients for these two variables are of similar magnitude but of opposite sign.

However, in counties with large percentages of group quarters residents under age 21, primarily college students and military personnel, the relationship between the IRS variable and the population estimate may be distorted. To the extent that college students and military personnel under age 21 live in a county that is not the same as the county in which their parents reside or file tax returns on their own behalf, they will not be recorded as child exemptions in their county of residence. Consequently, there will be an overestimate of the number of people under age 21 in families that do not file returns in these counties and a corresponding overestimate, through the model, of the number of school-age children in poverty.

Model (b) replaces the population estimate for the under 21 age group as a predictor variable with the population estimate for the under 18 age group. This change not only eliminates the pattern of overpredicting the number of poor school-age children as a function of the percentage of group quarters residents that is so pronounced in model (a), but it also causes model (b) to perform better than model (a) on a number of other characteristics (e.g., population size). For reasons that are not clear, the under 18 formulation does not improve the performance of the log rate model; in fact, the log rate (under 18) model (d) generally performs worse than the log rate (under 21) model (c).

Interestingly, the procedures that rely more heavily on the 1980 census (i-iv)—even the stable shares procedure—perform reasonably well in predicting the number of poor school-age children for counties categorized by percent group quarters residents (see Figure 4-5).

## Census Division

All four candidate models show differences from the census for counties categorized by census division. In particular, the four models overpredict the number of poor school-age children in counties in the West (in the Mountain Division and, particularly, in the Pacific Division) relative to counties in other areas. The spread between the largest positive and negative differences is 11 percentage points.

Because the county estimates from the four candidate models are raked to the state estimates from the Census Bureau's state model, category differences on this characteristic must be attributable to the state model.[16] The state model

[16]The category differences are the same for all four candidate models because they are raked to the same set of state estimates; see Table 4-3. The average proportional category differences shown in Appendix D vary somewhat because they are calculated relative to each county's 1990 census estimated number of poor school-age children before being summed (see Table D-2).

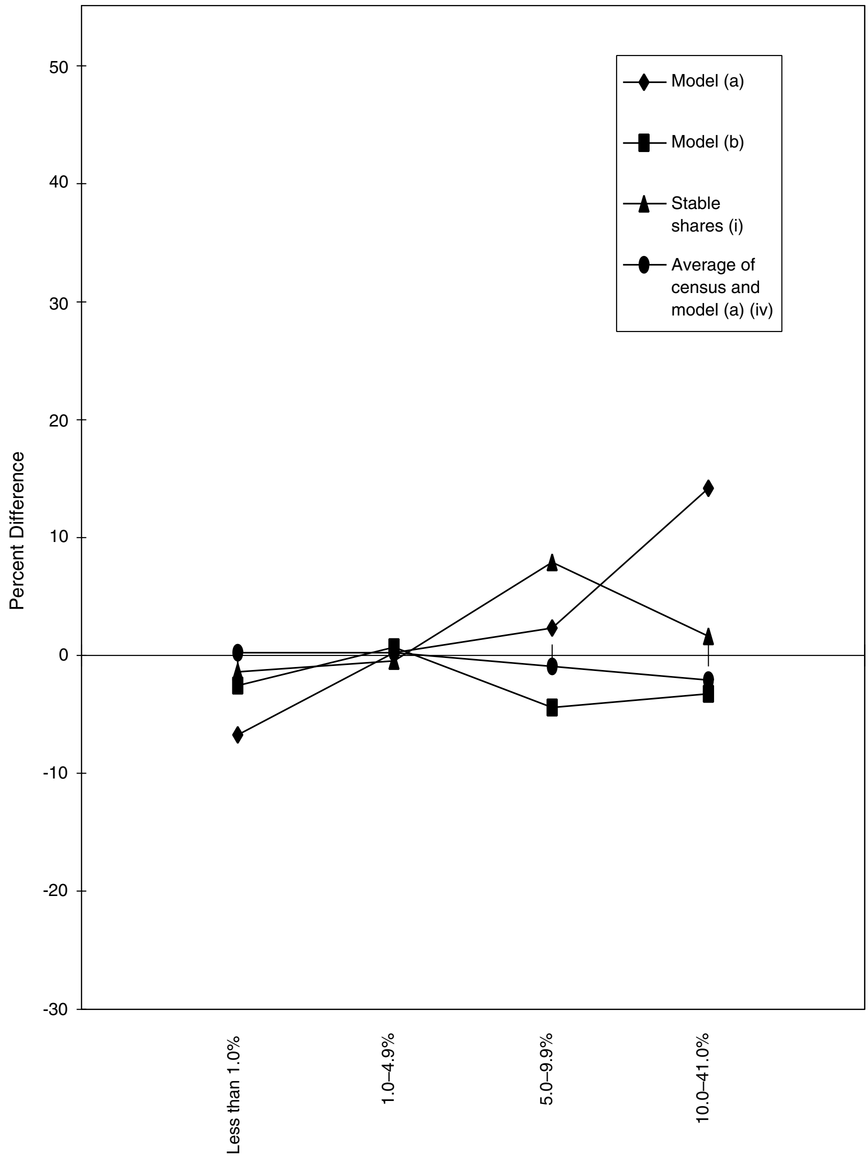


FIGURE 4-5 Percent group quarters residents, 1990: Category differences from the 1990 census.

needs to be more fully investigated to determine if category differences by area of the country are random occurrences in a single year or whether they persist across years, and why (see below, "State Model"). Yet the state raking procedure, which is done for the four candidate models and for the procedures that assume stable shares within state and stable rates within state (ii, iii), results in substantially better performance on this characteristic than the stable shares procedure (i). The averaging procedure (iv), which partly reflects the effects of the state raking, also performs better than the stable shares procedure (see Figure 4-6).

### Differences in Proportions of Poor School-Age Children

Examining the data for category differences in estimates of proportions of poor school-age children similar to those in Table 4-3, the panel reaches the same conclusions. Comparisons were performed only for the four candidate models, not for the other procedures.

First, the performance of the four candidate models is similar. Second, the two models that performed best in estimating the number of poor school-age children—log number (under 18) model (b) and log rate (under 21) model (c)—also perform best in estimating the proportion of poor school-age children. However, model (c) performs slightly better than model (b) in estimating proportions, while model (b) performs slightly better than model (c) in estimating numbers of poor school-age children. This reversal is expected because the use of population estimates for children aged 5-17, which themselves contain errors, to convert estimated numbers to estimated proportions from the log number models puts these models at a disadvantage for comparisons of proportions. Conversely, the use of population estimates for children aged 5-17 to convert estimated proportions to estimated numbers from the log rate models puts these models at a disadvantage for comparisons of numbers (see below, "Use of Postcensal Population Estimates").

Poverty rates (proportions poor) of school-age children enter the Title I allocation formulas as thresholds, so the panel and the Census Bureau examined the correspondence between each of the four candidate models and the 1990 census in classifying counties and school-age children into three poverty rate categories: 0 to 15 percent; 15 to 30 percent; and 30 percent or higher. A poverty rate of 15 percent or higher is an eligibility threshold for concentration grants; 15 percent and 30 percent poverty rates are thresholds for hold-harmless provisions of the allocation formulas (see Table 4-4); no comparisons were performed for the other procedures.

When there are two poverty rate categories, 0 to 15 percent and 15 percent or higher, each of the four candidate models performs equally well, assigning about 87 percent of the counties, which include about 92 percent of the poor school-age children, to the same category as the 1990 census (column 5, top half and bottom half of Table 4-4). When there are three poverty rate categories, 0 to 15 percent,

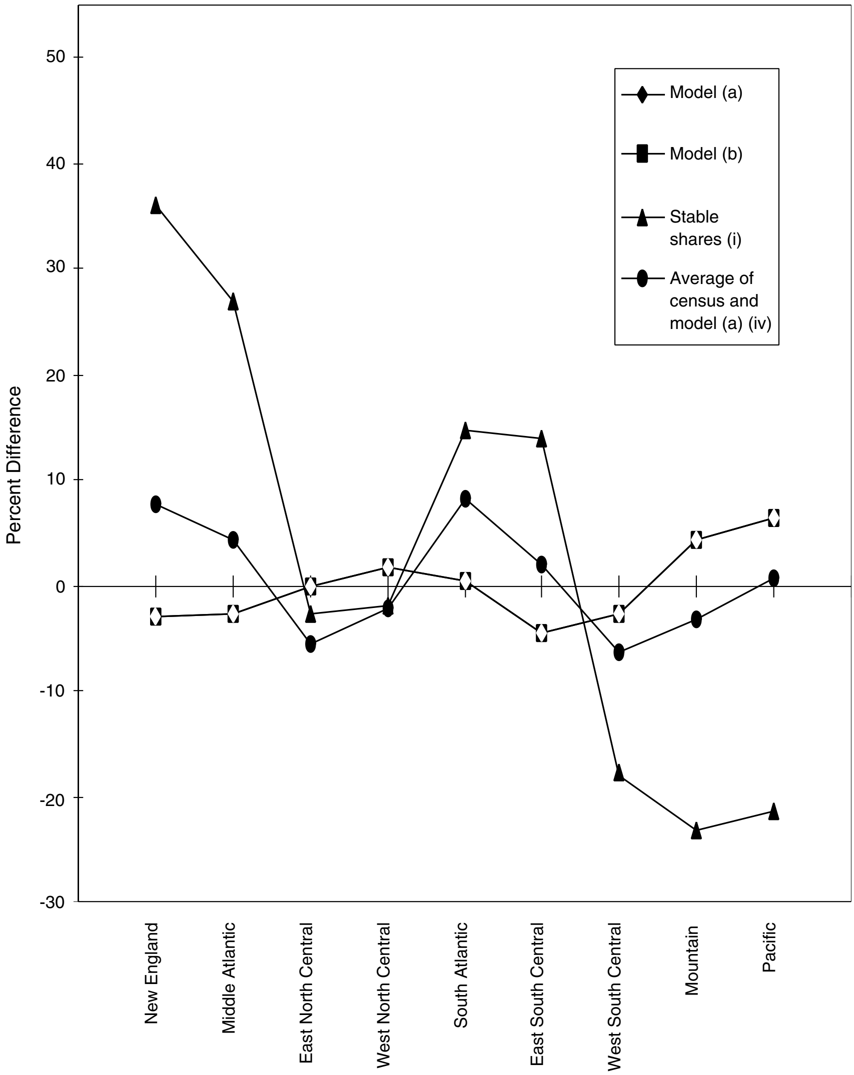


FIGURE 4-6 Census division: Category differences from the 1990 census.

TABLE 4-4 Agreement Between Model Estimates for 1989 and 1990 Census County Estimates for Proportions of School-Age Children in Poverty in 1989 (in percent)

| | Counties | | | | | | |
|---|---|---|---|---|---|---|---|
| | Model and Census Estimate in Same Poverty Rate Category | | | | Percent in Agreement | | |
| Model | Under 15% | 15% or More (1) | 15-30% (2) | 30% or More (3) | Under 15% and 15% or More (4) | Under 15%, 15-30%, and 30% or More (5) | (6) |
| (a) Log Number (Under 21) | 30.1 | 57.0 | 39.2 | 11.4 | 87.1 | 80.7 | |
| (b) Log Number (Under 18) | 30.5 | 57.1 | 38.4 | 11.9 | 87.6 | 80.8 | |
| (c) Log Rate (Under 21) | 28.8 | 58.6 | 40.1 | 12.9 | 87.4 | 81.8 | |
| (d) Log Rate (Under 18) | 28.4 | 58.6 | 39.5 | 13.0 | 87.0 | 80.9 | |
| | School-Age Children | | | | | | |
| | Model and Census Estimate in Same Poverty Rate Category | | | | Percent in Agreement | | |
| Model | Under 15% | 15% or More | 15-30% | 30% or More | Under 15% and 15% or More | Under 30%, 15-30%, and 30% or More | |
| (a) Log Number (Under 21) | 40.7 | 51.0 | 39.9 | 7.3 | 91.7 | 87.9 | |
| (b) Log Number (Under 18) | 40.9 | 50.7 | 38.5 | 7.3 | 91.6 | 86.7 | |
| (c) Log Rate (Under 21) | 40.1 | 51.5 | 41.0 | 7.6 | 91.6 | 88.7 | |
| (d) Log Rate (Under 18) | 40.3 | 51.1 | 40.3 | 7.5 | 91.4 | 88.1 | |

NOTE: Census estimates are controlled to the CPS national estimate for 1989.
SOURCE: Data from the Bureau of the Census.

15 to 30 percent, and 30 percent or higher, each of the four candidate models assigns about 81 percent of the counties, which include about 88 percent of the poor school-age children, to the same category as the 1990 census (column 6, top half and bottom half of Table 4-4).

### Summary

Keeping in mind the limitations of a single census-based validation opportunity, the panel concludes that the four candidate models perform substantially better in predicting the number and proportion of poor school-age children for counties for 1989 than the simple stable shares procedure (i), which relies solely on estimates from the previous (1980) census and the current (1989) CPS national total. Using the state model to rake the 1980 census county estimates for consistency with updated estimates of poor school-age children in each state, as is done in procedures (ii) and (iii), is an improvement over procedure (i). However, the four candidate models, which use a county regression model together with the state model, perform much better than procedures (ii) and (iii). Finally, the four candidate models perform better in many respects than procedure (iv), which averages the 1980 census estimates and the 1989 estimates from the log number (under 21) model (a), although this averaging procedure shows good performance on some characteristics. Overall, the comparisons with the procedures that rely more heavily on the 1980 census provide significant evidence in favor of a model-based approach for updated estimates of poor school-age children and against using estimates that derive solely or largely from the previous census.

The panel further concludes that, while the performance of the four candidate models in comparison with the 1990 census is broadly similar, when consideration is given to measures of overall absolute difference and differences for categories of counties, for estimates of numbers and estimates of proportions of poor school-age children, the log number (under 18) model (b) and the log rate (under 21) model (c) perform better than the other two. Comparing models (b) and (c), model (b) performs somewhat better, and the Census Bureau used this model to prepare the revised county estimates of poor school-age children in 1993. The comparisons also identify areas of performance of model (b) that deserve further examination in an ongoing research program to continue to improve model-based estimates of poverty for small geographic areas (see Chapter 6).

## CPS-CENSUS DIFFERENCES

A possible explanation of some of the category differences identified in the 1990 census comparisons just described may be, not that a model is in error, but that measurement of poverty differs systematically between the census and the CPS because of the many differences in data collection procedures (see National

Research Council, 1997:App. B). The Census Bureau performed chi-square tests to determine if there were significant differences between estimates from the March 1990 CPS and the 1990 census of the number of school-age children and the number and proportion poor in this age group in 1989 for county groupings (Fay, 1997).[17] More specifically, the tests determined if the ratios of the CPS and census estimates for categories of a characteristic, such as county population size, were significantly different from each other. The characteristics tested were those examined in the 1990 census comparisons.

The tests generally show inconclusive results. However, there is some evidence that, when compared with the 1990 census, the March 1990 CPS estimates higher numbers and proportions of poor school-age children in metropolitan counties and larger size counties relative to medium-size counties. (CPS estimates for small-size counties have low reliability because of the relatively small proportion of the population in such counties and the small number of these counties in the CPS sample.) Also, while not significant, a pattern is evident in which the March CPS, when compared with the 1990 census, tends to estimate higher numbers and proportions of poor school-age children in counties with higher percentages of Hispanic population. These results for population size and percent Hispanic population parallel the results from the 1990 census comparisons described above. They suggest that at least some portions of the category differences for the candidate models for these two characteristics arise from differences in the CPS measurement of poverty and are not due to model error as such. Whether similar CPS-census differences would be present for 1993 is, of course, not known.

## EXTERNAL EVALUATION: LOCAL ASSESSMENT OF 1993 COUNTY ESTIMATES

The panel performed another type of external evaluation of the original 1993 county estimates of poor school-age children—the use of local knowledge.[18] Using the original 1993 model estimates for all 3,143 counties in the United States, the analysis first sought to identify groups of counties for which the 1993 estimates seemed unusually high or low in relation to prior levels and trends (e.g., from 1980 to 1990) in the number and proportion of poor school-age children and known social and economic trends for these groups of counties. Then, local informants—including staff and members of local councils of government, eco-

[17]The March 1990 CPS estimates for the categories involved are direct estimates produced using the CPS weights.

[18]This evaluation was carried out at the University of Wisconsin-Madison by Dr. Paul Voss, a member of the panel, with the assistance of Richard Gibson and Kathleen Morgen (see Voss, Gibson, and Morgen, 1997).

nomic development authorities, welfare agencies, state demographic units, state data centers, and other agencies—were contacted to obtain their assessment of the reasonableness of the implied trends in poverty for school-age children given their knowledge of local socioeconomic conditions.[19]

## County Analysis

Changes in the number and proportion of poor school-age children implied by the 1993 estimates were examined for counties categorized by several characteristics, including: population size and metropolitan status; population change; percent immigrant population; college-dominated counties; reservation and Native American counties; for nonmetropolitan counties, whether predominantly agricultural; and several classifications by geographic location (e.g., state and the regions identified by the U.S. Department of Agriculture).

The analysis identified a number of categories of counties for which further investigation of the reasonableness of the 1993 estimates seemed warranted:

- Large metropolitan central city counties had a high implied percentage change in the number of school-age children in poverty between 1989 and 1993—42 percent. This change declined systematically with decreasing size for metropolitan counties and continued to decline to the most remote, rural nonmetropolitan counties, for which the implied change in the number of school-age children in poverty was –6 percent.
- Counties with higher levels of international immigration had higher implied increases in the number and proportion of poor school-age children.
- Counties with higher percentages of Native Americans had lower implied increases in the number and proportion of poor school-age children. There was no particular pattern for counties with reservations.
- Farm counties had an implied decline in the number and proportion of poor school-age children, while nonfarm metropolitan counties had an implied increase.
- When the country was divided into the 26 regions identified by the U.S. Department of Agriculture, several regions were identified on the extremes of change in the number and proportion of poor school-age children. High implied increases were found in the Northern Metropolitan Belt, the Florida Peninsula, the Southwest, Northern New England, Mohawk New York and Pennsylvania, Lower Great Lakes Industrial, Southern Piedmont, and the Northern Pacific Coast. Small implied increases were found in the Central Corn Belt, the Southern Appalachian Coal Region, the Coastal Plain Cotton Region, the Northern Great Plains,

[19]The discussion refers to "implied" trends because the Census Bureau's county model is not designed to directly estimate change over time.

and the Rockies, Mormon, Columbia River Region. The single region with an implied decrease in the number and proportion of poor school-age children was the Mississippi Delta.

Some of these implied changes are apparently related to the general effect of population size, discussed above. However, the findings in this regional analysis, in particular, suggested which states and counties to follow up in discussions with local officials.

## Local Input

When counties that share certain characteristics appeared also to share a common pattern of change in the number and proportion of poor school-age children, a variety of individuals with local knowledge were contacted. Initially, 70 individuals associated with state data centers or state data center affiliate units were contacted; they provided a series of responses and referrals to other state and local officials. In addition, 26 states that appeared to have a sizable number of counties that shared a common implied trend in poverty for school-age children were targeted for intensive contact.

The nature of responses varied considerably. In some states, the original 1993 county estimates released by the Census Bureau had not been examined, and there appeared to be little interest in discussing them. In other states, the estimates had been looked at, but the general admonitions about standard errors that accompanied their release had dampened interest in studying them in detail. In contrast, several states had carried out in-depth analyses of the estimates. Of the 26 states targeted for intensive follow up, 8 provided detailed explanations (supported by examples) of trends suggested by the original 1993 county estimates, and 7 more states provided in-depth responses supported by their own analyses.

Almost every state agency contacted expressed specific doubts about the original 1993 estimates for one or more counties—too high here, too low there. In general, however, there was no consensus that the trends implied by the original 1993 county estimates were wrong, even in states for which large numbers of counties experienced apparent declines in the number and proportion of poor school-age children. Of the 26 states, 21 provided explanations as to why the original 1993 estimates appeared to show poverty trends in a specific direction or why the direction of change is too difficult to know. The most common explanations included comments about the size of the county, its rural agricultural nature, the fact that it is a diverse metropolitan county, immigration from abroad, and economic growth or economic decline. Occasionally, reference was made to a military base, an Indian reservation, or a university as an explanation for an apparent trend in poverty for school-age children. In three states, concern was expressed about the role of food stamp program data in the estimation model, as these data were deemed to be unreliable.

In summary, a high level of concern was expressed by individuals with local knowledge about the statistical reliability of the original 1993 county estimates, which is largely due to the Census Bureau's own cautions in this regard, coupled with specific county estimates that seem on the basis of local knowledge to be highly doubtful. These concerns notwithstanding, no categories of counties were identified that experienced apparent trends in the number and proportion of poor school-age children between 1989 and 1993 that were not accepted by local informants. Although the trends for a few counties were not accepted locally, the analysis found no strong indicators of potential bias for groups of counties sharing common characteristics in the county model.

## STATE MODEL

The state model plays an important role in the production of county estimates of poor school-age children. Evaluations conducted of the state model include an internal evaluation of the regression output for 1989 and 1993 and an external evaluation through comparing 1989 estimates from the model with 1990 census estimates of the proportion of poor school-age children by state. The results in each case, which are summarized below, support the use of the model. However, the state model evaluations have been more limited than the county model evaluations, as alternative state model formulations have not been evaluated explicitly. Further evaluation of the state model would be useful (see Chapter 6), particularly to examine the relationship between the state and county models and what factors may underlie the variations in the state estimates from the state model and the state estimates formed by summing the estimates from the county models (see below, "State Raking Factors").

### State Model Regression Output

The state regression model is a poverty rate model with the variables not transformed (see Chapter 2). The analysis of the regression output for the state model for 1989 and 1993 examined the same assumptions that were examined for the four candidate county models. The analysis is somewhat less informative for the state model than for the county models for two reasons. First, few explicit alternatives were developed for the state regression model. A log rate model was developed, and comparisons of that model with the rate model demonstrated that the log rate formulation had no particular advantage. Work was also started on a multivariate state model.[20] However, no formal analysis of regression output

[20]A multivariate formulation could be advantageous not only for the state model, but also for the county model, as an extension of the bivariate formulation for which initial development work was carried out (see Chapter 6).

was performed for either of these alternatives or for other alternatives that were explored early on in the development of the model. Second, there are more than 3,000 counties but only 50 states, and states vary much less than counties with respect to poverty rates and other characteristics. Hence, comparisons for categories of states are less informative than comparisons for categories of counties, and some categories of states do not contain enough states for analysis.

Nonetheless, examination of the regression output for the state model helps assess the validity of its assumptions. Overall, the analysis finds strong support for the assumptions underlying the state model (see below); there is no evidence of significant problems with the model formulation (although there may be other models that fit just as well).

***Linearity*** Plots of standardized residuals against the four predictor variables in the state model—the proportion of child exemptions reported by families in poverty on tax returns, the proportion of people receiving food stamps, the proportion of people under age 65 who did not file a tax return, and a residual from the analogous regression equation using the previous census as the dependent variable—support the assumption of linearity. Furthermore, the standardized residuals, when plotted against the model predicted values, provide no evidence of the need for any transformation of the variables. This result helps justify the decision not to use the log transformation of the proportion of poor school-age children as the dependent variable.

***Constancy over Time*** Table 4-5 shows the regression coefficients for the predictor variables for the state model for 1989 and 1993. The coefficients for all four poverty rate predictor variables are positive in both years. Generally, the coefficients are similar for 1989 and 1993, with the exception that the coefficient of the residual from the previous census (column 4) is large and significant for 1993 but fairly small and not significant for 1989.

***Inclusion or Exclusion of Predictor Variables*** The standardized residuals for the state regression model were grouped into four categories for each of the following characteristics: census region; 1990 population size; 1980 to 1990 population growth; percent black population in 1990; percent Hispanic population in 1990; percent group quarters residents in 1990; and percent poor school-age children in 1979 (from the 1980 census). The distributions of the standardized residuals for each category were then displayed using box plots. For none of these box plots was there an obvious pattern to the standardized residuals across categories. The model slightly overpredicts the proportion of poor school-age children for large states in 1993 (i.e., the model estimates are somewhat higher than the CPS direct estimates for large states relative to other categories), but this pattern is not evident in 1989. The model also slightly overpredicts the proportion of poor school-age children for states with a moderate percentage of Hispan-

TABLE 4-5 Estimates of Regression Coefficients for the State Model for 1989 and 1993

| | Predictor Variables[a] | | | |
|---|---|---|---|---|
| Year | 1 | 2 | 3 | 4 |
| 1989 | 0.53<br>(.10) | 0.57<br>(.21) | 0.33<br>(.10) | 0.37<br>(.32) |
| 1993 | 0.31<br>(.11) | 0.98<br>(.22) | 0.52<br>(.13) | 1.36<br>(.39) |

NOTES: All predictor variables are in terms of rates. Standard errors of the estimated regression coefficients are in parentheses.

[a]Predictor variables: (1) ratio of child exemptions reported by families in poverty on tax returns to total child exemptions; (2) ratio of people receiving food stamps to total population; (3) ratio of people under age 65 who did not file an income tax return to total population under age 65; (4) residual from a regression of poverty rates for school-age children from the prior decennial census (1980 or 1990) on the other three predictor variables.

ics in 1989 and slightly underpredicts the proportion of poor school-age children for states in the West census region in 1993, but in neither case is the pattern observed for the other year. Therefore, there is no strong reason to suggest that these variables need to be incorporated in the state regression model.

***Normality, Homogeneous Variances, and Outliers*** The distribution of the standardized residuals from the state regression model appears to follow a normal distribution. Also, although there is less information available for the state model than for the county regression models, the residual plots and the box plots of the distributions of the standardized residuals against the categories of states show little evidence of any heterogenous variance. Finally, there is no evidence of outliers from examination of the residual plots or displays of the distributions of the standardized residuals from the state regression model.

### 1990 Census Comparisons

Fay and Train (1997) compare 1989 estimates of the proportion of poor school-age children from the state model with 1990 census estimates. They find that the differences between the model and census estimates are much smaller than the differences between the 1989 CPS direct estimates and the 1990 census estimates and considerably smaller than the differences between the 1980 census estimates and the 1990 census estimates. These findings, which are presented graphically in Fay and Train (1997), support the use of a model-based approach to producing updated state estimates of poor school-age children instead of relying on estimates from the previous census or from the CPS alone. Similarly, a formal hypothesis test performed for the state model (Fay, 1996) supports the conclusion that the model-based estimates for 1993 are preferable to estimates from the 1990 census.[21] Comparable evaluations have not been performed for alternative state models or for categories of states.

### State Raking Factors

The final stage in producing updated estimates of the number of poor school-age children for counties is to rake the estimates from the county model for consistency with the estimates from the state model. The raking procedure is clearly beneficial to the county estimates. Thus, the 1990 census comparisons for the two procedures (ii and iii) that raked the 1980 census estimates to the estimates from the state model for 1989 showed better performance than the stable shares procedure (i), which did not entail raking. Also, an evaluation that was performed of the original log number (under 21) model (a) found a smaller overall average absolute difference from the 1990 census when the county model estimates were raked to the state model estimates for 1989 than when the county model was used without raking (National Research Council, 1997:31).

On the assumption that a county model is performing well, one would expect the state raking factors to be tightly distributed around 1.0—that is, one would expect relatively minor differences between the estimates for states formed by summing the county estimates before raking and the estimates from the state model. However, the raking factors vary considerably across states. For example, the log number (under 18) model (b), which shows somewhat less variation than the other three candidate models, has raking factors that range from 0.86 to 1.31 in 1989 (two-thirds falling between 0.96 and 1.17) and from 0.84 to 1.29 in 1993 (two-thirds falling between 0.98 and 1.15). This degree of variation suggests that there may be state effects not captured in the county model, which, in turn, could possibly affect the behavior of the model in estimating poor school-age children for counties within states. Also, the state model uses 1 year of CPS data, while the candidate models use 3 years: this difference could contribute to the variation in raking factors and also to the fact that they average greater than 1.

Implementation of a fixed state effects formulation of the county model in which state indicator variables are included as predictor variables in the regres-

[21]The test assumes that the objective is to predict poverty rates that reflect the CPS measurement of poverty and not the decennial census measurement.

sion (see Chapter 3) widened rather than narrowed the range of the state raking factors. Technical reasons having to do with the transformation of the predicted log values of poor school-age children to estimated numbers probably explain the increased variation in the state raking factors with a fixed state effects model. However, further investigation of the state raking factors and how to account for state effects in the county model should be topics for research in the near term (see Chapter 6). The investigation should include consideration of whether there is any feature of the state model that might explain the variation in the raking factors.

## USE OF POSTCENSAL POPULATION ESTIMATES

The process for producing updated estimates of school-age children in poverty at the county level and the use of those estimates in the Title I allocation formulas require population totals by age in noncensus years for two purposes: as a variable in the county regression equation (population under age 18 or under age 21, depending on the model), and as the basis for computing the estimated number or proportion of poor school-age children, depending on the model specification (log number or log rate). Population totals by age are also required for the state model.

The Census Bureau's log number (under 18) model (b) produces estimates of the number of poor school-age children in each county. Because the Title I allocations require both numbers and proportions, the Census Bureau provides the Department of Education with population estimates for the 5-17 age group to use as denominators for calculating the proportion of poor school-age children.[22]

The Census Bureau currently develops county age estimates within the framework of total population estimates for counties and age estimates for states (see Appendix B). Briefly, in a process that begins anew with each decennial census, total population estimates for counties are developed by updating the population estimates for the preceding year with data on births, deaths, net immigration from abroad, and net internal migration. (Net internal migration is estimated from a year-to-year match of federal income tax returns for people under age 65 and from the change in Medicare enrollment records for people aged 65 and over.) Estimates are developed separately for the population over and under

[22]The population estimates of school-age children that accompany the 1993 county model estimates pertain to July 1994. In addition, the Census Bureau makes available on its web site estimated proportions of poor school-age children in which the denominators are estimates of *related* children aged 5-17 in each county. These estimates are developed by adjusting the estimates from the Census Bureau's population estimates program for the noninstitutionalized population aged 5-17 on the basis of the ratio of related children aged 5-17 to noninstitutionalized children aged 5-17 for each county in the 1990 census.

age 65 in households and in group quarters. County total population estimates are aggregated to form state total population estimates.

State estimates by single years of age are developed by similar demographic methods, in which the preceding year's estimate for each cohort (single year of age) is updated with data on births, deaths, and migration. (For people under age 65, net internal migration is estimated from school enrollment data.) The state estimates for single ages are then raked to equal the state population totals. Finally, county estimates by age are developed by ratio-adjusting the 1990 census county age estimates to both the updated county total population estimates and the updated state age estimates in an iterative proportional fitting (raking) procedure. This procedure assumes that the age distribution of each county within a state changes in the same manner as that state's age distribution.

The Census Bureau has an active program to develop and review the performance of its demographically based population estimates, including evaluating the estimates at 10-year intervals by comparing them with the decennial census as a measure of the true values. These comparisons provide an indication of the differences, but they are not complete measures of accuracy and precision because the standard (i.e., the decennial census) itself is flawed, notably from net population undercount, which varies by age group across time and place (see Robinson et al., 1993).

The Census Bureau's methods and data for producing postcensal population estimates have generally improved over time, but three patterns of differences, which are practically inevitable, continue to affect the county and state estimates (see Davis, 1994). First, the proportional differences of the estimates in comparison with the census are larger on average for small areas than for large ones. Second, the proportional differences tend to be larger for areas in which the population is changing rapidly than for areas that are more stable. Third, the proportional differences for age groups tend to be higher than those for the total population.

The Census Bureau recently completed an evaluation of the county estimates of total population and children aged 5-17 by comparison with the 1990 census for all counties and for categories of counties similar to the categories used in the 1990 census model evaluations described above (see Appendix B). The procedure to develop updated estimates for counties by age for 1990 was to ratio adjust the 1980 census county age estimates to 1990 county total population estimates and 1990 state age estimates. The overall average proportional absolute difference in the 1990 county estimates of the population aged 5-17 was 6.3 percent, unweighted by county population size, and 4.9 percent, weighted by size. By comparison, the overall average absolute difference in the 1990 county estimates of the total population was 3.6 percent unweighted and 2.3 percent weighted.

Population size markedly affects the accuracy of the estimates for children aged 5-17. For counties with more than 1 million people in 1990, the average proportional absolute difference in the estimate for this age group was 5.2 percent, but it was 12.4 percent for counties with fewer than 2,500 people. This relationship is expected given the likelihood that errors in the input data will be disproportionately greater for smaller counties than for larger counties.

In terms of bias measured by the average proportional algebraic difference for categories of counties, the population estimates procedure, in comparison with the 1990 census, tends to overestimate children aged 5-17 in larger counties relative to smaller counties and in metropolitan counties relative to nonmetropolitan counties. The estimation procedure also tends to underestimate children aged 5-17 in counties with larger percentages of group quarters residents, to overestimate children aged 5-17 in counties with larger percentages of blacks, and to underestimate children aged 5-17 in counties with larger percentages of Hispanics relative to other counties. However, the differences are small for each characteristic.

The issue in the context of Title I allocations is the extent to which differences from the census in the population estimates for children aged 5-17 affect the estimates of the proportion of poor school-age children from log number models (a, b), or how they affect the estimates of the number of poor school-age children from log rate models (c, d). In the aggregate, the use of population estimates to convert estimated numbers from log number models to estimated proportions adds about 1 percentage point to the overall average proportional absolute difference between the model estimates for 1989 and the 1990 census estimates (compare column 3 with column 2 of Table 4-2 for the two log number models). The use of population estimates to convert estimated proportions from log rate models to estimated numbers has even less effect overall (compare column 2 with column 3 of Table 4-2 for the two log rate models).

In addition, although a rigorous analysis was not done, there seems to be little systematic contribution of errors in the population estimates to category differences in the model estimates of poor school-age children from the 1990 census estimates (see Appendix D). For the three single-equation rate models that were examined for 1989 in the first round of evaluations, including the log rate (under 21) model (c), the use of population estimates instead of 1990 census estimates ("true values") to convert estimated proportions to estimated numbers of poor school-age children worsened the performance of the models for some characteristics (e.g., by increasing the spread between the largest negative and positive category differences compared with the census), improved their performance for other characteristics, and made essentially no difference for other characteristics. None of the category differences between the model estimates of poor school-age children developed with population estimates and those developed with 1990 census estimates was large.

The evaluations of the effects of the population estimates on estimates of poor school-age children relate to a 10-year period: the population estimates for 1990 were developed on the basis of 1980 census data updated with other sources. The 1994 population estimates that are used to convert estimated numbers to estimated proportions of poor school-age children in 1993 from the log number (under 18) model (b) were developed on the basis of 1990 census data. Because of the 4-year instead of 10-year period for updating, it is likely that errors in the 1994 population estimates are smaller than errors in the 1990 population estimates and that they have even smaller effects on the estimates of the number and proportion of poor school-age children.

# 5

# Recommendation for Title I Allocations for the 1998-1999 School Year

## BACKGROUND

On the recommendation of the Panel on Estimates of Poverty for Small Geographic Areas (National Research Council, 1997), the Department of Education allocated Title I funds to counties for the 1997-1998 school year by averaging estimates of poor school-age children from two sources, the 1990 census and the Census Bureau's original county model for 1993.[1] The panel's recommendation was designed to meet the need for an immediate decision on allocating funds given that there had not been time to thoroughly evaluate the county model and resulting estimates. The recommendation took advantage of the Census Bureau's work to develop model-based estimates that are more up to date than the census estimates but reduced the impact of possible limitations in the model. The panel further recommended a series of evaluations of the Census Bureau's model and alternatives to it, many of which the Census Bureau had begun but had not had time to complete.

Between June and October 1997 the Census Bureau carried out an extensive set of evaluations of its model and a range of alternative models with input and review from the panel. On the basis of those evaluations, the Census Bureau made some changes to its county model and provided to the panel and the Department of Education a revised set of 1993 county estimates of poor school-age children in late October.

[1] "County model" is used in the broad sense to include the entire estimation procedure; see Chapter 2.

The panel commends the Census Bureau staff for their work to investigate thoroughly the properties and pros and cons of a number of alternative county model specifications, some of which represented innovative and methodologically challenging formulations. The work was carried out under severe time constraints to meet the schedule requested by the Department of Education.

Thorough evaluation is a critical component of any estimation approach. The evaluation of county models for estimating the number and proportion of poor school-age children pinpointed strengths and weaknesses of alternative models and identified immediate solutions for some problems. The evaluation results also suggested useful avenues for future research and development (see Chapter 6).

## RECOMMENDATION

The Department of Education requested that the panel assess the revised 1993 county estimates of poor school-age children for use in Title I allocations for the 1998-1999 school year. (The alternative would be to use the average of the 1990 census estimates and the 1993 estimates from the Census Bureau's original county model for a second round of allocations, that is, not only for the 1997-1998 school year, but also for the 1998-1999 school year.)

The extensive evaluations of the model and alternatives to it that were completed between June and October 1997 provided the information that the Census Bureau needed to assess and refine its county model and for the panel to judge the suitability of the resulting revised estimates for the next round of Title I allocations. The panel's recommendation is based on its review of the county model evaluation results.

**The panel recommends to the Secretaries of Education and Commerce that the Census Bureau's revised 1993 county estimates of poor school-age children be used in the Title I allocations for the 1998-1999 school year.**

The revised estimates should not be averaged with estimates from the 1990 census, as was done for the allocations for the 1997-1998 school year.

The evaluation work initially considered a broad range of alternative formulations and then focused on four similar models that were considered practicable candidates for use in producing revised 1993 county estimates of poor school-age children by October 1997. The panel's recommendation follows from analysis of the final round of evaluations:

- The four candidate models are similar in their performance when evaluated internally in terms of the features of the underlying regression models. When evaluated externally by estimating each model for 1989 and comparing the results with 1990 census estimates of numbers and proportions of poor school-age children, the four candidate models also show broadly similar performance. However, considering all of the measures of model-census differences that were used in the 1990 census comparisons, two of the four models—the log number (under 18) model (b) and the log rate (under 21) model (c)—perform better than the other two models.[2] Furthermore, model (b), which is the model the Census Bureau used to produce the revised estimates of poor school-age children in 1993 that were provided to the panel in October 1997, performs somewhat better than model (c).

- When compared with the 1990 census, the Census Bureau's revised county model, model (b), performs substantially better than simple estimation procedures that are based solely or largely on data from the previous (1980) census. (The other three candidate models also perform better than the simple estimation procedures.) Moreover, the revised model performs better than a procedure in which 1989 county estimates are developed by averaging 1980 census estimates with 1989 estimates from the Census Bureau's original log number model (under 21)—the same sort of averaging procedure that was used for the 1997-1998 Title I allocations.

The conclusions from comparisons with census estimates are necessarily based on evaluations for a single year (1989) that use 1990 census estimates of poor school-age children for comparison purposes; there is no comparable basis of evaluation for 1993. Nevertheless, the performance of the Census Bureau's revised county model (and the other three candidate models) in comparison with the estimation procedures that rely more heavily on 1980 census estimates to predict poor school-age children in 1989 makes the panel confident that a model-based approach for 1993 is preferable to using the outdated 1990 census estimates or to averaging 1990 census estimates with model-based estimates for 1993. The major changes in the distribution of children in poverty that have occurred since 1989 make heavy reliance on 1990 census estimates for current allocations highly problematic. There was a 20 percent increase in the number of children in poverty between 1989 and 1993, and there is clear evidence that the geographic distribution of such children changed markedly in those years (National Research Council, 1997:Table 2-3).

The justification for use of the revised 1993 county estimates of poor school-age children in the Title I allocations for the 1998-1999 school year is clear. At the same time, continuing work on research, evaluation, and development for the

[2]The other two models are the log number (under 21) model (a), which the Census Bureau used to produce the original 1993 county estimates, and the log rate (under 18) model (d). The "under 21" and "under 18" designations refer to the specification of one of the predictor variables in the county regression model; see Chapter 3.

county model (and state model) is warranted to further improve the methods for producing updated county estimates of poor school-age children in the future (see Chapter 6). The evaluation work identified areas of potential improvement in which progress could be made in the near term—for example, in the estimation of the sampling error variance of the county regression model and in clarifying the relationship between the state and county regression models. The evaluation work also identified promising alternative specifications for the county model, such as the bivariate formulation, that will require longer term research to develop.

Over time, changes in the nature of the available data for model-based estimates will require evaluation and are likely to lead to modification of the county model. For example, changes in welfare programs may affect the comparability of food stamp program data across states and even across counties within states. Also, data from the 2000 census and the possibility of a continuing large-scale household survey in the next decade, the American Community Survey, will require rethinking the best approach for producing county estimates of poor school-age children. Consequently, research and development for the county estimates will need to continue even while the Census Bureau works on the challenging task of developing updated estimates of poor school-age children for school districts.

## SPECIAL CASE: PUERTO RICO

The Title I allocations include Puerto Rico, which is treated as a county equivalent. While the commonwealth's 1990 decennial census provides estimates for 1989, no estimates of Puerto Rican children in poverty can be made for 1993 from the Census Bureau's model because the appropriate IRS and food stamp program data are not available for Puerto Rico. The Census Bureau computed 1993 estimates for Puerto Rico from data collected in a Family Income Survey that was conducted in the commonwealth in February and March 1995. Several adjustments had to be made to produce the estimates of school-age children in poverty in 1993. The original 1993 estimates for Puerto Rico, which were averaged with 1990 census estimates for Title I allocations for the 1997-1998 school year, have not been revised.

The panel concluded in its first interim report that the approach adopted by the Census Bureau for producing 1993 estimates of poor school-age children in Puerto Rico seems a reasonable one given the data available, although there is limited information about the quality of the data (see National Research Council, 1997:App. F). Because there is no alternative at this time and for consistency with the treatment of U.S. counties, the panel recommends that the original 1993 estimates for Puerto Rico be used in the Title I allocations for the upcoming 1998-1999 school year. They should not be averaged with 1990 census estimates.

The Puerto Rico income survey was repeated in 1997 and is expected to be conducted at 2-year intervals in the future. It will presumably be the basis of updated estimates of poor school-age children in Puerto Rico for 1995 and later years. Further investigation should be carried out of the quality of the estimates, and their comparability with the model-based estimates for U.S. counties, to determine if there are ways in which the data and estimation procedures for Puerto Rico can be improved.

# 6

# Future Research and Development for County Estimates

It is important to continue a significant program of research and development for methods of estimating poverty for school-age children at the county level for three reasons. First, although there is clear justification for using the revised 1993 estimates of poor school-age children for Title I allocations for the 1998-1999 school year, the county (and state) model evaluations identified issues that warrant investigation to determine how to further improve the estimation procedures. Second, with a model-based approach, it is important to examine carefully the continued applicability of a model each time it is used. Third, research is needed to take account of likely future developments in the availability of data that have implications for the modeling effort.

For the immediate future, a pressing requirement for the Census Bureau is to produce poverty estimates for school districts. The task of developing updated estimates of poor school-age children for school districts is challenging in many respects. Just some of the factors that complicate the work (see Siegel, 1997) are the scarcity of relevant data (e.g., IRS and food stamp program data are not currently available for school districts); the small size of many school districts (66% of the 15,227 school districts in 1989-1990 had a 1990 census total population of less than 10,000); the variations among states in the ways in which school districts are defined (e.g., 26% of 1989-1990 school districts included certain grade levels only, and 27% of 1989-1990 school district boundaries crossed county lines); and the changes in school district boundaries that occur over time.

The panel looks forward to working with the Census Bureau on school district estimation, but specific methods for developing school district estimates of poor school-age children are not considered further in this chapter. Because improvements to the county model are likely to have important benefits for school district estimates, the research discussed below is also relevant for school district estimation.

## OVERVIEW OF RESEARCH NEEDS

Setting priorities for research on the county model should take account of the production schedule for updated small-area estimates of poor school-age children that are required, directly or indirectly, by the Improving America's Schools Act of 1994. The deadlines for the Census Bureau to deliver small-area estimates of poor school-age children to the Department of Education for use in Title I allocations are as follows for 1998-2002:

- May 1998 deliver county estimates for 1995
- October 1998 deliver school district estimates for 1995
- October 2000 deliver school district estimates for 1997
- October 2002 deliver school district estimates for 1999

The 1995 county estimates to be delivered in May 1998 will include total and poor school-age children. The 1995, 1997, and 1999 school district estimates to be delivered subsequently will include total and poor school-age children, as well as total population.[1] Although county estimates are not required for future years by the legislation, they are likely to be needed in order to have a system of estimates that is consistent for different levels of geography. Just as estimates from the state model are currently used to control the estimates from the county model, so it is likely that, for the foreseeable future, estimates from the county model will play an important role in producing estimates for school districts. Also, there is likely to be interest in state and county estimates of poor children for other important public policy uses, such as evaluating the effects of changes in welfare programs.

Research on methods to improve the 1995 county estimates of poor school-age children must be relatively straightforward, given the May 1998 deadline. Longer term research will be useful for improving the county estimates in connection with the school district estimates to be delivered in October 2000 and later. Priorities for longer term research should consider the important changes that are likely to occur in the availability of data for modeling over the 1998-2002 period and beyond, which include:

[1]Total population estimates are needed because the legislation includes a provision that Title I allocations for school districts that have less than 20,000 total population in a state may be aggregated, and the state may then reallocate the funds to school districts on the basis of data other than the Census Bureau's estimates.

- current and future changes to welfare programs and tax systems that may affect the comparability of IRS and food stamp program data;
- the income and poverty estimates for small areas that will be available from the 2000 decennial census long-form sample of about 17 million households (such estimates will likely be available in 2002 for counties but not until later for school districts); and
- the planned introduction of the American Community Survey (ACS) as a large-scale, continuing sample survey of U.S. households, conducted primarily by mail, that will provide estimates similar to those provided by the decennial census long-form sample, including income and poverty estimates for small areas. The ACS is currently being tested in 4 sites; it will be implemented in 40 sites in 1999-2001 for comparison with the 2000 census. For each year from 2000 to 2002, the ACS will sample about 70,000 households nationwide. Beginning in 2003, the ACS will sample 250,000 households each month throughout the decade, for an annual sample size of about 3 million households (see Alexander, Dahl, and Weidman, 1997).

Changes in welfare programs and the accompanying data systems (especially those resulting from the 1996 Personal Responsibility and Work Opportunity Reconciliation Act) will almost certainly affect the comparability of food stamp data over geographic areas. For example, legal immigrants, who are no longer eligible for benefits, are very unevenly distributed geographically. Comparability is an important assumption in both the county and state regression models, and, therefore, the way in which food stamp program data are used as a predictor variable in the models may need to be modified. Changes in the tax system could also affect the usefulness of IRS data for small-area poverty estimation.

The American Community Survey, when it is fully operational, will be an important component of any approach to providing updated estimates of poor school-age children for small areas. It is possible that several months (or years) of data from the ACS might be used to provide direct estimates of poor school-age children for small areas. Alternatively, ACS data could be used indirectly as a dependent variable in a model-based approach for counties and school districts, parallel to the manner in which the CPS data are currently used for counties.

However, given that each year of the CPS and the 2000 census will also provide information on poverty,[2] it will be important to find ways to use all three sources of information together, for multiple time periods (for the CPS and ACS), to produce the best small-area estimates. Furthermore, given that all three data

[2]If the ACS is implemented as planned, it is likely that the 2010 census and subsequent censuses will not include a long form and, hence, will not provide income and poverty information.

sources will have their own measurement biases[3] and that they are available for different time periods—the decennial census year, multiple years of the CPS March Income Supplement, and many months of the American Community Survey—it is unlikely that simply pooling estimates from the three data sources can be justified. Hence, some adjustment or modeling procedure will be needed. Such a procedure will have to take account of available information about the variances and biases of the estimates from each data source (including not only measurement errors, but also bias because a data source covered a different time period).

Continued research and development for measurement error and time-series models will be needed to develop effective multivariate models for small-area poverty estimates that use multiple data sources for multiple time periods. A specific research issue is to determine how best to use the 2000 census information, which has low sampling variance but possibly substantial measurement bias and which may be biased if the economic conditions during the census reference period differ markedly from the period for which estimates are needed.

How best to combine information from disparate data sources is a general problem that has received considerable attention recently in statistical research. To prepare for future improvements in small-area estimates of poverty, research should start now on combining census and CPS data. Research should begin on combining data from all three data sources as soon as sufficient data are available from the American Community Survey.

The research required to take account of changes in data sources, as well as some of the other research recommended by the panel, is time consuming and will likely require additions to the staff who are currently working on small-area income and poverty estimation at the Census Bureau. More generally, the production of small-area estimates is a major effort that involves data acquisition and review, database development, geographic mapping and geocoding of data, methodological research, model development and testing, and evaluation. Since the production of small-area poverty estimates supports a range of important public policy needs for federal, state, and local governments, it is essential that adequate staff and other resources be available for the estimation program. The Census Bureau should consider ways of augmenting staff resources by engaging experts from outside the Census Bureau, by making use of contracts, interagency personnel transfer agreements, the research fellowship program of the Bureau and the American Statistical Association, and other arrangements.

[3] The data collection methods for the census, CPS, and ACS differ in many respects, including the length of the questionnaire, the primary data collection technique (face-to-face interviews versus mail questionnaires), the definitions of variables, the reference period for income measurement, and editing and imputation methods. Any of these differences can lead to different measurement biases.

## SHORT-TERM RESEARCH PRIORITIES FOR THE COUNTY MODEL

The panel identified four types of research that should be pursued to determine if the current estimation procedure can be improved before the next delivery of updated county estimates of poor school-age children in 1995, scheduled for May 1998. These four areas are discussed further below: generalized variance function modeling of CPS sampling variances for counties; further investigation of the state model; further research and development for models that incorporate state effects in the county model; and further examination of factors associated with large category differences and characteristics of outlier counties. All of this work appears to be feasible in the short term.

If it turns out that this research leads to one or more changes in the county model, then the kinds of internal and external evaluations conducted for the four models that were candidates to produce revised 1993 county estimates of poor school-age children (see Chapter 4) should be repeated for the new model and close competitors, to ensure that the changes to the model have not introduced any new problems. Even if the model remains unchanged, there is still a need to conduct a full internal evaluation of the 1995 county model because it will use different data than the 1993 county model. Generally, evaluation work should be a regular part of the development of updated county and school district estimates of poor school-age children.

***Generalized Variance Function Modeling of CPS County Sampling Variances*** The total squared error, or residual variance, for the revised county model—log number (under 18) model (b)—comprises two components, the model error variance and the sampling variance of the dependent variable. These two components need to be estimated separately for the application of the model, particularly for determining the relative weights of the regression estimate and the direct estimate in the shrinkage procedure. The current approach for estimating these components is to assume that the model error variance from the 1989 regression equation with the dependent variable formed from 1990 census data is the same as the model error variance formed from 1993-1995 CPS data. The sampling variance is then obtained by subtraction from the total squared error. The sampling variance for a particular county is assumed to be inversely proportional to the CPS sample size in that county.

An alternative approach is to estimate the CPS sampling variances on the basis of direct calculations of these variances that take account of the clustered sample design within counties, and then use a generalized variance function for modeling the sampling variances. With this approach, the model variance is calculated by subtracting the sampling variance from the total squared error. It thus avoids the questionable assumption that the model variances for the 1989 census equation and the 1993 CPS equation are equal.

Census Bureau staff have already begun work on fitting a generalized variance function to the CPS sampling variances. This work has a number of benefits. It should lead to improved relative weights for use in the shrinkage estimation, although this is likely to have only a modest impact on the final estimates. It may reduce, or help explain, the variance heterogeneity of the standardized residuals from the county model as a function of the poverty rate and the CPS sample size; in particular, it may address the pattern of increasing standardized residual variances with CPS sample size. The work also may shed some light on why the methods of moments and maximum likelihood produce different estimates of the sampling variances with the current approach.

***Further Investigation of the State Model*** There are a number of research issues that remain concerning the state model. First, the possible bias of the county estimates in the Pacific Division (see Chapter 4), which is due to raking the county estimates to the state model, should be investigated. If the overprediction for the Pacific Division is due to CPS-census measurement differences or to sampling variation in the CPS for 1989, it is not necessary to make any adjustments in the model. However, if CPS-census differences or sampling variation are not the cause, work is needed to improve the state model.

Second, the current state model uses only one year of CPS data. It could be beneficial to use the information on poverty at the state level that is contained in the CPS samples for previous years. Multivariate modeling is one approach that may be used to incorporate CPS data from other years. Work along these lines has been initiated at the Census Bureau (see Otto and Bell, 1997).

Finally, the detailed evaluations that were carried out for alternative county models have not been carried out to the same extent for the state model. Additional evaluations could include examination of regression output for other years of the state model and external validation vis-à-vis the 1990 census for alternative versions of the state model, including comparisons for categories of states to the extent feasible.

***Further Research and Development for County Models that Incorporate State Effects*** The magnitude of the state raking factors for the county estimates is of concern to some panel members (see Chapter 4). When the sums of county estimates often diverge substantially from the corresponding state estimates, it is important to understand why. To the extent that the variation in the raking factors is due to problems of the county or state model—for example, if there are idiosyncratic state effects that the county model does not capture—then it may be possible to improve the county estimates by modifying one or both models. In particular, there could be substantial benefits from a county model that incorporates state effects, possibly through the use of a fixed or random state effects formulation.

The Census Bureau did some preliminary research on adding fixed state effects to alternative formulations of the county model (see Appendix A). The fixed state effects models had raking factors that varied even more widely than the factors for other models. Also, while the addition of fixed state effects reduced some nonrandom residual patterns in the regression output, a fixed state effects log number model (under 21) did not perform better than other models in comparison with the 1990 census estimates (see Appendices C and D). It is important to study more thoroughly the discrepancies between the state and county models and to try out various methods for incorporating state effects in the county model in a more integrated way, such as through a two-level nested model.[4] Another part of this work could be to examine the effect of using a single year of CPS data for the state model and 3 years of CPS data for the county model.

***Further Examination of Factors Associated with Large Category Differences and Characteristics of Outlier Counties*** The internal and external evaluations (see Chapter 4) demonstrated that the log number (under 18) model (b) was generally reasonably well behaved with respect to the estimates for various categories of counties. However, some of the residual patterns and category differences from the 1990 census are worth investigating further to determine if the regression model could be improved either through a modification of the model form or through the addition of predictor variables. For example, the standardized residuals of model (b) exhibited some unusual behavior for urban counties and for counties with a high percentage of Hispanics, and this was also the case with the other three candidate models.

More generally, it is important to consider any anomalies in the model output (such as the variation in the state raking factors, discussed above), to understand their cause and to take corrective action when necessary. In that regard, it is somewhat surprising that the four candidate models, which are very similar, are so sensitive to relatively minor changes in specification. Two examples are the large change in the estimated sampling variance that resulted from fitting the models with maximum likelihood estimation instead of the method of moments, and the benefits of using the estimated population under 18 rather than under 21 as a predictor variable for the log number model (b) but not the log rate model (d). To investigate these kinds of anomalies, it would be useful to explore the data set using graphical analysis tools to further assess if there are county clusters that behave unusually.

Finally, some of the category differences in the county models could be due to differences between the CPS and census measurements of poverty. Some

[4]One approach is to estimate two components of variance, one for state and one for county within state, for the model fitted to the 1990 census data. An exploratory analysis of state effects can be conducted by estimating a state variance component, using the residuals from the model fitted to the CPS data.

work has been done to understand these differences (see Chapter 4), but more could be done to compare direct census and CPS estimates for different geographic and socioeconomic subgroups of the population. Much more powerful analyses of CPS-census differences could have been obtained had a March 1990 CPS-1990 census match been conducted (as had been done in previous decades). From such a match, CPS and census responses for the same households could be compared.

Looking to the future, it would be wise to plan now for an exact match of the 2000 census and the March 2000 CPS. Given that the American Community Survey will be in the field in 2000 with a national sample of about 70,000 households, an exact match could also be performed of the ACS in that year and the 2000 census. These matches would provide a wealth of information about the three different income measurement systems and would also provide key inputs to the development of a CPS-census measurement error model. Such a model could help resolve remaining issues about the differences between the state and county models (e.g., the overprediction of the number of poor school-age children in the Pacific Division). Such a model could also provide information from which to determine how to use data from the 2000 census with the currently employed CPS-based estimation procedure to minimize discontinuities in the Title I fund allocations that are based on estimates for income year 1999.

## LONGER TERM RESEARCH AND DEVELOPMENT FOR THE COUNTY MODEL

In the medium and longer term, research on some other areas could likely result in improvements to the county model, perhaps as early as the October 2000 release of estimates of poor school-age children for 1997. Four such longer term research areas are multivariate approaches to county estimation; investigation of models that make use of all the counties in the CPS sample, including those with no sampled poor school-age children; examination of ways to reduce the time lag in the estimates; and improvements in small-area population estimates for school-age children.

***Multivariate Approaches to County-Level Estimation*** The Census Bureau proposed, as an alternative to the separate use of CPS and census county regression equations (with the census equation being used only to estimate the model error variance for the CPS model), a *bivariate* county regression model, in which the two dependent variables are the CPS and census estimates of poor school-age children. This formulation has some very real advantages. First, the internal evaluation of the regression output for the bivariate models indicates that they are as good as or possibly better than their single-equation analogues. Also, additional analysis of the bivariate and single-equation formulations showed the benefit of the bivariate approach (see Appendix C). Unfortunately, lack of administrative records data for 1979 prevented the Census Bureau from conducting an external evaluation of the bivariate models, and, therefore, given their novelty and relative lack of evaluation, the panel did not recommend them for the production of the revised 1993 county estimates of poor school-age children. Research into this approach should continue, including an external evaluation as soon as that is feasible using the 2000 census data.

Similarly, integrating multiple years of the March Income Supplement of the CPS into the estimation procedure by means of a multivariate model, as opposed to the current procedure of averaging the data for several years, may be advantageous. A multivariate model, with estimates from more than one CPS year and the census as dependent variables in a linear system of equations, might provide an effective way of using all of the available information. In the future this model could also incorporate data from the American Community Survey by adding equations for the estimates from that survey.[5]

***Investigation of Discrete Variable Models that Use Counties with No Sampled Poor School-Age Children*** When using a logarithmic transformation of the number or proportion of poor school-age children as the dependent variable in a regression model, all counties in the CPS sample for which none of the sampled households has poor school-age children—304 of 1,488 counties for the 1993 model—have to be removed from the regression analysis (see Chapter 2). The dropped counties are generally smaller counties with small CPS sample sizes.

Although there are reasons to believe that the current approach provides reasonable estimates for 1989 and 1993,[6] the exclusion of 20 percent of the counties in the CPS sample is a cause of concern. Also, the consistency of modeling for small and large counties observed for 1989 and 1993 may not characterize future years for which the county model is estimated. It is important to investigate the development of discrete variable regression models, such as Poisson regression or other forms of generalized linear models, that permit the inclusion of data for those counties that have no sampled families with children in poverty. Although a satisfactory approach may not be fully developed and tested by May 1998, when the next round of county estimates is to be completed, work should begin now on this topic.

---

[5]An alternative approach to a bivariate or multivariate model that could be investigated is a single-equation model in which estimates from more than one census and CPS year are included as predictor variables.

[6]Graphical analysis by the Census Bureau suggests that smaller counties are fairly well fit by the regression equation for both years. Also, comparisons with 1990 census estimates (see Chapter 4) show that the counties not in the CPS sample, which are typically smaller than counties in the sample, are well predicted by the county model. Finally, since the counties in the CPS sample that are excluded from the regression estimation generally have small samples, they would have less influence in any regression analysis.

There are two factors that complicate the development of discrete variable models in this context: the lack of fully developed shrinkage procedures for most models of this form and the treatment of CPS sampling variances. However, Markov chain Monte Carlo implementation of hierarchical models can be used to address the first issue, and, with additional research and development, can also probably address the second issue.

***Examination of Ways to Reduce the Time Lag of the Estimates*** The Title I fund allocations for the 1997-1998 school year were based on estimates of poor school-age children in 1993, and these estimates will also be used for the 1998-1999 school year allocations. It is important to explore the extent to which this time lag can be reduced. (The Census Bureau began some exploratory work on this topic in June 1997 but had to put it aside in order to complete the evaluations of the original 1993 county model and alternative models.)

One of the causes of the lag is the availability of food stamp data, which must be obtained from individual states in some instances and which are not available until almost 2 years after the year to which they refer. It might be possible to overcome this problem, without seriously harming the model performance, by using food stamp data for the year prior to the estimation year. Another possibility could be to control the estimates from the county model to the state model estimates for the latest of the 3 years of CPS data used in the county model, instead of to the middle year. These ideas and others need to be evaluated to determine if the lag between the time period of the estimates and the year of allocation of funds can be reduced.

***Improvements in Small-Area Population Estimates*** The Census Bureau has work under way, which should continue, to improve the procedures for estimating the population by age at the county level and to develop estimates of the total population and the school-age population for school districts. In addition, the current approach to producing population and poverty estimates for school-age children through separate estimation procedures may not make full use of the obvious correlation that the two kinds of estimates should have. Therefore, it would be useful to explore the possibility of a bivariate model of population and poverty for school-age children or of other methods that more fully integrate the development of these two sets of estimates.

# Appendices

# APPENDIX A
# Models for County and State Poverty Estimates

*William R. Bell, Statistical Research Division, Bureau of the Census*

This appendix reviews the models investigated by the Census Bureau for the county poverty estimates for children aged 5-17; the state model is also reviewed briefly. The same model forms can be used for poverty statistics for other age groups, with appropriately defined dependent and regression variables.

## NOTATION

The following notation is used in the estimation program:

- $y_{it}$ = CPS 5-17 poverty estimate for county $i$ in year $t$;
- $\text{Cen}_i$ = previous census estimate for county $i$ (where necessary, a specific census is distinguished by writing $\text{Cen90}_i$ or $\text{Cen80}_i$);
- $Y_{it}$, $Z_i$ = "true" quantities estimated by $y_{it}$ and $\text{Cen}_i$ (i.e., $Z_i$ is not assumed to be true poverty, since the census could be biased relative to CPS);
- $e_{it}$, $\epsilon_i$ = sampling errors in $y_{it}$ and $\text{Cen}_i$, assumed independent $N(0, v_e/n_{it})$ and $N(0, c_i)$, with $c_i$ and $n_{it}$ known, and $v_e$ a parameter to be estimated;
- $n_{it}$ = CPS sample size (number of households) in county $i$ in year $t$;
- $\mathbf{x}_{it}$, $\mathbf{x}_{i,89}$ = vectors of a constant term (i.e., 1) and regression variables from administrative records for county $i$ in income years $t$ and 1989, respectively;
- $\beta$, $\eta$ = corresponding vectors of regression parameters.

The CPS data that are modeled are for income year ($t$) 1993 or 1989 (for CPS samples taken in March 1994 and March 1990, respectively). The census data modeled are from the 1990 census and are for income year 1989. The 1980 census data (for income year 1979) enter SAIPE models as regression variables in the equation for the 1990 census data but are not themselves the dependent variable in any model (because the corresponding regression variables $\mathbf{x}_{i,79}$ are not available.)

Note that $y_{it} = Y_{it} + e_{it}$ and $\text{Cen}_i = Z_i + \epsilon_i$. The nature of $Y_{it}$ and $Z_i$, and their estimators, $y_{it}$ and $\text{Cen}_i$, varies. They can be log(numbers of poor), log(poverty rates), or unlogged poverty rates, depending on the model. Similarly, $\mathbf{x}_{it}$ and $\mathbf{x}_{i,89}$ vary over models. These variations are noted below for the specific models.

The CPS estimates $y_{it}$ and sample sizes $n_{it}$ are 3-year "averages" of CPS estimates centered on year $t$. The specific formulation depends on whether log(numbers of poor children) are being modeled, as opposed to either child poverty rates or their logarithms (see below for details). Given that $y_{it}$ involves a 3-year average, the corresponding "sample size" $n_{it}$ is defined by counting the number of households in sample in county $i$ in each year of the average ($t - 1$, $t$, $t + 1$) and adding the three numbers together. For counties with a CPS sample in only 2 of the 3 years, $y_{it}$ is defined from just a 2-year average, and the corresponding $n_{it}$ is defined by summing the households in sample for the 2 years. For counties with a sample in just 1 of the years, the estimate and sample size for just that year are used.

## MODELS

### SAIPE Model for Log Number Poor

Let $y_{it}$ and $\text{Cen}_i$ denote CPS and census estimates of log(number of poor related children, 5-17). The 1993 SAIPE model (using CPS data for income year 1993) is

$$y_{it} = (\mathbf{x}'_{it}\beta + \gamma \text{Cen90}_i + w_{it}) + e_{it} \tag{1}$$

$$\text{Cen90}_i = (\mathbf{x}'_{i,89}\eta + \tilde{\gamma}\text{Cen80}_i + \tilde{z}_i) + \epsilon_i. \tag{2}$$

The model errors $w_{it}$ and $\tilde{z}_i$ are both assumed *i.i.d.* $N(0, \sigma_w^2)$ and independent of each other.[1] The basic regression variables $\mathbf{x}_{it}$ are defined below. Recall that $e_{it}$ and $\epsilon_i$, the sampling errors in $y_{it}$ and $\text{Cen90}_i$, are assumed independent $N(0, v_e / n_{it})$ and $N(0, c_i)$, with $c_i$ and $n_{it}$ known, and $v_e$ a parameter to be estimated.

[1]Assuming $w_{it}$ independent of $\tilde{z}_i$ is not entirely necessary, but serves as a partial justification for fitting equations (1) and (2) separately. The normality assumption stated here and for other models is also not entirely necessary, as the model fitting and smoothing procedures used can be justified without it.

The unknown parameters to be estimated in (1) and (2) are thus the regression parameters $\beta$, $\gamma$, $\eta$, and $\tilde{\gamma}$; the common model error variance $\sigma_w^2$; and the sampling error variance parameter $v_e$. Decennial census sampling error variances for estimates of number of poor are available from published formulas (generalized variances). If $R_i = \exp(\text{Cen90}_i)$ is the census estimated number of poor, then from a Taylor series linearization, $c_i$, the sampling error variance in $\text{Cen90}_i$, is approximately

$$c_i = \text{Var}(\in_i) \approx \text{RelVar}(R_i) \equiv \text{Var}(R_i) / R_i^2. \tag{3}$$

Actually, a slight refinement of (3), based on properties of the lognormal distribution was used, as described by Fisher and Siegel (1997). Practically speaking, the results are not materially different from (3).

The key distinguishing feature of the SAIPE model is the use of the previous census data as a regression variable—the $\gamma\text{Cen90}_i$ term in (1) and the $\tilde{\gamma}\,\text{Cen80}_i$ term in (2). This SAIPE model form contrasts with the bivariate model form, discussed in the next section. In the SAIPE model form the model error variance, denoted here by $\sigma_w^2$, can be essentially thought of as $\text{Var}(Y_i \mid \mathbf{x}_i, \text{Cen90}_i)$, which differs from the model error variance for the bivariate model form, $\sigma_u^2 = \text{Var}(Y_i \mid \mathbf{x}_i)$. The two are not comparable; one would expect $\sigma_w^2 < \sigma_u^2$.

The 1989 SAIPE model (using CPS data for income year 1989) is

$$y_{it} = (\mathbf{x}_{it}'\beta + \gamma\text{Cen80}_i + w_{it}) + e_{it} \tag{4}$$

$$\text{Cen90}_i = (\mathbf{x}_{i,89}'\eta + \tilde{\gamma}\text{Cen80}_i + \tilde{z}_i) + \in_i, \tag{5}$$

with $t = 1989$. Notice that $\mathbf{x}_{it} = \mathbf{x}_{i,89}$, and the regression variables in (4) and (5) are the same. The regression parameters, $(\beta, \gamma)$ and $(\eta, \tilde{\gamma})$, are still allowed to be different, however. The same assumptions as above are made about the model errors. Assuming that $w_{it}$ and $\tilde{z}_i$ are independent makes less sense here, since both equations refer to the same year and $\text{Cen90}_i$ does not enter (4) as a regression variable. Fortunately, this assumption is unnecessary. Since (4) and (5) contain "identical explanatory variables," regression fitting of these two equations separately produces the same results as fitting them jointly (Theil, 1971:309-310). Finally, notice that the second (census) equations of both the 1993 and 1989 SAIPE models—(2) and (5)—must be the same. Although it might be more appropriate for the 1989 model to replace (5) by the corresponding equation for $\text{Cen80}_i$, this cannot be done because the required regression variables $\mathbf{x}_{i,79}$ are not available.

For this and other models of log(number poor), the CPS estimates $y_{it}$ are defined using 3 years of CPS data for each county $i$ as follows:

$$y_{it} = \log([\text{3-yr weighted avg poverty rate}] \times [\text{3-yr weighted avg poverty universe}]). \tag{6}$$

The weights given to data from years $t - 1$, $t$, and $t + 1$ for the weighted averages in (6) are proportional to the numbers of interviewed housing units in county $i$ that contain at least one child aged 5-17 for the year in question. The CPS poverty rate in (6) for county $i$ in year $j$ (= $t - 1$, $t$, $t + 1$) is

$$\frac{\text{estimated number poor related children 5–17 in county } i, \text{ year } j}{\text{estimated total related children 5–17 (CPS poverty universe) in county } i, \text{ year } j} \tag{7}$$

Note that the second term in (6) is the 3-year weighted average of the denominators in (7) for $j = t - 1, t, t + 1$. The CPS poverty universe, and the number of poor related children aged 5-17, are estimated from CPS data for each year using CPS weights modified to make each county "self-representing."

For counties with a CPS sample in only 1 or 2 of the 3 years, the values for only that year, or for the 2-year average corresponding to (6), are used. For counties with no poor children observed in the CPS sample, the direct CPS estimate of the number of poor children is 0. Since logarithms cannot be taken when the direct estimate is 0, $y_{it}$ is not defined, and these counties must be dropped from the model fitting. The same problem arises with the census data, though only for a few counties.

The basic regression variables, $\mathbf{x}_{it} = (x_{0it}, ..., x_{4it})'$, are defined as follows, all but $x_{0it}$ derived from tabulating certain data for each county $i$:

$x_{0it}$ = 1 (constant term)
$x_{1it}$ = log (number of IRS dependent child tax exemptions on tax returns with income below poverty);
$x_{2it}$ = log (number of food stamp program participants) (from USDA);
$x_{3it}$ = log (resident population aged 0-21);
$x_{4it}$ = log (number of IRS total dependent child tax exemptions). (8)

More recently, Census Bureau analysts have experimented with changing the age limits defining $x_{3it}$ to 0-17. This removed some bias found in evaluations and regression diagnostics for counties with high group quarters populations (usually because of college dorms and military barracks).

### Bivariate Model for Log Number Poor

Let $y_{it}$ and $\text{Cen}_i$ denote estimates of log(number of poor), as above. The bivariate model form is

$$y_{it} = (\mathbf{x}'_{it}\beta + u_{it}) + e_{it} \tag{9}$$

$$\text{Cen}_i = (\mathbf{x}'_{i,89}\eta + z_i) + \epsilon_i. \tag{10}$$

The model errors $u_{it}$ and $z_i$ are both *i.i.d.* $N(0, \sigma_u^2)$, with $\text{Cov}(u_{it}, z_i) = \sigma_{uz} = \rho\sigma_u^2$ constant over $i$. This is the "constrained" bivariate model. The "unconstrained" bivariate model, allowing $\text{Var}(z_i) \equiv \sigma_z^2 \neq \sigma_u^2$, was investigated and found to produce unreasonable results, and it is not considered further here. As above, the sampling errors $e_{it}$ and $\epsilon_i$ are assumed independent $N(0, v_e / n_{it})$ and $N(0, c_i)$, with $c_i$ and $n_{it}$ known, and $v_e$ a parameter to be estimated. Parameters in (9) and (10) to be estimated are thus the regression parameter vectors $\beta$ and $\eta$; the common model error variance $\sigma_u^2$; the model error correlation $\rho$; and the sampling error variance parameter $v_e$.

Note that the bivariate model form differs from the SAIPE model form in that it does not include the previous census data as a regression variable, and it also allows the model errors to be correlated. These two differences in model form are related. In fact, by making a linear transformation, one could replace (9) by

$$y_{it} = [\mathbf{x}'_{it}\beta + \gamma_i(\text{Cen}_i - \mathbf{x}'_{i,89}\eta) + \tilde{w}_{it}] + e_{it}, \tag{11}$$

where

$$\gamma_i = \sigma_{uz} / (\sigma_u^2 + c_i) = \rho\sigma_u^2 / (\sigma_u^2 + c_i) \tag{12}$$

$$\begin{aligned} \text{Var}(\tilde{w}_{it}) &= \sigma_u^2 - \sigma_{uz}^2 / (\sigma_u^2 + c_i) = \sigma_u^2[1 - \rho\gamma_i] \\ \text{Cov}(\tilde{w}_{it}, z_i) &= 0. \end{aligned} \tag{13}$$

Replacing (9) by (11) makes the bivariate model form look more like the SAIPE model form, in that both now have the census data on the right-hand side of the CPS equation, and the model errors of the two equations are now uncorrelated. The two differences between (11) and (1) are that (11) uses the regression residuals $\text{Cen}_i - \mathbf{x}'_{i,89}\eta$ instead of just $\text{Cen}_i$, and that $\gamma_i$ and $\text{Var}(\tilde{w}_{it})$ for (11) vary over counties $i$. The latter feature makes (11) inconvenient for model estimation relative to (9). However, having fitted a bivariate model using (9) and (10), one can compute estimates of $\gamma_i$ and $\text{Var}(\tilde{w}_{it})$ and compare them to the corresponding quantities $\gamma$ and $\sigma_w^2$ from the SAIPE model (which assumes they are constant over counties). (Histograms of $\gamma_i$ and $\text{Var}(\tilde{w}_{it})$ are provided as part of the regression diagnostics for the fitted bivariate models.)[2]

[2] More details related to this transformation of the bivariate model are given in Bell (1997a). To interpret (11), it may help to note that $\mathbf{x}'_{it}\beta + \gamma_i(\text{Cen}_i - \mathbf{x}'_{i,89}\eta) = E(Y_{it}|\mathbf{x}_{it}, \text{Cen}_i)$ and $\text{Var}(\tilde{w}_{it}) = \text{Var}(Y_{it}|\mathbf{x}_{it}, \text{Cen}_i)$.

Because the bivariate model uses previous census data $\mathrm{Cen}_i$ by jointly modeling it with the CPS data $y_{it}$, it could not be applied for $t$ = 1989 because the regression variables $\mathbf{x}_{i,79}$ needed for modeling the 1980 census data are not available. Consequently, the bivariate model was applied only for $t$ = 1993, and $\mathrm{Cen}_i$ in (10) always denotes $\mathrm{Cen90}_i$. (The bivariate model approach can be applied to jointly model 1990 CPS and 1990 census data, but this is a different exercise, since the resulting smoothed estimates of $Y_{it}$ would use current year census data, rather than previous census data.)

## Adding Fixed State Effects to Models

Any of the basic models discussed here can be augmented to include fixed state effects by replacing $x_{0it}$ = 1 by a set of 51 state indicator variables, constructed alphabetically: $I_{1i}$ = 1 for all counties in Alabama and 0 otherwise, $I_{2i}$ = 1 for all counties in Alaska and 0 otherwise, etc., through $I_{51,i}$ = 1 for all counties in Wyoming and 0 otherwise. The resulting regression effect can be written as $\sum_{j=1}^{51} \alpha_j I_{ji}$, where the $\alpha_j$ are state intercept parameters. Alternatively, the regression can be reparameterized as follows to maintain the overall constant term $\beta_0 x_{0it}$, but with 50 state contrast variables added to the regression variables for each equation:

$$\begin{aligned}\sum_{j=1}^{51} \alpha_j I_i &= \sum_{j=1}^{51} (\alpha_j - \bar{\alpha}) I_{ji} + \bar{\alpha} x_{0it} \\ &= \bar{\alpha} x_{0it} + \sum_{j=1}^{51} (\alpha_j - \bar{\alpha})[I_{ji} - I_{51,i}] + \sum_{j=1}^{51} (\alpha_j - \bar{\alpha}) I_{51,i} \\ &= \beta_0 x_{0it} + \sum_{j=1}^{50} \tilde{\alpha}_j M_{ji},\end{aligned}$$

where $\bar{\alpha} = \beta_0 = (1/51) \sum_{j=1}^{51} \alpha_j$ is the mean of the 51 state intercepts; $\tilde{\alpha}_j = \alpha_j - \bar{\alpha}$ are the differential state effects; and $M_{ji} = I_{ji} - I_{51,i}$ are 50 contrast variables that are 1 when county $i$ is in state $j$, –1 when county $i$ is in Wyoming, and 0 otherwise. The differential state effect for Wyoming is $\tilde{\alpha}_{51} = -(\tilde{\alpha}_i + \ldots + \tilde{\alpha}_{50}$, which is obtained from the constraint $\sum_{j=1}^{51} \tilde{\alpha}_j = 0$.

Two sets of state indicator variables (or state contrast variables) are used—one set for the CPS equation and one set for the census equation. These can be denoted $I_{jit}$ ($M_{jit}$) and $I_{ji,89}$ ($M_{ji,89}$), which lets the state intercepts $\alpha_j$ (or effects $\tilde{\alpha}_j$) be distinct for the CPS and census equations. (The two sets of intercepts could be denoted $\alpha_{jt}$ and $\alpha_{j,\mathrm{cen}}$, or the two sets of contrasts could be denoted $\tilde{\alpha}_{jt}$ and $\tilde{\alpha}_{j,\mathrm{cen}}$.) Thus, adding state effects to a model adds 100 additional parameters, 50 in each of the two equations: this holds even when modeling CPS data for $t$ =

1989, the same income year as for the census. This approach avoids assuming that state effects are the same for the CPS and census data (though I and my colleagues did do some experimentation with common state effects in the bivariate model).

## SAIPE and Bivariate Models for Poverty Rates

All the models that have been investigated are of either the SAIPE or bivariate form, with or without fixed state effects; they are simply applied to different data than discussed above. For modeling poverty rates, $\text{Cen}_i$ denotes the census estimated poverty rate for county $i$ (for related children, 5-17). The CPS data $y_{it}$ are defined as an aggregate 3-year "poverty rate," using CPS data for years $t - 1$, $t$, and $t + 1$:

$$y_{it} = \frac{\sum_t(\text{estimated number poor related children 5–17 in county } i)}{\sum_t(\text{estimated total children 5–17 in county } i)}, \quad (14)$$

where $\sum_t$ indicates the 3-year sum over $t - 1$, $t$, and $t + 1$. The estimated numbers for the numerator and denominator of (14) are produced by using CPS weights modified to make each county "self-representing." CPS sample sizes $n_{it}$ are defined as before.

Notice that the denominator of (14) is not the CPS poverty universe (poor related children 5-17 in families), as it was for the single year poverty rates defined in (7); rather, it is the CPS total number of children 5-17. This choice of denominator for the "poverty rate" in (14) is necessary because county population estimates are available for all children 5-17, but not for the 5-17 CPS poverty universe (restricted to related children in families). Population estimates corresponding to the denominator of (14) are needed to convert smoothed poverty rate estimates to estimates of the number of poor children.

For some counties with very small CPS sample sizes there may be no related children aged 5-17 observed in the sample. For these counties, the poverty rates are not defined, and they cannot be used in the model fitting. However, it is not necessary to drop counties just because no *poor* 5-17 children are found in the sample, as it is with the models for log number poor and log poverty rate; the poverty rate models use the most CPS observations for model fitting; 304 counties had CPS sample but no poor age 5-17 in the sample in 1993.

The basic regression variables $\mathbf{x}_{it} = (x_{0it}, \ldots, x_{3it})'$ used in poverty rate models are three other rate variables and an intercept, defined as follows:

$x_{0it} = 1$ (constant term); (15)

$x_{1it}$ = (number of IRS dependent child tax exemptions on returns with income below poverty)/(total IRS dependent child tax exemptions);

$x_{2it}$ = (number of food stamp participants) / (resident population, all ages);

$x_{3it}$ = (total IRS dependent child tax exemptions) / (resident pop. age 0-21).

Except for the constant term, the numerators and denominators of these variables derive from tabulations of administrative records data or population estimates for county *i*. It should be noted that for a significant number of counties (292 in 1993 and 82 in 1989) the IRS dependent child exemption "rate," $x_{3it}$, exceeds 1: this is partly due to errors in geocoding the IRS tax return data, and partly due to differences between IRS and census residence definitions.

Having thus defined the data and regression variables, either the SAIPE model form given by (1) and (2) or the bivariate model form given by (9) and (10) can be used for the estimates. In doing so the same assumptions about the error structure are used. Thus, for SAIPE poverty rate models the model errors $w_{it}$ and $\tilde{z}_i$ in (1) and (2) are both assumed *i.i.d.* $N(0, \sigma_w^2)$ and independent of each other. For bivariate poverty rate models, both model errors $u_{it}$ and $z_i$ in (9) and (10) are assumed *i.i.d.* $N(0, \sigma_u^2)$, with $\mathrm{Cov}(u_{it}, z_i) = \sigma_{uz} = \rho\sigma_u^2$ constant over *i*. And for both SAIPE and bivariate models the CPS sampling errors $e_{it}$ are assumed *i.i.d.* $N(0, v_e / n_{it})$, and the census sampling errors $\epsilon_i$ are assumed *i.i.d.* $N(0, c_i)$. Obviously, the values of the variance parameters will be different from those in the log number poor models: in particular, the census sampling error variances $c_i$ are obtained from published census generalized variances for rate estimates.

To assume that the CPS sampling errors of direct poverty rate estimates have variance of the form $v_e / n_{it}$ is inconsistent with making the same assumption for CPS direct estimates of log number poor or log poverty rate. Simple Taylor series approximations suggest that if $v_e / n_{it}$ is the appropriate variance for poverty rate estimates, then the sampling error variance for log poverty rates will depend on the underlying true poverty rate *p*, and vice versa. (The sampling error variance for log poverty rates will be the same as that for log number poor, ignoring, as a crude approximation, variability in the denominator of the poverty rates.) In fact, considerations of the binomial distribution suggest that sampling error variances of poverty rates and log poverty rates could both depend on *p* (see Bell (1997b) for a little more discussion.) The form $v_e / n_{it}$ of the sampling error variances was chosen not because it was believed to be exactly correct for any of the various data being modeled (poverty rates, log poverty rates, or log number poor), but because it is the simplest form that allows sampling error variance to depend inversely on sample size. Because of the need to estimate $v_e$ from the fitting of the CPS equation, it is doubtful that much more involved sampling error variance formulations could be effectively estimated. Since the Census Bureau now has direct estimates of county sampling error variances (Fay, 1997b), there is more information for exploring alternative sampling variance formulations, and that work has begun. (Fixed state effects can also be added to the poverty rate models, as discussed above.)

### SAIPE and Bivariate Models for Log Poverty Rates

Models for log poverty rates are of the same form as those for poverty rates just discussed, except that the models are applied with the logarithms of all the rates involved. That is, $y_{it}$ and $\text{Cen}_i$ are defined to be the logarithms of the CPS and census poverty rates (defined above) and $(x_{1it}, \ldots, x_{3it})$ are defined to be the logs of the rates given in (15). The $y_{it}$ are not defined for counties for which there are no poor children 5-17 in the CPS sample, so they must be dropped from the model fitting, as is done with the log number poor models.

As with the models discussed above, the assumptions about the covariance structure of (1) and (2) (for a SAIPE model of log poverty rates), or about the covariance structure of (9) and (10) (for a bivariate model), remain unchanged. The parameter values will change, of course: in particular, the sampling variances $c_i$, which now refer to the log census poverty rates, can be approximated from those for the census poverty rates. Thus, if $\tilde{c}_i$ are the sampling variances in census estimates $\hat{p}_i$ of poverty rates $p_i$, and $c_i$ are the corresponding sampling variances in the $\text{Cen}_i = \log(\hat{p}_i)$, from Taylor series linearization the two are approximately related by

$$c_i = \text{Var}\big(\log(\hat{p}_i)\big) \approx \text{RelVar}(\hat{p}_i) \equiv \text{Var}(\hat{p}_i) / p_i^2 = \tilde{c}_i / p_i^2 .$$

## D-Revised Models for Log Poverty Rates

The "D-Revised" models for log poverty rates are a hybrid: they use CPS and census log poverty rates for $y_{it}$ and $\text{Cen}_i$, as defined above, but with regression variables as defined for the log number poor models in (8).[3] Only the SAIPE form of this model was tried, and fixed state effects were not used. (Alternatives using the bivariate model form or fixed state effects, or both, could be investigated.) For the D-Revised model form there is one additional difference between (1) and (2): the census data appearing on the right-hand side of the equations are—analogous to the other regression variables—defined as log number poor children 5-17, whereas $\text{Cen90}_i$ appearing on the left-hand side is the log census poverty rate. With the data thus defined, the model fitting proceeds in the same fashion as for the other models discussed.

## State Poverty Rate Models

Models for state poverty rates are discussed in detail in Fay and Train (1997). Here I provide only brief summary remarks relating their model to the forms just discussed. The model developed was of the form of (11), but with the coefficient $(\gamma_i)$ on the census residuals assumed constant over states *i*:

$$y_{it} = [\mathbf{x}'_{it}\beta + \gamma(\text{Cen}_i - \mathbf{x}'_{i,89}\eta) + \tilde{w}_{it}] + e_{it} . \tag{16}$$

[3] "D-Revised" was the term originally used by the panel for the hybrid log rate-number model.

The model error variance, $\mathrm{Var}(\tilde{w}_{it}) = \sigma_w^2$, was also assumed constant over states. For states, the census sampling error variances $c_i$ are effectively 0. Thus, examining (12) and (13) for states, a bivariate model does indeed lead to the model form (16), with a constant $\gamma$ and $\sigma_w^2$. In Fay and Train (1997), the equation (16) and corresponding census equation of form (10) were fitted separately. Because the census data have negligible sampling error variance, the census equation for states can be fitted by OLS. Fay and Train then fitted (16) by maximum likelihood to estimate $\beta, \gamma$, and $\sigma_w^2$, given previous estimates of the $\mathrm{Var}(e_{it})$.

The estimates of $\mathrm{Var}(e_{it})$ were developed by Mark Otto and myself (see Otto and Bell, 1995). These estimates used generalized variance functions fitted to direct estimates of state sampling error variances developed in Fay and Train (1995). In their later paper on the state modeling, Fay and Train (1997) refined the estimates of $\mathrm{Var}(e_{it})$ as their iterative estimation proceeded by updating the dependence of the $\mathrm{Var}(e_{it})$ on the poverty rate being estimated.

## MODEL FITTING

Once the data for a given model have been defined, model fitting proceeds in the same fashion for all models. Thus, model fitting can be discussed in general terms, with one qualification: for models for log number poor or log poverty rates, counties with no CPS sample poor are omitted from the model fitting, as discussed above. Small numbers of other counties may also be eliminated due to census no poor or problems in defining the regression variables.

First, consider estimation of the regression parameters given estimates of the model variance parameters. Let **y** and **Cen** (similarly, **Cen90** and **Cen80**) be vectors containing the county CPS and census data to be used for model fitting, and let $\mathbf{X}_t$ and $\mathbf{X}_{89}$ be the corresponding matrices of regression variables for their respective equations. The SAIPE model form given by (1) and (2) can be written in a rather obvious matrix-vector notation as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{Cen90} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_t & \mathbf{Cen90} & 0 & 0 \\ 0 & 0 & \mathbf{X}_{89} & \mathbf{Cen80} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \\ \eta \\ \tilde{\gamma} \end{bmatrix} + \begin{bmatrix} \mathbf{w}_t \\ \tilde{\mathbf{z}} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_t \\ \in \end{bmatrix}. \qquad (17)$$

The error vectors $\mathbf{w}_t$, $\tilde{\mathbf{z}}$, $\mathbf{e}_t$, and $\in$ are all assumed uncorrelated with each other, and there are also no correlations among their elements (i.e., each has a diagonal covariance matrix). Thus, $\mathrm{Var}(\mathbf{w}_t + \mathbf{e}_t) = \sigma_w^2 I + v_e K$, where $K$ is a diagonal matrix with elements $1/n_{it}$. Also, $\mathrm{Var}(\tilde{\mathbf{z}} + \in = \sigma_w^2 I + C$, where $C$ is a diagonal matrix with elements $c_i$. Given $\sigma_w^2$, $v_e$, and the $n_{it}$ and $c_i$ (always assumed known), (17) can be fitted by weighted least squares to estimate the regression parameters ($\beta, \gamma, \eta\tilde{\gamma}$). In fact, since there is no correlation between the error terms in the equations for **y** and **Cen90**, these two equations can be fitted separately.

For the bivariate model, the corresponding equation to (17) is

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{Cen90} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_t & 0 \\ \mathbf{0} & \mathbf{X}_{89} \end{bmatrix} \begin{bmatrix} \beta \\ \eta \end{bmatrix} + \begin{bmatrix} \mathbf{u}_t \\ \mathbf{z} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_t \\ \in \end{bmatrix}. \tag{18}$$

In (18) the vectors $\mathbf{u}_t$ and **z** have, in general, nonzero correlations for observations corresponding to the same county. Thus, while $\mathrm{Var}(\mathbf{w}_t + \mathbf{e}_t) = \sigma_u^2 I + v_e K$ and $\mathrm{Var}(\mathbf{z} + \in) = \sigma_u^2 I + C$, similar to the SAIPE model (17), one also needs to allow for the correlations between $\mathbf{u}_t$ and **z** when estimating the regression parameters $(\beta, \eta)$. This can be done by applying generalized least squares to (18). In fact, it is simpler to structure the equations for the bivariate model so that the CPS and census data are paired off (for those counties with CPS data available for model fitting), for which the covariance matrix for the resulting equation is block diagonal, with blocks no larger than 2 × 2. (For counties with only census data available for model fitting, the "block" is a scalar.) (This process is straightforward, but the notation is tedious and details are omitted here.)

Fixed state effects are easily added to (17) or (18) by simply augmenting the regression matrix and parameter vector as appropriate. For example, for the bivariate model (18), with 50 state contrast variables $M_{ji}$ and corresponding parameters $\tilde{\alpha}_j$ added to each equation, the resulting model can be written

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{Cen90} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_t & \mathbf{M}_t & 0 & 0 \\ 0 & 0 & \mathbf{X}_{89} & \mathbf{M}_{89} \end{bmatrix} \begin{bmatrix} \beta \\ \tilde{\alpha}_t \\ \eta \\ \tilde{\alpha}_{\text{cen}} \end{bmatrix} + \begin{bmatrix} \mathrm{u}_t \\ \mathrm{z} \end{bmatrix} + \begin{bmatrix} \mathrm{e}_t \\ \in \end{bmatrix}.$$

Finally, it is necessary to discuss how the covariance parameters are estimated and how this estimation is integrated with that for the regression parameters. Two approaches have been taken. One approach (implemented in SAS IML) was used in fitting models to produce the evaluations against the 1990 census. This approach used basically a method of moments approach (see Fisher and Siegel, 1997).

The second approach (implemented in Splus) was used in fitting the models for producing the regression diagnostics. This approach uses Gaussian maximum likelihood. For bivariate form models, for given values of the model parameters

($\beta, \eta, \sigma_u^2, \rho, v_e$), the joint density of the data (the likelihood function) can be evaluated, and thus numerically maximized over the parameters to produce the maximum likelihood estimates (MLEs). This is done by iterating between GLS estimation of ($\beta, \eta$) for given values of ($\sigma_u^2, \rho, v_e$) and maximization of the likelihood over ($\sigma_u^2, \rho, v_e$), using the regression residuals $y_{it} - \mathbf{x}_{it}'\beta$ and $\text{Cen}_i - \mathbf{x}_{i\,89}'\eta$ as data. This approach can be called iterative GLS. Asymptotic inference (approximate standard errors, etc.) about ($\beta, \eta$) follows from standard GLS results by plugging in MLEs of ($\sigma_u^2, \rho, v_e$), and inference about ($\sigma_u^2, \rho, v_e$) uses standard asymptotic results for MLEs (use of an approximate normal distribution with covariance matrix given by the inverse negative Hessian of the log-likelihood evaluated at the MLEs).

This second approach can also fit models of the SAIPE form. For these models, $\rho = 0$, so the CPS and census equations are independent. However, these two equations are linked by the common variance, $\sigma_w^2$, assumed for the model errors $w_{it}$ and $\tilde{z}_i$. Thus, fitting the two equations jointly combines their information for the estimation of $\sigma_w^2$. Practically speaking, this makes little difference, as the information from the census data swamps that from the CPS data, so that essentially the same results would be obtained by fitting the census equation first to estimate $\sigma_w^2$ and then treating $\sigma_w^2$ as known when estimating the CPS equation. This latter strategy was used in the first approach (implemented in SAS IML).

The SAS program differs from the Splus program in another related respect: in the SAS program the census equation is fitted only to data from the counties that also provide data for the CPS equation. The reasoning behind this decision was that the model error variance might differ for counties without a CPS sample (which are smaller, on average, than counties included in the CPS), and thus it may be appropriate to exclude them from the fitting of the census equation. As noted in the next section, an important role of the model error variance relates to how weights are assigned to the regression predictions and the direct CPS estimates in constructing the smoothed estimates. Since this calculation is irrelevant to counties without a CPS sample, it may be appropriate to avoid their influence on estimates of the model error variance. In the Splus bivariate model software, all the census data are used in the model fitting, along with as much CPS data as are available for the year and the poverty statistic being modeled. This approach assumes that the model applies equally well to counties with and without a CPS sample.

The two different model fitting approaches were adopted because some analysts use SAS and others use Splus and because the SAS code was developed for the original SAIPE model and could not be used to fit models of bivariate form, necessitating development of a second program. Generalization of the Splus bivariate model software is a recent development, and there has not been time to make extensive comparisons of the two programs for models they can both fit. For the comparisons that have been made, the differences in results appear to be small.

## SMOOTHED ESTIMATES

Smoothed estimates from an estimated 1993 SAIPE model form are determined from the CPS equation (1), treating $\text{Cen90}_i$ the same way as the other regression variables in $\mathbf{x}_{it}$. (For $t$ = 1989, the same approach is applied to (4).) Recall that the true quantity of interest for county $i$ is $Y_{it} = \mathbf{x}'_{it}\beta + \gamma\text{Cen90}_i + w_{it}$, and the direct CPS estimate is $y_{it} = Y_{it} + e_{it}$. The estimate of $Y_{it}$ and its variance are

$$\hat{Y}_{it} = h_{it}y_{it} + (1-h_{it})(\mathbf{x}'_{it}\hat{\beta} + \hat{\gamma}\text{Cen90}_i) \tag{19}$$

$$\begin{aligned}\text{Var}(Y_{it} - \hat{Y}_{it}) &= \sigma_w^2(1-h_{it}) \\ &\quad + (1-h_{it})^2[\mathbf{x}'_{it}\ \text{Cen90}_i]\ \text{Var}\begin{pmatrix}\hat{\beta} \\ \hat{\gamma}\end{pmatrix}\begin{bmatrix}\mathbf{x}_{it} \\ \text{Cen90}_i\end{bmatrix},\end{aligned} \tag{20}$$

where

$$h_{it} = \sigma_w^2/(\sigma_w^2 + v_e/n_{it}),$$

and $\text{Var}(\hat{\beta}, \hat{\gamma})$ is obtained from the weighted least squares results. From (19) the smoothed estimate $\hat{Y}_{it}$ is a weighted average of the regression prediction $\mathbf{x}'_{it}\hat{\beta} + \hat{\gamma}\text{Cen90}_i$ and the direct estimate $y_{it}$. The first term in (20), $\sigma_w^2(1-h_{it})$, is the variance that would result if all model parameters were known. The second term in (20) accounts for additional error due to estimating the regression parameters $(\beta, \gamma)$. One can also augment (20) to account for additional error due to estimating some or all of the variance parameters ($\sigma_w^2$ and $v_e$), using either the approach of Prasad and Rao (1990:47-59), or by simulation. These calculations have been done for some of the models, and this addition to the variance was found to be small. (Note that the models have a small number of variance parameters relative to the amount of data.)

For models with fixed state effects, smoothed estimates and their variances are obtained from expressions analogous to (19) and (20) by appropriately augmenting the regression variables and parameters with the state effect regression variables and parameters.

For counties without a CPS sample or that have a CPS sample with no poor children and are dropped from the fitting of log(number poor) or log(poverty rate) models, the estimate $\hat{Y}_{it}$ is defined to be just the regression prediction $\mathbf{x}'_{it}\hat{\beta} + \hat{\gamma}\text{Cen90}_i$, which has variance

$$\mathrm{Var}(Y_{it} - \hat{Y}_{it}) = \sigma_w^2 + [\mathbf{x}_{it}' \ \mathrm{Cen90}_i] \ \mathrm{Var}\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \begin{bmatrix} \mathbf{x}_{it} \\ \mathrm{Cen90}_i \end{bmatrix} .$$

Smoothed estimates and their variances for the bivariate model are a little more complicated, but follow the same principles; they are discussed in Bell (1997a).

When log(numbers of poor) or log(poverty rates) are modeled, smoothed estimates on the original scale (of numbers of poor or of poverty rates, unlogged) can be obtained by exponentiating $\hat{Y}_{it}$. However, it is useful to use the following modified estimate, based on the mean of the lognormal distribution, to remove bias:

$$\exp\left( \hat{Y}_{it} + \frac{1}{2} \mathrm{Var}(Y_{it} - \hat{Y}_{it}) \right). \tag{21}$$

Prediction intervals on the original scale can be obtained by exponentiating prediction interval limits on the transformed (log) scale, yielding asymmetric intervals on the original scale.

When poverty rates are modeled, the resulting smoothed rate estimate for county *i* must be multiplied by the population estimate of total children 5-17 in county *i* (see (14) and discussion following) to convert it to a smoothed estimate of the number of poor children. This is also necessary for smoothed poverty rate estimates from the state model, and, similarly, when log(poverty rates) for counties are modeled, with smoothed rate estimates produced using (21). Prediction error variances in these cases could be taken to be those for the smoothed poverty rates multiplied by the square of the population estimates, though this ignores error in the 5-17 population estimates. Formal measures (variances) of error in state and county population estimates are not available, so there is no ready way to recognize this additional uncertainty. Treating error in the population estimates as ignorable is more tenable for states than it is for counties.

As a final step, smoothed county estimates of number of poor related children aged 5-17 are "raked" to agree with the corresponding smoothed estimates from the state model. Thus, the smoothed county estimates are aggregated to states, and then the individual county estimates are multiplied by the ratio of their state model estimate to the aggregated county estimates for that state. These ratios, or "raking factors," one for each state for a given model, have been developed for the 1989 models. Deriving variances for the raked, smoothed estimates is complicated, but an approximate procedure (described in Fisher and Siegel, 1997) has been implemented in conjunction with the SAS estimation software.

# APPENDIX B

# Population Estimates

The Census Bureau has long had an active program of using demographic analysis to develop updated estimates of total population and population by age for various levels of geography, such as states, counties, and cities. The Census Bureau's state and county models of school-age children who were poor in 1993 use state and county postcensal population estimates for age groups as of July 1994. These estimates were developed within the framework of the Census Bureau's population estimates program (Long, 1993).[1]

## TOTAL POPULATION ESTIMATES

Total population estimates are developed by the component method of demographic analysis. In general, the component method starts from an area's population in the previous census. That number is then updated by the net demographic change—adding births and international immigration and subtracting deaths and emigration. The final component, internal migration or migration to and from other parts of the United States, is currently estimated from administrative records. No adjustments are made to the population estimates to allow for the estimated net population undercount in the census.

[1]Estimates for Puerto Rico are developed separately. The basic methodology uses registered births by sex, registered deaths by age and sex, and estimates of annual intercensal net migration by age and sex from an analysis using the natural rate of increase for the 1980-1990 period and the reported 1990 census population by age and sex (Reed, 1996).

Postcensal county estimates of total population are produced by the component method: (1) the numbers of births and deaths are based on reported vital statistics for each county; (2) reports of the Immigration and Naturalization Service are used to estimate net immigration from abroad; and (3) administrative records are used to estimate net migration among counties. Net migration of people under 65 years of age is estimated for each county from a year-to-year match of IRS federal income tax returns; for people aged 65 and over, net migration is estimated for each county from the change in Medicare enrollment (Bureau of the Census, 1995). Estimates are developed separately for household and group quarters populations.

The county population totals are summed for each state to provide estimates of the total population of each state. All county and state population totals are then adjusted to sum to independently derived estimates of the total U.S. population. The county estimates are also reviewed locally under the Census Bureau's Federal-State Cooperative Population Estimates (FSCPE) program.

Operationally, the county total population estimates are the sum of the estimates for four groups:

- Household population under age 65 (HHP<65);
- Household population age 65 and over (HHP65+);
- Group quarters population under age 65 (GQ<65); and
- Group quarters population age 65 and over (GQ65+).

## HHP<65

The estimates for HHP<65 use a component method for year $t$ to measure change in each component of population change during the 12-month period preceding the estimate date, as follows::

$$\text{HHP}<65_t = \text{HHP}<65_{t\text{-}1} + \text{NI} + \text{NMIG} + \text{NETMOVE} - \text{AGE}. \tag{1}$$

NI is natural increase (births and deaths for people under age 65), which is estimated from a combination of vital statistics data from the National Center for Health Statistics (NCHS) and the FSCPE. Each of these sources has some problems. The FSCPE does not include all states, and the NCHS data exhibit some peculiarities (e.g., birth records are not always properly assigned to place of residence in such areas as Washington, D.C., in which births often occur in hospitals that are not in the county of residence, and in areas with military bases).

NMIG, net internal migration, is estimated from data on IRS tax returns matched year to year on the basis of the Social Security number of the filer. A migration rate is developed from the number of exemptions on the matched tax returns [(inmigrants – outmigrants) / (nonmigrants + outmigrants)], and this rate is applied to the migration base [HHP<65$_{t\text{-}1}$ + 1/2(NI + NETMOVE) – AGE].

Coverage of the IRS data (i.e., the proportion of exemptions to estimated population) varies across counties, as do matching rates.

NETMOVE is nondomestic net movements, mainly international immigration and emigration. It is estimated with a variety of data, and the totals generally are small. *Legal immigrants and refugees* (about 800,000 per year nationwide) are assigned to a county of residence on the basis of Immigration and Naturalization Service data about their intended place of residence although they may not reside at the indicated place. *Undocumented immigrants* (estimated at 225,000 annually) are assigned to a county on the basis of the 1990 census distribution of the foreign born population. Estimates are also made of *emigrants* (about 195,000 per year). *Net inmigrants from Puerto Rico* (only about 7,000 annually because there is almost an equal number of outmigrants each year) used to be estimated from passenger traffic data from the San Juan airport. However, this method became increasingly untenable, and the current procedure uses estimates of migration of Puerto Ricans to the rest of the world, which include an assumption of the U.S. share. The U.S. share is allocated to counties on the basis of 1990 census data on place of residence. Estimates of the net movement in and out of the country of *military and federal civilian and military dependents* are based on data from the Department of Defense (DoD) and the Office of Personnel Management. County station strength data from DoD, which are used to allocate military personnel to counties, are modified in some locations (e.g., the Washington, D.C., area).

Lastly, AGE is an estimate of the number of persons in the county who aged from 64 to 65 during the year.

Except for internal migration, all components are controlled to national totals.

## HHP65+

The estimates for HHP65+ use a component method in which:

$$\text{HHP65+}_t = \text{HHP65+}_{t\text{-}1} + \text{NI65+} + \text{NMIG65+} + \text{NETMOVE65+}. \qquad (2)$$

NI65+ is natural increase (decrease) of persons aging into the cohort (AGE in the first equation) minus deaths in the population aged 65 and over.

NMIG65+ is internal migration, which is estimated from Medicare enrollment data. A migration rate is estimated as [(actual Medicare enrollment$_{t\text{-}1}$ – expected Medicare enrollment) / actual enrollment$_{t\text{-}1}$]. Expected Medicare enrollment is [actual enrollment$_{t-1}$ + (NI65+$_{t-1}$ × the 1990 Medicare coverage ratio)].[2] The estimated migration rate is then applied to the migration base, HHP65+$_{t-1}$ + 1/2(NI$_{t\text{-}1}$ + NETMOVE$_{t-1}$).

[2]Previously, the method simply used the change in Medicare enrollment to estimate the migration rate for the population aged 65 and over directly; the current method preserves the county variation in Medicare coverage.

NETMOVE65+ is other net movements (legal immigrants, undocumented immigrants, refugees, emigrants, net entrants from Puerto Rico).

## GQ<65 and GQ65+

Group quarters population for both age groups (GQ<65 and GQ65+) is estimated as the 1990 census group quarters population plus the difference between the current group quarters report (GQR) minus the 1990 GQR figure. The GQR is compiled annually from data obtained from the FSCPE, DoD, Veterans Administration, and colleges by type of group quarters: correctional facility, juvenile facility, nursing home, other institutional, college, military quarters, other noninstitutional.

## ESTIMATES BY AGE

Estimates by age are prepared separately, but within the framework of the total population estimates for states and counties.

County age estimates are prepared in a two-step procedure. In the first step, estimates of total county population are developed as described above. Separately, estimates of state populations by single years of age, sex, race, and Hispanic origin are developed. The state age estimates (which are controlled to the state total population estimates) use a component method in which migration rates by age for people under age 65 are derived from school enrollment data (Bureau of the Census, 1987).

In the second step, the county age estimates are developed by using a raking-ratio adjustment of the estimates from the previous census. In this approach, the beginning matrix of counts for each county by age, sex, race, and Hispanic origin from the previous census is simultaneously adjusted to agree with the postcensal estimate of the total county population and the postcensal estimates for the applicable state by age, sex, race, and Hispanic origin (Sink, 1996).[3]

This procedure assumes that the age distribution of each county within a state changes in the same manner as that state's age distribution. Errors in the county estimates of an age group can arise from errors in this assumption, errors in the derivation of the state estimates of age groups, and errors in the derivation of the county estimates of total population.

---

[3]The revised county age estimates for 1994 that were used in producing the revised county estimates of poor school-age children in 1993 were developed by the ratio-raking procedure just described with the following refinement: the ratio-raking procedure was applied separately to the 1990 census figures for school-age children in group quarters and school-age children not in group quarters.

## USE OF POSTCENSAL POPULATION ESTIMATES

The process for estimating school-age children in poverty at the county level in 1993 and the Title I allocation formulas for using those estimates require population totals by age in a noncensus year. For estimating numbers of poor school-age children, the Census Bureau's county model uses population estimates for the population under age 18 in 1994, and the state model uses population estimates for the population under age 65 and the noninstitutionalized population aged 5-17 in 1994. Estimates for the noninstitutionalized population are developed by subtracting administrative record counts of institutionalized people in the relevant age groups from the demographic estimates developed as described above.

As required by the Title I legislation, the Census Bureau provided to the Department of Education estimates of *total* children aged 5-17 for counties as of July 1994 to use as denominators in calculating county proportions of poor school-age children to use in the Title I allocations. (The numerators are the Census Bureau's estimates of the number of related school-age children in each county who were poor in 1993, developed as described in Chapter 2, with the addition by the Department of Education of estimates for several other groups of children, such as those in foster care, as specified by the legislation.)

## EVALUATION OF COUNTY AGE ESTIMATES

The Census Bureau maintains an ongoing program to develop and review the performance of its population estimates, including evaluating the estimates at 10-year intervals by comparing them with decennial census figures. These comparisons are not complete measures of the accuracy and precision of the population estimates because the standard (i.e., the decennial census) itself is flawed, notably from net population undercount, which varies by age group across time and place (see Robinson et al., 1993).

To evaluate the county age estimates developed with the current raking-ratio procedure, the Census Bureau raked the 1980 census county age figures to the 1990 estimates of county total population and state population by age. The resulting 1990 county age estimates were compared with the 1990 census county age figures, which were assumed to be the true values in each case.

Tables B-1 to B-8 show the average proportional algebraic difference and the average proportional absolute difference, expressed as percents, between the 1990 county population estimates for people aged 5-17, developed by raking the 1980 census estimates as described above, and the 1990 census figures. Following Census Bureau terminology, these difference measures are termed mean algebraic (i.e., signed) percent error and mean absolute percent error.[4] The two measures are shown for all counties and for counties grouped into categories for the following characteristics: population size in 1990; population growth from 1980 to 1990; percent black and other nonwhite population in 1990; percent Hispanic population in 1990; percent poor population in 1990; percent group quarters residents in 1990; census division; and metropolitan status. Also shown are the percentage of counties with negative errors (underpredictions relative to the census) and the number of counties with more than a 20 percent negative or positive error.

## FINDINGS

The overall mean absolute percent error in the 1990 county estimates of people aged 5-17 is 6.3 percent (shown in Table B-1).[5] By comparison, for 1990 county estimates of total population, prepared using the Census Bureau's current estimation procedure, it is 3.6 percent (Davis, 1994).

The mean absolute percent errors do not seem to be concentrated in any particular types of counties (Tables B-2 to B-8). However, as would be expected, the smallest counties (those with populations under 2,500) have errors running at twice the overall average: 12.4 percent, compared with 6.3 percent overall (see Table B-1).

There may be a systematic prediction bias by county population size (Table B-1). The mean algebraic percent error is negative for counties in the smaller population size groups (except for those under 2,500 with a 0.0 value) and positive for counties in the larger population size groups. The percentage of counties with negative prediction errors generally increases as county population size decreases. Overall, the mean algebraic percent error is –0.4.

Nonmetropolitan counties also have a negative mean algebraic percent error (see Table B-8), with 60 percent of these counties having negative prediction errors, which is consistent with the pattern of negative prediction errors for smaller counties. Negative mean algebraic percent errors also characterize counties with negative or lower rates of population growth (Table B-2); with lower percent black and other nonwhite population (Table B-3); with average or higher than average percent Hispanic population (Table B-4); with smaller percent poor popu-

[4]The mean absolute percent error is computed as the sum for all counties (or all counties in a category) of the absolute difference between the estimate and the 1990 census figure for each county as a proportion of the census figure for each county, divided by the number of counties. The mean algebraic percent error is computed similarly, except that the sign of the difference (positive or negative) is considered in the computation.

[5]The estimates in all the tables are unweighted by population size. The overall weighted mean absolute percent error for 1990 county estimates of children aged 5-17 is 4.9 percent, as compared with 6.3 percent, unweighted.

lation (Table B-5); with higher percent group quarters residents (Table B-6); and for counties in the Mountain, Pacific, North Central (East and West), and New England Divisions (Table B-7).

The 72 counties with mean algebraic percent errors of greater than 20 percent, whether positive (36 counties) or negative (36 counties), do not have any common features except that they are almost all nonmetropolitan counties and generally have smaller populations. There are, however, a few counties with populations of 50,000 or more and many counties with populations of 10,000 or more that have large prediction errors. Most of the large errors are between 20 and 30 percent.

An issue in examining the mean algebraic percent errors in the 1990 county estimates of children aged 5-17 for categories of counties (see Tables B-1 through B-8) is whether the patterns observed—for example, the tendency for smaller(larger)-sized counties to have negative(positive) prediction errors—are statistically significant, suggesting the possibility of a systematic bias. Tests of significance were conducted to determine whether there is evidence of possible bias with respect to the characteristics in Tables B-1 to B-8. Since most of these characteristics have ordered categories, a test of a linear trend was conducted using the Abelson-Tukey test procedure (Abelson and Tukey, 1963). Because the number of degrees of freedom is large, the test statistic has essentially a normal distribution under the null hypothesis of no trend. The categories for census division do not have an ordering, so a one-way analysis of variance was performed for that characteristic.

The test results suggest the possibility of some bias associated with the estimates of children aged 5-17 for several categories of counties: county population size, percent black and other nonwhite population, percent Hispanic population, percent group quarters residents, metropolitan status, and census division. However, the results are not conclusive given that there is only a single year—1990—for which it is possible to evaluate the estimates by comparison with figures from the census or another source.

TABLE B-1 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Population Size in 1990

| Population Size, 1990 | Counties (Number)[a] | Mean Algebraic Percent Error[b] | | Mean Absolute Percent Error[b] | | Percentage of Counties with Negative Errors | Counties with More Than 20 Percent Error (Number) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | – | + |
| All | 3,140 | –0.4 | (8.7) | 6.3 | (6.1) | 56.2 | 36 | 36 |
| 1,000,000 and over | 30 | 1.5 | (6.5) | 5.2 | (4.1) | 46.7 | 0 | 0 |
| 500,000 to 1,000,000 | 67 | 1.7 | (5.1) | 4.4 | (3.2) | 29.9 | 0 | 0 |
| 100,000 to 500,000 | 361 | 0.9 | (5.9) | 4.6 | (3.8) | 50.4 | 0 | 1 |
| 50,000 to100,000 | 384 | 0.6 | (7.3) | 5.6 | (4.7) | 51.3 | 2 | 2 |
| 10,000 to 50,000 | 1,543 | –0.5 | (7.7) | 6.0 | (4.8) | 56.9 | 10 | 12 |
| 5,000 to 10,000 | 457 | –1.5 | (9.0) | 7.2 | (5.6) | 61.7 | 6 | 8 |
| 2,500 to 5,000 | 180 | –3.2 | (10.5) | 8.4 | (7.0) | 67.2 | 8 | 3 |
| Less than 2,500 | 118 | 0.0 | (21.2) | 12.4 | (17.2) | 59.0 | 10 | 10 |

[a]Excludes Kalawayo County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Standard deviations are in parentheses.

TABLE B-2 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Growth Rate, 1980-1990

| Growth Rate, 1980-1990 | Counties (Number)[a] | Mean Algebraic Percent Error[b] | | Mean Absolute Percent Error[b] | | Percentage of Counties with Negative Errors | Counties with More Than 20 Percent Error (Number) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | – | + |
| All | 3,140 | –0.4 | (8.7) | 6.3 | (6.1) | 56.2 | 36 | 36 |
| Decrease of 5% or more | 834 | –0.7 | (10.1) | 6.7 | (7.5) | 58.4 | 12 | 14 |
| –5% to 0% | 595 | –1.0 | (7.8) | 5.8 | (5.3) | 60.7 | 1 | 4 |
| 0 to 5% | 583 | –0.4 | (7.8) | 5.9 | (5.0) | 55.8 | 7 | 5 |
| 5 to 10% | 386 | –0.1 | (7.7) | 5.7 | (5.1) | 57.5 | 2 | 4 |
| 10 to 15% | 208 | 0.6 | (7.5) | 6.1 | (4.5) | 49.0 | 2 | 0 |
| 15 to 25% | 247 | 0.3 | (9.0) | 6.7 | (6.0) | 51.8 | 3 | 3 |
| 25% and over | 287 | –0.2 | (10.0) | 7.5 | (6.5) | 48.4 | 9 | 6 |

[a]Excludes Kalawayo County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Standard deviations are in parentheses.

TABLE B-3 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Percent Black and Other Nonwhite Population, 1990

| Percent Black and Other Nonwhite, 1990 | Counties (Number)[a] | Mean Algebraic Percent Error[b] | | Mean Absolute Percent Error[b] | | Percentage of Counties with Negative Errors | Counties with More Than 20 Percent Error (Number) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | – | + |
| All | 3,140 | –0.4 | (8.7) | 6.3 | (6.1) | 56.2 | 36 | 36 |
| Less than 0.5% | 304 | –0.7 | (13.1) | 7.4 | (10.8) | 61.6 | 9 | 7 |
| 0.5 to 1.0% | 405 | –2.1 | (8.3) | 6.1 | (6.0) | 67.2 | 3 | 5 |
| 1.0 to 2.0% | 468 | –1.6 | (8.4) | 6.8 | (5.2) | 62.4 | 5 | 5 |
| 2.0 to 5.0% | 550 | –1.0 | (7.8) | 6.2 | (4.8) | 58.6 | 4 | 3 |
| 5.0 to 15.0% | 641 | 0.3 | (8.2) | 6.3 | (5.3) | 49.1 | 9 | 7 |
| 15.0 to 40.0% | 546 | 0.8 | (7.6) | 5.7 | (5.1) | 48.7 | 3 | 3 |
| 40.0% and over | 226 | 1.6 | (8.0) | 6.1 | (5.5) | 48.5 | 3 | 6 |

[a]Excludes Kalawayo County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Standard deviations are in parentheses.

TABLE B-4 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Percent Hispanic Population, 1990

| Percent Hispanic, 1990 | Counties (Number)[a] | Mean Algebraic Percent Error[b] | | Mean Absolute Percent Error[b] | | Percentage of Counties with Negative Errors | Counties with More Than 20 Percent Error (Number) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | – | + |
| All | 3,140 | –0.4 | (8.7) | 6.3 | (6.1) | 56.2 | 36 | 36 |
| Less than 0.5% | 983 | 0.7 | (9.2) | 6.1 | (6.9) | 52.9 | 5 | 13 |
| 0.5 to 1.0% | 760 | 0.2 | (7.0) | 5.5 | (4.3) | 52.5 | 4 | 3 |
| 1.0 to 2.0% | 485 | –0.1 | (8.3) | 6.5 | (5.2) | 56.1 | 3 | 10 |
| 2.0 to 5.0% | 385 | –1.4 | (8.8) | 6.5 | (6.2) | 60.3 | 7 | 3 |
| 5.0 to 15.0% | 291 | –3.2 | (9.8) | 7.5 | (7.0) | 63.4 | 11 | 4 |
| 15.0 to 40.0% | 162 | –3.8 | (10.4) | 8.4 | (7.3) | 70.4 | 5 | 3 |
| 40.0% and over | 74 | –1.0 | (8.1) | 6.5 | (4.9) | 56.8 | 1 | 0 |

[a]Excludes Kalawayo County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Standard deviations are in parentheses.

TABLE B-5 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Percent Poor Population, 1990

| Percent Poor, 1990 | Counties (Number)[a] | Mean Algebraic Percent Error[b] | | Mean Absolute Percent Error[b] | | Percentage of Counties with Negative Errors | Counties with More Than 20 Percent Error (Number) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | – | + |
| All | 3,140 | –0.4 | (8.7) | 6.3 | (6.1) | 56.2 | 36 | 36 |
| None | 50 | –1.5 | (8.2) | 7.0 | (4.4) | 58.0 | 0 | 0 |
| Less than 5% | 253 | 1.8 | (13.9) | 7.3 | (11.9) | 45.5 | 11 | 6 |
| 5 to 10% | 1,046 | –1.4 | (7.7) | 6.0 | (5.1) | 62.1 | 8 | 8 |
| 10 to 15% | 929 | –1.1 | (8.4) | 6.7 | (5.2) | 58.3 | 13 | 8 |
| 15 to 25% | 688 | 0.9 | (8.2) | 6.1 | (5.6) | 50.0 | 2 | 11 |
| 25 to 40% | 157 | 0.8 | (7.0) | 5.4 | (4.4) | 49.7 | 2 | 1 |
| 40% and over | 17 | 3.1 | (11.4) | 8.9 | (7.5) | 35.3 | 0 | 2 |

[a]Excludes Kalawayo County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Standard deviations are in parentheses.

TABLE B-6 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Percent Group Quarters Residents, 1990

| Percent Group Quarters, 1990 | Counties (Number)[a] | Mean Algebraic Percent Error[b] | | Mean Absolute Percent Error[b] | | Percentage of Counties with Negative Errors | Counties with More Than 20 Percent Error (Number) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | – | + |
| All | 3,140 | –0.4 | (8.7) | 6.3 | (6.1) | 56.2 | 36 | 36 |
| Less than 0.5% | 175 | 1.0 | (17.3) | 9.9 | (14.2) | 50.6 | 9 | 11 |
| 0.5 to 1.0% | 372 | 1.4 | (8.1) | 6.4 | (5.2) | 43.0 | 6 | 4 |
| 1.0 to 1.5% | 636 | 0.6 | (7.6) | 5.9 | (4.7) | 49.7 | 4 | 3 |
| 1.5 to 2.0% | 591 | –0.3 | (7.4) | 5.7 | (4.6) | 55.3 | 1 | 3 |
| 2.5 to 3.0% | 535 | –1.8 | (8.3) | 6.3 | (5.7) | 64.7 | 7 | 5 |
| 3.0 to 5.0% | 431 | –0.9 | (7.0) | 5.5 | (4.4) | 60.8 | 1 | 4 |
| 5.0% and over | 400 | –2.1 | (9.0) | 7.1 | (5.9) | 66.1 | 8 | 6 |

[a]Excludes Kalawayo County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Standard deviations are in parentheses.

TABLE B-7 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Census Division

| Census Division | Counties (Number)[a] | Mean Algebraic Percent Error[b] | | Mean Absolute Percent Error[b] | | Percentage of Counties with Negative Errors | Counties with More Than 20 Percent Error (Number) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | – | + |
| All | 3,140 | –0.4 | (8.7) | 6.3 | (6.1) | 56.2 | 36 | 36 |
| New England | 67 | –1.2 | (5.3) | 4.1 | (3.6) | 62.7 | 1 | 0 |
| Middle Atlantic | 150 | 0.6 | (5.2) | 4.1 | (3.3) | 54.0 | 0 | 0 |
| East North Central | 437 | –1.4 | (5.7) | 4.7 | (3.5) | 64.5 | 0 | 0 |
| West North Central | 618 | –3.0 | (7.5) | 6.4 | (5.0) | 72.0 | 4 | 5 |
| South Atlantic | 591 | 2.4 | (8.3) | 6.5 | (5.7) | 39.6 | 7 | 7 |
| East South Central | 364 | 2.9 | (7.2) | 6.0 | (5.0) | 37.4 | 0 | 7 |
| West South Central | 470 | –0.4 | (9.6) | 7.0 | (6.5) | 50.9 | 9 | 8 |
| Mountain | 281 | –2.8 | (14.1) | 9.0 | (11.2) | 68.7 | 10 | 7 |
| Pacific | 161 | –2.5 | (7.9) | 6.5 | (5.1) | 68.5 | 5 | 2 |

[a]Excludes Kalawayo County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Standard deviations are in parentheses.

TABLE B-8 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Metropolitan Status, 1990

| Metropolitan Status, 1990 | Counties (Number)[a] | Mean Algebraic Percent Error[b] | | Mean Absolute Percent Error[b] | | Percentage of Counties with Negative Errors | Counties with More Than 20 Percent Error (Number) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | – | + |
| All | 3,140 | –0.4 | (8.7) | 6.3 | (6.1) | 56.2 | 36 | 36 |
| Nonmetropolitan | 2,393 | –1.2 | (9.0) | 6.5 | (6.3) | 60.0 | 34 | 26 |
| Metropolitan | 747 | 1.9 | (7.2) | 5.6 | (4.9) | 43.9 | 2 | 10 |

[a]Excludes Kalawayo County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Standard deviations are in parentheses.

# APPENDIX C

# Regression Diagnostics on Alternative County Regression Models

An internal evaluation of a regression model, or "regression diagnostics," involves an assessment of its underlying assumptions and features. Chapter 4 reports the results of such an evaluation for four county models, estimated for 2 years, 1989 and 1993. These four models, which were considered serious candidates to produce revised county estimates of poor school-age children in 1993, have the following designations: (a) log number model (under 21, the original county model); (b) log number model (under 18, the revised county model); (c) log rate model (under 21); and (d) log rate model (under 18).

This appendix summarizes the results of an internal evaluation for 13 county models, listed below (see Chapter 3 and Appendix A for the model specifications). Twelve of the models were considered in the first round of model evaluations; they include models (a), (b), and (c). The other model, the log rate (under 18) model (d), was added for the second round of evaluations, which considered the four candidate models (a-d).

Of the 13 county models, 7 are single-equation models, in which the dependent variable is from 3 years of the CPS. For 1993 estimates of poor school-age children, the dependent variable is a weighted average of data from the March 1993, 1994, and 1995 CPS, covering income years 1992, 1993, and 1994. For 1989 estimates of poor school-age children, produced for evaluation purposes, the dependent variable is a weighted average of data from the March 1989, 1990, and 1991 CPS, covering income years 1988, 1989, and 1990.

The other 6 county models are bivariate models in which two equations are jointly estimated to develop estimates of poor school-age children in 1993. In one equation, the dependent variable is a weighted average of data from the

March 1993, 1994, and 1995 CPS, covering income years 1992, 1993, and 1994. In the second equation, the dependent variable is from the 1990 census, covering income year 1989.

The regression coefficients for all the CPS models are presented in Table C-1; Table C-2 shows the regression coefficients for the 1990 census equation for the 6 bivariate models (see pages 128-130).

| **Single-Equation Models** | **Bivariate Models** |
|---|---|
| Log number under 21 (1989, 1993) | Log number under 21 (1993) |
| Log number under 18 (1989, 1993) | |
| Log number under 21, fixed state effects (1989, 1993) | Log number under 21, fixed state effects (1993) |
| Log rate under 21 (1989, 1993) | Log rate under 21 (1993) |
| | Log rate under 21, fixed state effects (1993) |
| Log rate under 18 (1989, 1993) | |
| Rate under 21 (1989, 1993) | Rate under 21 (1993) |
| | Rate under 21, fixed state effects (1993) |
| Hybrid log rate-number under 21 (1989, 1993) | |

NOTE: The years for which coefficients were fit are in parentheses; for the bivariate models, the year shown is for the CPS equation.

## REGRESSION DIAGNOSTICS METHODS

Regression diagnostics is an analysis of the extent to which the various assumptions on which a regression model is based are supported by the data. The following six assumptions were examined for the 13 county models of poor school-age children (see Chapter 4):

(1) linearity of the relationship between the dependent variable and the predictor variables;

(2) constancy over time of the assumed linear relationship and in the estimated coefficients of the predictor variables;

(3) which variables are needed in the model, specifically, whether any of the included predictor variables are *not* needed in the model and, conversely, whether other potential predictor variables *are* needed in the model;

(4) normality (primarily symmetry and moderate tail length) of the distribution of the standardized residuals;[1]

(5) whether the standardized residuals have homogeneous variances, that is, whether the variability of the standardized residuals is constant across counties and does not depend on the values of the predictor variables; and

(6) the absence of outliers, which can be considered to be the absence of an extremely long tail to the error distribution.

Various techniques are useful for examining the degree to which each of these six assumptions obtain. The following techniques that were implemented by the panel and the Census Bureau to evaluate the 13 county models are certainly not the only ones that can be used to examine each of the above assumptions, but they are usually included. In addition to these general techniques, specific analyses were conducted to evaluate the bivariate model formulation in comparison with the single-equation model formulation and the use of the population under age 18 in comparison with the population under age 21 as a predictor variable in the log number model.

***Linearity*** Linearity of the relationships between the dependent variable and the predictor variables was assessed graphically, by observing whether there was evidence of curvature in the plots of standardized residuals against predictor variables in the model. In addition, plots of residuals against CPS sample size and against the predicted values from the regression model were examined for curvature.

***Constancy*** For the single-equation models that could be fit for both 1989 and 1993, the regression coefficients were compared to determine if the values remained roughly constant over time.

***Inclusion or Exclusion of Predictor Variables*** The possibility that one or more predictor variables should be excluded from a model was assessed by looking for insignificant *t*-statistics for the estimated values of individual regression coefficients. The need to include additional predictor variables was assessed by looking for nonrandom patterns, indicative of possible model bias, in the distributions of standardized residuals displayed for various categories of counties. (See Chapter 4 for the categories examined in various model evaluations; the distributional displays examined for this and other model assumptions were box plots.)

[1]See Chapter 4 for the procedure used to standardize the residuals, which are the differences between the predicted and reported values of the dependent variable for each observation.

***Normality*** The normality of the standardized residuals was evaluated through use of Q-Q plots, histograms, and box plots of the standardized residuals. While some skewness of the distribution of standardized residuals may be acceptable, extreme skewness can change the regression fit so that a relatively small number of counties have more influence on the estimation of the regression coefficients. In addition, extreme skewness can indicate the need for a transformation of the variables, which might, in turn, reveal the need for additional predictor variables.

***Homogeneous Variances*** The homogeneity of the variance of the standardized residuals was assessed using several statistics and graphical displays. The statistics included: Spearman's rank correlation coefficient of absolute standardized residuals with the predicted values and also with the CPS sample size, and a robust regression of the log absolute standardized residuals on CPS sample size. The graphical displays included: scatterplots of absolute standardized residuals versus model predictor variables; box plots of absolute standardized residuals for categories of counties; plots of the median absolute deviation of the standardized residuals in a category by categories; plots of absolute standardized residuals versus log CPS sample size; and plots of standardized residuals to the two-thirds power (the Wilson-Hilferty transformation) versus log CPS sample size.

***Outliers*** The assumption of the absence of outliers was evaluated through examination of plots of the distributions of the standardized residuals and plots of standardized residuals against the predictor variables and through analysis of patterns in the distribution of the 30 largest absolute standardized residuals for the various characteristics used to categorize the counties.[2] Any patterns observed among the 30 largest absolute standardized residuals for a characteristic may suggest that a predictor variable should be added to a model.

## FINDINGS

***Linearity*** There is no evidence of any strong nonlinearity between the predictor variables and the dependent variable in any of the 13 models. Thus, there is no reason to suggest a transformation of the dependent variable in any of the models, nor is there reason to include any higher order polynomial terms as additional predictor variables.

***Constancy*** The regression coefficients for the 7 single-equation models for 1989 and 1993 are shown in Table C-1. All of these models have some coefficients that differ substantially between 1989 and 1993.

[2]All the outlier statistics examined are based on the residuals from a least squares model fit, so they may miss influential outliers. It would be useful to look for outliers from a robust fit of the models. It would also be useful to compare the predictions from models with extreme outliers removed.

TABLE C-1 Estimates of Regression Coefficients for the CPS Equation for 13 County Models

| | Predictor Variables[a] | | | | |
|---|---|---|---|---|---|
| Model | 1 | 2 | 3 | 4 | 5 |
| Log Number (under 21) | | | | | |
| 1989 | 0.52 | 0.30 | 0.76 | –0.81 | 0.27 |
| | (.07) | (.05) | (.22) | (.22) | (.07) |
| 1993 | 0.31 | 0.30 | 0.03 | 0.03 | 0.40 |
| | (.08) | (.07) | (.21) | (.21) | (.09) |
| Log Number (under 18) | | | | | |
| 1989 | 0.50 | 0.23 | 1.79 | –1.80 | 0.32 |
| | (.06) | (.05) | (.27) | (.27) | (.07) |
| 1993 | 0.38 | 0.27 | 0.65 | –0.59 | 0.34 |
| | (.08) | (.07) | (.24) | (.24) | (.09) |
| Log Number (under 21), Fixed State Effects | | | | | |
| 1989 | 0.36 | 0.27 | 0.45 | –0.56 | 0.51 |
| | (.13) | (.07) | (.25) | (.25) | (.10) |
| 1993 | 0.50 | 0.17 | –0.03 | –0.07 | 0.45 |
| | (.12) | (.09) | (.25) | (.25) | (.11) |
| Hybrid Log Rate-Number (under 21) | | | | | |
| 1989 | 0.55 | 0.27 | 0.35 | –1.34 | 0.25 |
| | (.06) | (.05) | (.21) | (.21) | (.06) |
| 1989 | 0.37 | 0.26 | –0.33 | –0.59 | 0.37 |
| | (.07) | (.06) | (.18) | (.18) | (.08) |
| Bivariate Log Number (under 21) | | | | | |
| 1993 | 0.57 | 0.45 | 0.19 | –0.20 | NA |
| | (.06) | (.05) | (.20) | (.20) | |
| Bivariate Log Number (under 21), Fixed State Effects | | | | | |
| 1993 | 0.83 | 0.34 | 0.21 | –0.38 | NA |
| | (.09) | (.07) | (.24) | (.24) | |

TABLE C-1 Continued

| Model | Predictor Variables[b] 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Log Rate (under 21) | | | | |
| 1989 | 0.32 | 0.29 | –0.73 | 0.40 |
| | (.07) | (.04) | (.19) | (.07) |
| 1993 | 0.23 | 0.31 | –0.07 | 0.41 |
| | (.08) | (.06) | (.18) | (.09) |
| Log Rate (under 18) | | | | |
| 1989 | 0.29 | 0.26 | –1.13 | 0.43 |
| | (.07) | (.04) | (.24) | (.07) |
| 1993 | 0.26 | 0.30 | –0.42 | 0.38 |
| | (.08) | (.06) | (.20) | (.09) |
| Rate (under 21) | | | | |
| 1989 | 0.25 | 0.46 | –0.16 | 0.56 |
| | (.06) | (.08) | (.03) | (.06) |
| 1993 | 0.09 | 0.60 | –0.05 | 0.52 |
| | (.06) | (.11) | (.03) | (.10) |
| Bivariate Log Rate (under 21) | | | | |
| 1993 | 0.57 | 0.40 | –0.12 | NA |
| | (.05) | (.04) | (.16) | |
| Bivariate Log Rate (under 21), Fixed State Effects | | | | |
| 1993 | 0.75 | 0.35 | –0.01 | NA |
| | (.08) | (.05) | (.19) | |
| Bivariate Rate (under 21) | | | | |
| 1993 | 0.38 | 0.89 | –0.05 | NA |
| | (.04) | (.06) | (.03) | |
| Bivariate Rate (under 21), Fixed State Effects | | | | |
| 1993 | 0.44 | 0.85 | –0.05 | NA |
| | (.06) | (.08) | (.04) | |

NOTES: All predictor variables are on the logarithmic scale for numbers and rates. Standard errors of the estimated regression coefficients are in parentheses. Estimated coefficients for the state indicator variables are not shown. The models were estimated with maximum likelihood. NA: not applicable.

[a]Predictor variables: (1) number of child exemptions reported by families in poverty on tax returns (1989 or 1993); (2) number of people receiving food stamps (1989 or 1993); (3) population (under age 21 or under age 18, 1990 or 1994); (4) total number of child exemptions on tax returns (1989 or 1993); (5) number of poor school-age children from previous (1980 or 1990) census.

[b]Predictor variables: (1) ratio of child exemptions reported by families in poverty on tax returns to total child exemptions (1989 or 1993); (2) ratio of people receiving food stamps (1989 or 1993) to total population; (3) ratio of total child exemptions on tax returns (1989 or 1993) to population (under age 21 or under age 18); (4) ratio of poor school-age children from previous (1980 or 1990) census.

TABLE C-2 Estimates of Regression Coefficients for the 1990 Census Equation for the 1993 Bivariate Models

| Model | Predictor Variables[a] | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Bivariate Log Number (under 21) | 0.71 | 0.31 | 0.48 | –0.51 |
| | (.01) | (.01) | (.03) | (.03) |
| Bivariate Log Number (under 21), Fixed State Effects | 0.71 | 0.33 | 0.45 | –0.48 |
| | (.02) | (.01) | (.03) | (.03) |
| | Predictor Variables[b] | | | |
| Bivariate Log Rate (under 21) | 0.66 | 0.30 | –0.23 | N.A. |
| | (.01) | (.01) | (.02) | |
| Bivariate Log Rate (under 21), Fixed State Effects | 0.67 | 0.30 | –0.22 | N.A. |
| | (.01) | (.01) | (.02) | |
| Bivariate Rate (under 21) | 0.56 | 0.75 | –0.05 | N.A. |
| | (.01) | (.01) | (.01) | |
| Bivariate Rate (under 21), Fixed State Effects | 0.55 | 0.78 | –0.05 | N.A. |
| | (.01) | (.02) | (.01) | |

NOTE: See notes to Table C-1.

[a]Predictor variables: (1) number of child exemptions reported by families in poverty on tax returns in 1989; (2) number of people receiving food stamps in 1989; (3) population under age 21 in 1990; (4) total number of child exemptions on tax returns in 1989.

[b]Predictor variables: (1) ratio of child exemptions reported by families in poverty on tax returns to total child exemptions in 1989; (2) ratio of people receiving food stamps in 1989 to total population; (3) ratio of total child exemptions on tax returns in 1989 to population under age 21.

***Inclusion or Exclusion of Predictor Variables*** All of the models with fixed state effects have a large fraction of state effects that are not significant at the 5 percent level. In addition, several other models, especially for 1993, had one or two predictor variables with regression coefficients that were not significant, but that was typically for only 1 of the 2 years that were analyzed. Therefore, except for the models with fixed state effects, there was little evidence of predictor variables that should be excluded from an equation. For the fixed state effects models, an examination of the extent to which the state effects cluster and could be estimated in groups might make it possible to reduce the number of coefficients that need to be estimated.

With respect to the need to include additional predictor variables in a model, nonrandom patterns of the distributions of the standardized residuals—especially a difference in the median standardized residual from 0 for the residuals in a county category—were observed for several characteristics: percent Hispanic population, location in a metropolitan area outside the central county, and population size. The models with the fewest nonrandom patterns of the distributions of the standardized residuals were the bivariate log rate, bivariate rate, and rate models.

***Normality*** Many of the models had distributions of the standardized residuals that were both asymmetric and long-tailed, especially to the side to which the distribution was skewed. It was difficult to distinguish between skewness and the presence of outliers. Often, movement from a log number dependent variable to a log rate dependent variable reduced an outlier problem, but it introduced a skewness problem. The rate models and the hybrid log rate-number model seemed to have both problems and to be particularly problematic in this respect. In contrast, the log number models behaved relatively well on this criterion.

***Homogeneous Variances*** All of the models exhibited nonconstant variances of the standardized residuals. One would expect the standardized residual variance to remain constant over the distribution of CPS sample size; however, for these models, it increased with increasing sample size. Most of the models also had some variance heterogeneity as a function of the predicted value (number or proportion of poor school-age children).

***Outliers*** The rate models and the hybrid log rate-number model exhibited both skewness and long-tailed error distributions. For all models, large urban counties, particularly those with large percentages of Hispanics, and counties that are in metropolitan areas but not the central county had somewhat more outliers than other counties. The bivariate log rate, bivariate log number, and the log rate models had fewer outliers that demonstrated these patterns.

***Additional Analysis*** Analysis that focused on a regression coefficient that is assumed to be constant in the single-equation formulation and is variable in the bivariate formulation demonstrated strong heterogeneity, thereby supporting the bivariate approach (see Appendix A). Also, Akaike's information criterion (AIC) confirmed the superiority of using the population under age 18 as a predictor variable in the log number model instead of the population under age 21.

## SUMMARY

Analysis of the regression output for the 13 county models for the most part supports the assumptions of the models; it does not strongly support one model over another. All of the models exhibit a few common problems. First, they all behave somewhat differently for larger urban counties, especially those with large percentages of Hispanics, than for rural counties. Second, all models show evidence of some variance heterogeneity, particularly with respect to CPS sample size and often with respect to the predicted value (number or proportion of poor school-age children). The rate models and the hybrid log rate-number model exhibit more problems with skewness and outliers than other model formulations. The bivariate approach appears promising due to the heterogeneity in the regression coefficient mentioned above, the lack of patterns in the analysis of the standardized residuals, and the correlation observed by corresponding residuals in the CPS and census regression equations. Finally, according to the internal evaluation, none of the alternative models is clearly superior to the log number model, and the use of the predictor variable for the population under age 18 instead of under age 21 is supported for the log number model.

# APPENDIX D
# County Model Comparisons with 1990 Census Estimates

An external evaluation of alternative models for producing county estimates of poor school-age children can be carried out by comparing the county estimates obtained from each model for 1989 with 1900 census estimates of related children 5-17 who were poor in 1989. Although this evaluation is not ideal, it serves as a valuable tool for model assessment.

Chapter 4 reports the results of such an evaluation for four candidate models and four procedures that rely more heavily on estimates from the 1980 census. This appendix supplements the material in Chapter 4 in two ways. First, it provides additional results for the four models and four procedures examined in Chapter 4. Second, it provides evaluation results for the six single-equation models that were considered in the first round of evaluations.

## EVALUATION MEASURES

Four measures are used for the evaluations in Chapter 4 and in this appendix. Two are overall measures of the differences between the county estimates from a model (or procedure) and the census, and two are measures for categories of counties. The four measures are defined as follows:

(1) *Average absolute difference*: the sum over all counties of the absolute (unsigned) difference between the model estimate of poor school-age children and the 1990 census estimate for each county, divided by the number of counties (3,141), or

$$\Sigma(|Y_{\text{model } i} - Y_{\text{census } i}|) \,/\, n \,.$$

(2) *Average proportional absolute difference*: the sum over all counties of the absolute difference between the model estimate of poor school-age children and the 1990 census estimate as a proportion of the census estimate for each county, divided by the number of counties,[1] or

$$\Sigma[(|Y_{\text{model } i} - Y_{\text{census } i}|) / Y_{\text{census } i}] / n.$$

(3) *Category algebraic difference*: the sum for all counties ($i$) in a category ($j$) of the algebraic (signed) difference between the model estimate of poor school-age children and the 1990 census estimate for each county in the category, divided by the sum of the census estimates for the counties in the category, or

$$\Sigma_i \ (Y_{\text{model } ij} - Y_{\text{census } ij}) \ / \Sigma_i \ Y_{\text{census } ij} \ .$$

(4) *Category average proportional algebraic difference*: the sum for all counties ($i$) in a category ($j$) of the algebraic difference between the model estimate of poor school-age children and the 1990 census estimate as a proportion of the census estimate for each county in the category, divided by the number of counties in the category, or

$$\Sigma_i \ [(Y_{\text{model } ij} - Y_{\text{census } ij}) \ / \ Y_{\text{census } ij} \ ] \ / \ n_j \ .$$

Measure (1) expresses overall absolute model-census differences in terms of numbers of poor school-age children; measure (2) expresses overall absolute model-census differences in terms of percentage errors for counties. Similarly, for categories of counties, measure (3) expresses model-census differences in terms of numbers of poor school-age children, while measure (4) expresses model-census differences in terms of percentage errors for counties. The two kinds of category differences are algebraic (not absolute) measures, in which positive differences offset negative differences.

For measures (3) and (4), the counties are grouped into categories of the following characteristics: census division; metropolitan status of county; population size in 1990; population growth from 1980 to 1990; percent poor school-age children in the 1980 census; percent Hispanic population in 1990; percent black population in 1990; persistent poverty from 1960 to 1990 for rural counties; economic type for rural counties; percent group quarters residents in 1990; whether the county had households in the CPS sample; and percent change from 1980 to 1990 in the proportion of poor school-age children.[2] Tables D-1 and D-2 show the number of counties in each category.

---

[1]An analogous measure, shown in Table 4-2, is the average proportional absolute difference in estimated proportions of poor school-age children.

[2]The characteristic of percent change in the proportion of poor school-age children from 1980 to 1990 was not included in the first round of evaluations.

## COMPARISONS FOR CANDIDATE MODELS AND OTHER ESTIMATION PROCEDURES

The four candidate models considered in Chapter 4 have the following designations: (a) log number model (under 21); (b) log number model (under 18); (c) log rate model (under 21); and (d) log rate model (under 18).[3] The four other procedures (see Chapter 4) are designated as follows: (i) stable shares; (ii) stable shares within state; (iii) stable rates within state (with conversion); and (iv) average of 1980 census estimates and estimates for 1989 from the log number (under 21) model (a).

Table 4-2 presents the overall measures of average absolute difference (measure 1) and average proportional absolute difference (measure 2) between the estimates from the four candidate models and four procedures and the estimates from the census. Table 4-3 presents the category algebraic differences (measure 3) for the four candidate models and procedures (i) and (iv). Table D-1 is identical to Table 4-3 except that it also includes results for procedures (ii) and (iii). Table D-2 presents the category average proportional algebraic differences for the four candidate models and the four procedures. For reasons given in Chapter 4, the 1990 census estimates used in these comparisons are ratio-adjusted by a constant factor to equal the CPS national estimate of poor school-age children in 1989.

The findings from these evaluations are discussed in Chapter 4. The additional detail in Tables D-1 and D-2 is presented without commentary.

[3]The estimates from the four candidate models and the models considered in the first round of evaluations, listed below, are the final estimates for all counties, after the initial estimates from the county regression model are combined in a “shrinkage procedure” with direct CPS estimates for those counties with households in the CPS sample and raked for consistency with the estimates from the state model; see Chapter 2.

TABLE D-1 Comparison of Model Estimates and Other Procedures with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989: Algebraic Difference by Category of County (in percent)

| | | Model | | | |
|---|---|---|---|---|---|
| Category | Counties[a] (Number) | Log No. Under 21 (a) | Log No. Under 18 (b) | Log Rate Under 21 (c) | Log Rate Under 18 (d) |
| Census Division[b] | | | | | |
| New England | 67 | –2.9 | –2.9 | –2.9 | –2.9 |
| Middle Atlantic | 150 | –2.8 | –2.8 | –2.8 | –2.8 |
| East North Central | 437 | –0.2 | –0.2 | –0.2 | –0.2 |
| West North Central | 618 | 1.7 | 1.7 | 1.7 | 1.7 |
| South Atlantic | 591 | 0.5 | 0.5 | 0.5 | 0.5 |
| East South Central | 364 | –4.5 | –4.5 | –4.5 | –4.5 |
| West South Central | 470 | –2.7 | –2.7 | –2.7 | –2.7 |
| Mountain | 281 | 4.3 | 4.3 | 4.3 | 4.3 |
| Pacific | 163 | 6.5 | 6.5 | 6.5 | 6.5 |
| Metropolitan Status | | | | | |
| Central county of metropolitan area | 493 | 2.4 | 1.6 | –0.1 | –0.5 |
| Other metropolitan | 254 | –6.6 | –5.0 | 5.1 | 6.3 |
| Nonmetropolitan | 2394 | –4.2 | –2.8 | –0.3 | 0.4 |
| 1990 Population Size | | | | | |
| under 7,500 | 525 | –9.0 | –2.3 | –1.9 | 2.3 |
| 7,500-14,999 | 630 | –4.4 | 0.5 | 2.5 | 5.5 |
| 15,000-24,999 | 524 | –5.1 | –2.6 | 0.3 | 1.9 |
| 25,000-49,999 | 620 | –4.2 | –2.9 | 0.6 | 1.3 |
| 50,000-99,999 | 384 | –3.5 | –5.1 | –1.2 | –2.3 |
| 100,000-249,999 | 259 | –1.8 | –4.4 | –1.8 | –3.5 |
| 250,000 or more | 199 | 3.3 | 3.2 | 0.5 | 0.5 |
| 1980 to 1990 Population Growth | | | | | |
| Decrease of more than 10.0% | 444 | –1.9 | 0.6 | –3.4 | –1.9 |
| Decrease 0.1-10.0% | 972 | –0.6 | –0.5 | –1.9 | –1.8 |
| 0.0-4.9% | 547 | –2.8 | –2.8 | –3.2 | –3.1 |
| 5.0-14.9% | 620 | 0.0 | –1.0 | 0.2 | –0.6 |
| 15.0-24.9% | 260 | 7.7 | 5.8 | 5.5 | 4.6 |
| 25.0% or more | 292 | –4.0 | –1.4 | 1.7 | 3.1 |

| Other Procedures | | | |
|---|---|---|---|
| Stable Shares (i) | Stable Shares in State (ii) | Stable Rates in State (iii) | Average of Census and (a) (iv) |
| 35.9 | –2.9 | –2.9 | 7.8 |
| 27.1 | –2.8 | –2.8 | 4.4 |
| –2.8 | –0.2 | –0.2 | –5.6 |
| –1.8 | 1.7 | 1.7 | –2.1 |
| 14.8 | 0.5 | 0.5 | 8.1 |
| 14.1 | –4.5 | –4.5 | 2.1 |
| –18.1 | –2.7 | –2.7 | –6.3 |
| –23.2 | 4.3 | 4.3 | –3.1 |
| –21.3 | 6.5 | 6.5 | 0.2 |
| | | | |
| –1.6 | –0.6 | –0.4 | 0.4 |
| 3.2 | –1.6 | 10.1 | 3.4 |
| 3.3 | 1.8 | –0.5 | –1.4 |
| | | | |
| 16.5 | 23.0 | 9.4 | 1.3 |
| 10.9 | 10.7 | 4.4 | 2.2 |
| 6.2 | 3.4 | 0.0 | –0.6 |
| 2.4 | –0.2 | –0.3 | –1.3 |
| –2.5 | –4.8 | –2.5 | –3.3 |
| –4.9 | –5.9 | –2.9 | –3.3 |
| –0.6 | 0.8 | 0.8 | 1.8 |
| | | | |
| 9.1 | 9.9 | –3.1 | –3.4 |
| 7.5 | 0.7 | –4.6 | –2.7 |
| 11.0 | –2.3 | –3.3 | –0.2 |
| 6.1 | 0.2 | 1.7 | 2.1 |
| –12.8 | 4.4 | 3.5 | 2.4 |
| –21.2 | –6.8 | 7.2 | 1.0 |

*continued on next page*

TABLE D-1 Continued

| Category | Counties[a] (Number) | Model: Log No. Under 21 (a) | Log No. Under 18 (b) | Log Rate Under 21 (c) | Log Rate Under 18 (d) |
|---|---|---|---|---|---|
| Percent Poor School-Age Children, 1980 | | | | | |
| Less than 9.4% | 516 | –4.0 | –4.5 | 0.0 | 0.2 |
| 9.4-11.6% | 524 | –0.5 | –1.0 | –1.6 | –1.8 |
| 11.7-14.1% | 530 | 3.6 | 2.3 | 1.8 | 1.0 |
| 14.2-17.2% | 523 | 0.9 | 1.2 | –1.2 | –1.4 |
| 17.3-22.3% | 519 | 1.8 | 1.7 | 0.3 | –0.1 |
| 22.4-53.0% | 523 | –2.2 | 0.8 | 1.3 | 2.8 |
| Percent Hispanic, 1990 | | | | | |
| 0.0-0.9% | 1770 | –3.4 | –3.3 | –1.6 | –1.5 |
| 1.0-4.9% | 847 | 0.5 | 0.1 | 0.4 | 0.1 |
| 5.0-9.9% | 193 | –1.4 | –0.6 | –1.1 | –0.8 |
| 10.0-24.9% | 181 | 2.2 | 1.8 | 0.7 | 0.5 |
| 25.0-98.0% | 150 | 3.9 | 4.6 | 2.2 | 2.7 |
| Percent Black, 1990 | | | | | |
| 0.0-0.9% | 1446 | –1.2 | 0.3 | 3.9 | 4.9 |
| 1.0-4.9% | 615 | –0.7 | –2.0 | 1.3 | 0.5 |
| 5.0-9.9% | 294 | –2.9 | –2.5 | –0.7 | –0.6 |
| 10.0-24.9% | 381 | 2.0 | 1.2 | –1.0 | –1.3 |
| 25.0-87.0% | 405 | 1.0 | 1.7 | –1.8 | –1.4 |
| Persistent Rural Poverty, 1960-1990[c] | | | | | |
| Rural, not poor | 1740 | –4.0 | –3.7 | –1.2 | –1.0 |
| Rural, poor | 535 | –5.0 | –2.1 | 0.7 | 2.1 |
| Not classified | 866 | 1.7 | 1.2 | 0.3 | 0.0 |
| Economic Type, Rural Counties[c] | | | | | |
| Farming | 556 | –5.5 | –2.5 | –1.6 | 0.7 |
| Mining | 146 | –10.7 | –5.1 | –6.3 | –3.6 |
| Manufacturing | 506 | –6.2 | –5.9 | –1.7 | –1.0 |
| Government | 243 | 2.1 | –1.3 | 6.3 | 3.2 |
| Services | 323 | –3.9 | –3.0 | –1.8 | –1.2 |
| Nonspecialized | 484 | –3.7 | –1.0 | –0.1 | 1.4 |
| Not classified | 883 | 1.7 | 1.2 | 0.3 | 0.0 |

| Other Procedures | | | |
|---|---|---|---|
| Stable Shares (i) | Stable Shares in State (ii) | Stable Rates in State (iii) | Average of Census and (a) (iv) |
| | | | |
| 2.4 | 0.8 | 5.1 | –1.1 |
| –9.9 | –4.0 | –1.9 | –3.6 |
| –4.2 | 1.8 | 0.7 | 0.2 |
| –5.0 | –3.0 | –5.3 | –1.8 |
| 10.7 | 1.9 | –0.1 | 4.2 |
| 12.3 | 4.1 | 1.8 | 4.1 |
| | | | |
| 10.7 | –0.6 | –1.4 | 0.2 |
| 0.2 | 0.1 | 1.1 | –0.4 |
| 6.7 | 1.2 | 1.4 | 1.7 |
| –5.7 | 1.7 | 1.3 | 0.1 |
| –16.8 | –1.2 | –1.3 | –0.4 |
| | | | |
| –3.7 | 3.9 | 6.0 | –0.5 |
| –6.3 | –1.6 | –0.4 | –2.9 |
| –8.4 | –2.3 | 2.2 | –1.8 |
| –2.6 | –0.7 | –2.1 | 0.2 |
| 16.5 | 1.2 | –2.4 | 3.7 |
| | | | |
| 0.1 | 0.2 | –1.4 | –3.4 |
| 9.8 | 5.4 | 0.1 | 1.2 |
| –1.2 | –0.7 | 0.4 | 0.7 |
| | | | |
| 13.2 | 18.0 | 7.9 | 1.1 |
| –8.9 | –6.6 | –13.1 | –10.6 |
| 12.1 | 0.8 | –1.1 | –0.2 |
| –0.9 | 4.6 | 4.1 | 0.0 |
| –5.8 | –4.0 | –3.4 | –4.3 |
| 2.2 | 1.6 | –2.0 | –1.5 |
| –1.2 | –0.7 | 0.4 | 0.7 |

*continued on next page*

TABLE D-1 Continued

| Category | Counties[a] (Number) | Model: Log No. Under 21 (a) | Model: Log No. Under 18 (b) | Model: Log Rate Under 21 (c) | Model: Log Rate Under 18 (d) |
|---|---|---|---|---|---|
| Percent Group Quarters Residents, 1990 | | | | | |
| Less than 1.0% | 545 | –6.7 | –2.7 | 2.0 | 4.7 |
| 1.0-4.9% | 2187 | 0.3 | 0.7 | –0.3 | 0.1 |
| 5.0-9.9% | 299 | 2.3 | –4.4 | 0.5 | –5.2 |
| 10.0-41.0% | 110 | 14.2 | –3.2 | 7.4 | –7.5 |
| Status in CPS, 1989-1991 | | | | | |
| In CPS sample | 1028 | 1.4 | 1.0 | –0.2 | –0.5 |
| In CPS, no poor children 5-17 | 246 | –2.6 | –1.9 | 7.3 | 7.8 |
| Not in CPS sample | 1867 | –4.1 | –2.8 | –0.1 | 0.6 |
| Change in Poverty Rate for School-Age Children, 1980-1990 | | | | | |
| Decrease of more than 3.0% | 536 | 7.5 | 10.4 | 16.2 | 18.1 |
| Decrease 0.1-3.0% | 649 | 2.1 | 1.9 | 3.1 | 2.9 |
| 0.0-0.9% | 272 | –2.6 | –0.8 | –0.4 | 0.5 |
| 1.0-3.4% | 621 | 3.8 | 2.2 | 3.4 | 2.6 |
| 3.5-6.4% | 532 | –1.2 | –2.4 | –3.8 | –4.3 |
| 6.5-38.0% | 523 | –7.2 | –5.2 | –8.7 | –7.8 |

NOTES: The census estimates are controlled to the CPS national estimate for 1989. The algebraic difference by category is the sum for all counties in a category of the algebraic (signed) difference between the model estimate of poor school-age children and the 1990 census estimate for each county, divided by the sum of the census estimates for all counties in the category. See Chapter 4 text for definitions of models.

[a]3,141 counties are assigned to a category for most characteristics; 3,135 counties are assigned to a category for 1980-1990 population growth and 1980 percent poor school-age children; 3,133 counties are assigned to a category for 1980-1990 percent change in poverty rate for school-age children.

| Other Procedures | | | |
|---|---|---|---|
| Stable Shares (i) | Stable Shares in State (ii) | Stable Rates in State (iii) | Average of Census and (a) (iv) |
| –1.4 | –0.9 | 3.7 | 0.3 |
| –0.4 | 0.3 | –0.1 | 0.1 |
| 7.8 | –1.4 | –2.8 | –0.8 |
| 1.8 | –0.9 | –1.4 | –2.2 |
| –0.6 | –0.7 | –0.4 | 0.5 |
| 10.0 | 3.7 | 12.0 | 5.9 |
| 0.6 | 2.3 | –0.3 | –2.3 |
| 51.6 | 30.1 | 32.8 | 30.0 |
| 29.2 | 8.0 | 9.8 | 12.1 |
| 4.3 | –0.9 | 3.3 | 3.1 |
| –5.1 | 3.7 | 3.4 | 0.2 |
| –14.3 | –7.7 | –9.5 | –8.3 |
| –25.2 | –14.2 | –16.5 | –14.5 |

[b]Census division states:

New England: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
Middle Atlantic: New York, New Jersey, Pennsylvania
East North Central: Ohio, Indiana, Illinois, Michigan, Wisconsin
West North Central: Missouri, Minnesota, Iowa, North Dakota, South Dakota, Nebraska, Kansas
South Atlantic: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida
East South Central: Kentucky, Tennessee, Alabama, Mississippi
West South Central: Arkansas, Louisiana, Oklahoma, Texas
Mountain: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada
Pacific: Washington, Oregon, California, Alaska, Hawaii

[c]The Economic Research Service, U.S. Department of Agriculture, classifies rural counties by 1960-1990 poverty status and economic type. Counties not classified are urban counties and rural counties for which a classification could not be made.

SOURCE: Data from the Bureau of the Census.

TABLE D-2 Comparison of Model Estimates and Other Proceduree with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989: Average Proportional Algebraic Difference for Counties in Each Category (in percent)

| | | Model | | | |
|---|---|---|---|---|---|
| Category | Counties (Number) | Log No. Under 21 (a) | Log No. Under 18 (b) | Log Rate Under 21 (c) | Log Rate Under 18 (d) |
| Census Division | | | | | |
| New England | 67 | 4.1 | 4.5 | 6.6 | 7.1 |
| Middle Atlantic | 150 | –5.9 | –8.4 | 0.7 | –1.0 |
| East North Central | 437 | –3.6 | –3.0 | 2.5 | 3.0 |
| West North Central | 618 | –3.1 | –0.6 | 0.5 | 2.3 |
| South Atlantic | 591 | 1.2 | 2.5 | 8.9 | 9.8 |
| East South Central | 364 | –4.6 | –3.0 | 0.5 | 1.3 |
| West South Central | 470 | –7.6 | –4.6 | –4.0 | –2.3 |
| Mountain | 281 | 0.6 | 5.4 | 7.2 | 10.4 |
| Pacific | 163 | 10.2 | 13.6 | 17.8 | 20.2 |
| Metropolitan Status | | | | | |
| Central county of metropolitan area | 493 | 0.6 | –2.0 | 1.0 | –0.6 |
| Other metropolitan | 254 | –3.6 | –0.8 | 11.6 | 13.7 |
| Nonmetropolitan | 2394 | –2.6 | 0.2 | 2.9 | 4.7 |
| 1990 Population Size | | | | | |
| under 7,500 | 525 | –5.9 | 1.6 | 2.6 | 7.6 |
| 7,500-14,999 | 630 | –1.0 | 3.0 | 5.7 | 8.4 |
| 15,000-24,999 | 524 | –3.2 | –1.8 | 2.1 | 3.2 |
| 25,000-49,999 | 620 | –1.5 | –0.7 | 4.2 | 4.6 |
| 50,000-99,999 | 384 | –1.4 | –3.3 | 2.5 | 1.2 |
| 100,000-249,999 | 259 | –0.7 | –3.4 | 1.5 | –0.3 |
| 250,000 or more | 199 | 1.0 | 0.4 | 1.3 | 1.1 |
| 1980 to 1990 Population Growth | | | | | |
| Decrease of more than 10.0% | 444 | –5.2 | –1.0 | –1.2 | 2.0 |
| Decrease 0.1-10.0% | 972 | –3.3 | –2.2 | 0.1 | 0.9 |
| 0.0-4.9% | 547 | –1.3 | 0.4 | 4.0 | 5.0 |
| 5.0-14.9% | 620 | –0.7 | 0.0 | 4.7 | 5.0 |
| 15.0-24.9% | 260 | 4.0 | 3.8 | 10.6 | 10.1 |
| 25.0% or more | 292 | –4.1 | 2.3 | 9.8 | 14.0 |

| Other Procedures | | | |
|---|---|---|---|
| Stable Shares (i) | Stable Shares in State (ii) | Stable Rates in State (iii) | Average of Census and (a) (iv) |
| 45.6 | 7.0 | 8.6 | 20.2 |
| 28.8 | –0.2 | 3.1 | 3.6 |
| 0.6 | 3.5 | 5.8 | –4.6 |
| 18.7 | 21.0 | 15.9 | 3.7 |
| 28.6 | 10.2 | 11.9 | 14.5 |
| 19.5 | 0.4 | 0.3 | 5.0 |
| –6.4 | 8.8 | –0.2 | –5.5 |
| –3.4 | 30.5 | 22.6 | 2.6 |
| –9.6 | 23.9 | 20.6 | 7.5 |
| | | | |
| 4.2 | –0.2 | 2.2 | 0.8 |
| 16.2 | 7.0 | 20.9 | 11.7 |
| 13.2 | 15.0 | 9.9 | 3.6 |
| | | | |
| 30.3 | 42.0 | 25.9 | 9.2 |
| 16.3 | 17.5 | 12.2 | 6.1 |
| 9.0 | 6.8 | 4.5 | 1.1 |
| 6.0 | 3.1 | 5.3 | 2.2 |
| 3.1 | –1.7 | 3.3 | 0.8 |
| 2.4 | –2.5 | 2.8 | 0.8 |
| 7.9 | 2.9 | 6.5 | 4.5 |
| | | | |
| 29.0 | 36.9 | 17.5 | 3.7 |
| 11.6 | 10.1 | 3.0 | –0.8 |
| 11.7 | 7.5 | 5.2 | 3.3 |
| 9.9 | 6.1 | 8.7 | 4.8 |
| 8.7 | 8.7 | 16.0 | 10.4 |
| –4.0 | 4.3 | 23.8 | 12.6 |

*continued on next page*

TABLE D-2 Continued

| Category | Counties (Number) | Model: Log No. Under 21 (a) | Log No. Under 18 (b) | Log Rate Under 21 (c) | Log Rate Under 18 (d) |
|---|---|---|---|---|---|
| Percent Poor School-Age Children, 1980 | | | | | |
| Less than 9.4% | 516 | –4.1 | –3.0 | 3.7 | 5.2 |
| 9.4-11.6% | 524 | –1.7 | –0.2 | 2.4 | 3.6 |
| 11.7-14.1% | 530 | –2.0 | –1.2 | 1.4 | 2.0 |
| 14.2-17.2% | 523 | –0.3 | 0.8 | 3.9 | 4.7 |
| 17.3-22.3% | 519 | –2.6 | –1.2 | 1.9 | 2.6 |
| 22.4-53.0% | 523 | –2.3 | 3.2 | 6.3 | 9.3 |
| Percent Hispanic, 1990 | | | | | |
| 0.0-0.9% | 1770 | –3.2 | –1.4 | 2.6 | 3.9 |
| 1.0-4.9% | 847 | 1.0 | 3.1 | 7.1 | 8.3 |
| 5.0-9.9% | 193 | –0.6 | 0.7 | 2.2 | 3.3 |
| 10.0-24.9% | 181 | –5.7 | –3.0 | –2.9 | –1.2 |
| 25.0-98.0% | 150 | –6.2 | –3.3 | –2.2 | –0.3 |
| Percent Black, 1990 | | | | | |
| 0.0-0.9% | 1446 | –2.4 | 1.4 | 4.0 | 6.7 |
| 1.0-4.9% | 615 | –1.4 | –2.1 | 3.1 | 2.4 |
| 5.0-9.9% | 294 | –2.4 | –2.4 | 2.6 | 2.6 |
| 10.0-24.9% | 381 | –0.7 | 0.6 | 4.7 | 5.4 |
| 25.0-87.0% | 405 | –3.8 | –2.7 | 0.0 | 0.9 |
| Persistent Rural Poverty, 1960-1990 | | | | | |
| Rural, not poor | 1740 | –2.6 | 0.0 | 2.3 | 4.1 |
| Rural, poor | 535 | –3.7 | 0.3 | 3.5 | 5.5 |
| Not classified | 866 | –0.4 | –1.1 | 5.2 | 4.8 |
| Economic Type, Rural Counties | | | | | |
| Farming | 556 | –5.2 | 0.3 | 0.3 | 4.2 |
| Mining | 146 | –8.6 | –1.2 | –1.7 | 2.2 |
| Manufacturing | 506 | –3.8 | –2.2 | 2.6 | 3.9 |
| Government | 243 | 5.8 | 5.1 | 11.8 | 10.5 |
| Services | 323 | –2.1 | –0.4 | 1.6 | 2.7 |
| Nonspecialized | 484 | –2.8 | –0.1 | 1.9 | 3.7 |
| Not classified | 883 | –0.1 | –0.8 | 5.4 | 5.1 |

| Other Procedures | | | |
|---|---|---|---|
| Stable Shares (i) | Stable Shares in State (ii) | Stable Rates in State (iii) | Average of Census and (a) (iv) |
| 1.9 | 2.9 | 8.1 | –0.4 |
| 3.5 | 6.0 | 6.1 | 0.6 |
| 5.6 | 8.3 | 6.2 | 0.5 |
| 15.6 | 17.0 | 13.6 | 6.0 |
| 17.0 | 15.1 | 9.8 | 5.1 |
| 28.7 | 22.4 | 13.6 | 11.1 |
| | | | |
| 20.7 | 12.1 | 10.2 | 5.4 |
| 4.7 | 10.4 | 11.0 | 4.5 |
| –0.6 | 15.4 | 10.2 | 1.0 |
| –7.1 | 14.8 | 5.1 | –3.5 |
| –10.0 | 11.7 | –1.2 | –5.8 |
| | | | |
| 12.7 | 19.9 | 15.9 | 4.1 |
| 5.3 | 5.1 | 3.8 | 0.3 |
| 5.7 | 3.2 | 4.9 | 2.4 |
| 13.8 | 5.9 | 8.0 | 8.0 |
| 23.1 | 6.2 | 0.5 | 5.3 |
| | | | |
| 12.5 | 16.4 | 11.4 | 3.0 |
| 16.2 | 12.0 | 4.0 | 4.4 |
| 8.6 | 3.0 | 9.3 | 5.1 |
| | | | |
| 29.0 | 37.3 | 22.6 | 7.5 |
| –2.4 | 11.9 | 3.3 | –4.0 |
| 17.3 | 7.0 | 5.1 | 4.0 |
| 5.8 | 12.1 | 9.3 | 5.0 |
| 2.6 | 6.4 | 5.9 | 0.4 |
| 6.8 | 7.1 | 3.7 | 0.8 |
| 8.8 | 3.5 | 9.6 | 5.3 |

*continued on next page*

TABLE D-2 Continued

| | | Model | | | |
|---|---|---|---|---|---|
| Category | Counties (Number) | Log No. Under 21 (a) | Log No. Under 18 (b) | Log Rate Under 21 (c) | Log Rate Under 18 (d) |
| Percent Group Quarters Residents, 1990 | | | | | |
| Less than 1.0% | 545 | –5.7 | 2.5 | 6.1 | 11.4 |
| 1.0-4.9% | 2187 | –3.1 | –0.6 | 1.7 | 3.7 |
| 5.0-9.9% | 299 | 5.2 | –0.6 | 6.7 | 1.7 |
| 10.0-41.0% | 110 | 13.8 | –5.0 | 11.5 | –3.9 |
| Status in CPS, 1989-1991 | | | | | |
| In CPS sample | 1028 | –0.9 | –1.3 | 1.9 | 1.7 |
| In CPS, no poor children 5-17 | 246 | –1.3 | 1.0 | 9.9 | 11.6 |
| Not in CPS sample | 1867 | –3.0 | 0.2 | 3.1 | 5.2 |
| Change in Poverty Rate for School-Age Children, 1980-1990 | | | | | |
| Decrease of more than 3.0% | 536 | 12.5 | 19.1 | 25.6 | 30.0 |
| Decrease 0.1-3.0% | 649 | 2.0 | 3.6 | 9.2 | 10.3 |
| 0.0-0.9% | 272 | –0.9 | –0.1 | 4.9 | 5.4 |
| 1.0-3.4% | 621 | –3.7 | –4.0 | –0.3 | –0.4 |
| 3.5-6.4% | 532 | –7.8 | –7.7 | –6.3 | –6.2 |
| 6.5-38.0% | 523 | –15.5 | –12.9 | –13.8 | –12.3 |

NOTE: See Notes to Table D-1.
SOURCE: Data from the Bureau of the Census.

## COMPARISONS FOR THE SINGLE-EQUATION MODELS CONSIDERED IN THE FIRST ROUND OF EVALUATIONS

Six single-equation models were considered in the first round of evaluations (see Chapter 3). For this appendix these models are labeled as follows: (D.1) log number model (under 21) (model (a) of the candidate models); (D.2) log number model (under 18) (model (b) of the candidate models); (D.3) log number model (under 21) with fixed state effects; (D.4) log rate model (under 21) (model (c) of the candidate models); (D.5) rate model (under 21, variables not transformed); and (D.6) hybrid log rate-number model (under 21).[4] Also included are compari-

[4]The "under 21" designation is retained in the discussion only for the log number model, D.1, to distinguish it from model D.2.

| Other Procedures | | | |
|---|---|---|---|
| Stable Shares (i) | Stable Shares in State (ii) | Stable Rates in State (iii) | Average of Census and (a) (iv) |
| 16.4 | 17.6 | 15.8 | 8.4 |
| 11.3 | 11.2 | 8.6 | 3.0 |
| 11.5 | 9.6 | 6.7 | 3.0 |
| 7.7 | 6.4 | 5.0 | –0.7 |
| 7.9 | 2.8 | 4.4 | 2.7 |
| 20.5 | 11.2 | 19.0 | 11.3 |
| 13.2 | 17.2 | 11.2 | 3.5 |
| 71.8 | 65.8 | 61.7 | 41.4 |
| 28.1 | 19.2 | 20.6 | 13.9 |
| 9.5 | 9.8 | 9.3 | 3.5 |
| –0.9 | 1.9 | 0.1 | –4.2 |
| –13.4 | –8.2 | –12.4 | –12.6 |
| –26.5 | –18.6 | –23.7 | –20.9 |

sons for a variant of each of the three rate models—D.4a, D.5a, and D.6a, respectively—in which 1990 census population figures instead of estimates from the Census Bureau's population estimates program are used to convert the estimated proportions of poor school-age children from each rate model to estimated numbers.

For the first round of evaluations the census estimates were not ratio-adjusted to make the census national estimate of poor school-age children in 1989 equal to the corresponding CPS total for 1989, unlike the situation with the evaluations of the candidate models and other procedures described above. Thus, the results of the first round of evaluations given in Tables D-3 to D-5 cannot be directly compared with those for the later round. However, knowing that the

TABLE D-3 Comparison of First-Round Model Estimates with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989

| Model | | Average Absolute Difference | Average Proportional Absolute Difference, in Percent |
|---|---|---|---|
| D.1 | Log number model (under 21) | 284 | 15.7 |
| D.2 | Log number model (under 18) | 284 | 17.1 |
| D.3 | Log number model (under 21), with fixed state effects | 289 | 17.4 |
| D.4 | Log rate model (under 21), rates converted to numbers with 1990 population estimates | 285 | 18.9 |
| D.4a | Log rate model (under 21), rates converted to numbers with 1990 census estimates | 263 | 17.9 |
| D.5 | Rate model (under 21), untransformed, rates converted to numbers with 1990 population estimates | 325 | 20.0 |
| D.5a | Rate model (under 21), untransformed, rates converted to numbers with 1990 census estimates | 299 | 18.8 |
| D.6 | Hybrid log rate-number model (under 21), rates converted to numbers with 1990 population estimates | 298 | 17.1 |
| D.6a | Hybrid log rate-number model (under 21), rates converted to numbers with 1990 census estimates | 270 | 15.3 |

NOTE: See text for definitions of models and measures.
SOURCE: Data from the Bureau of the Census.

ratio-adjustment increased the census estimates by about 5 percent, it could be possible to make some rough comparisons.

## Overall Differences

Table D-3 presents the average absolute difference (measure 1) and the average proportional absolute difference (measure 2) between model estimates and 1990 census estimates of the number of poor school-age children in 1989 for the six single-equation models, D.1-D.6, that were included in the first round of county model evaluations. It also shows the two absolute difference measures for the variant of the three rate models, D.4a, D.5a, and D.6a, in which 1990 census population figures instead of estimates from the Census Bureau's population estimates program are used to convert estimated proportions to estimated numbers of poor school-age children.

For models D.1, D.2, D.3, D.4, D.5, and D.6, the average absolute difference ranges from 284 to 325, or 11-13 percent of the average number of poor school-age children per county for 1989 (about 2,500 children). For these six models, the average proportional absolute difference ranges from 15.7 to 20.0 percent. The log number (under 21) model (D.1) performs best; it has the lowest average proportional absolute difference and is tied with the log number (under 18) model (D.2) for the lowest average absolute difference. The rate model (D.5) performs worst; it has the largest differences on both measures.

Because the 1990 census estimates used in the comparisons for models D.1-D.6 are not ratio-adjusted to the CPS national estimate of poor school-age children in 1989, the absolute difference measures in Table D-3 are about 5 percent higher than they would be if the ratio-adjustment had been made.[5] For an evaluation of overall differences, controlling the 1990 census estimates to the CPS national estimate does not affect comparisons across models. However, for evaluation of category differences, there could be an effect.

## Use of 1990 Population Estimates

For rate models, it is necessary to use population estimates of the number of school-age children to convert estimated proportions to estimated numbers of poor school-age children. The population estimates themselves differ from 1990 census figures (see Appendix B). The use of 1990 population estimates instead of 1990 census figures to convert estimated proportions from the three rate models to estimated numbers increases the average absolute difference in the estimated number of poor school-age children by 8-10 percent and increases the average proportional absolute difference by about 6 percent for the log rate and rate models and 12 percent for the hybrid log rate-number model. (Compare the measures in Table D-3 for model D.4 and D.4a, for D.5 and D.5a, and for D.6 and D.6a.)

## Differences by Categories of Counties

Tables D-4 and D-5 (on pages 154-165) show the category algebraic differences (measure 3) and the category average proportional algebraic differences

[5]Comparing Tables D-3 and 4-2, the average absolute differences for models D.1, D.2, and D.4 from Table D-3 are 4 to 6 percent higher than the corresponding differences for models (a), (b), and (c) from Table 4-2; the average proportional absolute differences are 2 to 8 percent higher.

(measure 4), respectively, between model estimates and 1990 census estimates of the number of poor school-age children in 1989 for the six single-equation models that were considered in the first round of county model evaluations and the variant of the three rate models. The discussion considers models D.1-D.6.

***Census Division*** The category algebraic differences in the predicted number of poor school-age children categorized by census division (measure 3, Table D-4) are the same for all of the models because they are raked to the same set of state estimates. They vary widely by census division. In particular, all of the models overpredict the number of poor school-age children for counties in the Mountain Division and, especially, the Pacific Division relative to other counties. The proportional category differences (measure 4, Table D-5) vary even more widely across divisions than do the category differences. For the Pacific Division, the proportional category difference is 1.3 to 2 times the category difference (16-26% versus 12%), indicating that the overprediction is more pronounced for smaller counties than larger counties in that geographic area.[6] Further investigation is required to determine the reasons for the variations across divisions, which could include sampling variability in the CPS for 1989 or a specification problem in the state model (see Chapter 4).

***Metropolitan Status*** The category differences and proportional category differences in the predicted number of poor school-age children vary somewhat for counties categorized by metropolitan status. There is no consistent pattern across models: for example, the log number (under 21) model (D.1) overpredicts the number of poor school-age children in central counties of metropolitan areas relative to other counties, while the log rate model (D.4) overpredicts the number of poor school-age children in "other metropolitan" counties relative to central counties or counties in nonmetropolitan areas.

***1990 Population Size*** The category differences in the predicted number of poor school-age children (Table D-4) show a systematic tendency for the log number (under 21) model (D.1) and the hybrid log rate-number model (D.6) to overpredict the number of poor school-age children for larger size counties relative to smaller size counties. The proportional category differences (Table D-5) show somewhat less variation. A statistical test established that the variations in the proportional differences for categories of counties classified by population size were significant for model D.6, but not for model D.1. However, the test used was not sensitive to monotonic patterns—for example, an increasing rate of

---

[6]The proportional category differences differ somewhat across models because they are calculated relative to each county's 1990 census estimated number of poor school-age children before being summed.

overprediction by county size. (The test was not performed for the category differences, measure 3.)

***Population Growth from 1980-1990*** The category differences and proportional category differences in the predicted number of poor school-age children show a tendency for most models to overpredict the number of poor school-age children in counties with larger rates of population increase from 1980 to 1990 relative to counties with smaller increases or with decreases.[7] However, the extent of overprediction does not increase monotonically. In particular, most models underpredict the number of poor school-age children for counties with the largest population increases (25% or more) relative to counties with the next largest increases (15-25%). In contrast to the pattern shown by other models, the log number model with fixed state effects (D.3) tends to overpredict the number of poor school-age children for counties that experienced a large population decrease relative to other counties.

***Percent Poor School-Age Children, 1980 Census*** The category differences and proportional category differences in the predicted number of poor school-age children show relatively little variation for most models for counties categorized by their proportion of poor school-age children in 1979. The exception is the log number model with fixed state effects (D.3), which overpredicts the number of poor school-age children for counties that had a higher proportion of such children in 1979 relative to counties with a lower proportion. The variation in the proportional category differences (Table D-5) for counties defined by their 1979 proportion of poor school-age children is statistically significant for this model.

***Percent Hispanic Population in 1990*** The category differences in the predicted number of poor school-age children (Table D-4) show a tendency for most models to overpredict the number of poor school-age children for counties with larger proportions of Hispanics relative to other counties. This pattern is particularly pronounced for the log number (under 21 and under 18) models (D.1, D.2). The proportional category differences (Table D-5) tend to show the opposite pattern, in which the number of poor school-age children is overpredicted for counties with *smaller* proportions of Hispanics relative to other counties. The variations in the proportional category differences for counties characterized by percent Hispanic population are statistically significant for all models with this pattern that were tested. The differences in the patterns for the two measures may occur because the models behave differently for small counties with many Hispanics (primarily rural border counties) than for large counties (cities).

[7]A statistical test established that the variations in the proportional category differences for categories of counties classified by population growth rate were significant for three of the four models tested: D.1, D.2, and D.3, but not D.6.

***Percent Black Population in 1990*** The category differences in the predicted number of poor school-age children (Table D-4) show a slight tendency for the log rate and rate models (D.4, D.5) to overpredict the number of poor school-age children for counties with smaller proportions of blacks relative to other counties. The proportional category differences (Table D-5) show little variation for any of the models for counties characterized by percent black population in 1990.

***Persistent Rural Poverty, 1960-1990*** The category differences in the predicted number of poor school-age children (Table D-4) vary little for most models for counties characterized as rural and persistently poor, rural and not persistently poor, and not classified (urban counties and rural counties for which a classification could not be made). However, the log number (under 21) model (D.1) underpredicts the number of poor school-age children for rural counties relative to not classified counties. Also, the hybrid log rate-number model (D.6) underpredicts the number of poor school-age children for rural counties, whether or not they are persistently poor, relative to not classified counties. This pattern, which appears for both category difference measures, is statistically significant for the proportional category difference measure (Table D-5).

***Economic Type, Rural Counties*** The category differences and proportional category differences in the predicted number of poor school-age children vary for all models for rural counties categorized by their principal economic activity. In particular, all of the models overpredict the number of poor school-age children in rural counties that have a large government presence relative to other types of rural counties.

***Percent Group Quarters Residents in 1990*** The category differences and proportional category differences in the predicted number of poor school-age children show that the log number (under 21) model (D.1), log number model with fixed state effects (D.3), and log rate model (D.4) tend to overpredict the number of poor school-age children in counties with larger percentages of group quarters residents relative to other counties. The pattern is particularly strong for model D.1. As discussed in Chapter 4, the replacement of the population under age 21 as a predictor variable in model D.1 by the population under 18 in model D.2 removed this pattern.

***Status in CPS, 1989-1991*** The category differences and proportional category differences in the predicted number of poor school-age children are similar in most models for counties categorized by their representation in the CPS sample. The log rate model (D.4) overpredicts the number of poor school-age children in counties with CPS sampled households, none of which contain poor school-age children (and thereby are excluded from the sample for estimating the model),[8] relative to other counties. The hybrid log rate-number model (D.6) somewhat overpredicts the number of poor school-age children in counties with CPS sampled households relative to counties with no CPS sampled households.

## Summary of Category Differences

Three of the eleven characteristics examined show no pronounced patterns of overprediction or underprediction of the number of poor school-age children for any of the models:

—percent poor school-age children from the 1980 census;
—percent black population in 1990; and
—persistent rural poverty from 1960 to 1990.

Four characteristics show patterns for all or all but one model in which some categories of counties are over(under)predicted relative to other counties:

—census division;
—percent change in population from 1980 to 1990 (population growth);
—percent Hispanic population in 1990; and
—economic type, for rural counties.

The remaining four characteristics exhibit mixed patterns, in which some models give evidence of over(under)prediction for counties in some categories and other models do not:

—metropolitan status of county;
—1990 population size;
—percent group quarters residents in 1990; and
—status in CPS sample.

Of these four characteristics, over(under)prediction for those models in which it occurs is most pronounced for population size and percent group quarters residents.

Overall, there is no clearly best or worst model in terms of differences from the 1990 census estimates for categories of counties. Each model exhibits strengths and weaknesses (keeping in mind that the analysis is based on a single evaluation). On balance, the log number (under 18) model (D.2) performs somewhat better than the other models.

---

[8] The only model that uses these counties in the estimation is the rate model for which the variables are untransformed (D.5).

TABLE D-4 Comparison of First-Round Model Estimates with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989: Algebraic Difference by Category of County (in percent)

| | Model | | |
|---|---|---|---|
| Category | Log Number Under 21 D.1 | Log Number Under 18 D.2 | Log Number Under 21, Fixed State Effects D.3 |
| Census Division[a] | | | |
| New England | 1.9 | 1.9 | 1.9 |
| Middle Atlantic | 2.0 | 2.0 | 2.0 |
| East North Central | 4.7 | 4.7 | 4.7 |
| West North Central | 6.8 | 6.8 | 6.8 |
| South Atlantic | 5.5 | 5.5 | 5.5 |
| East South Central | 0.3 | 0.3 | 0.3 |
| West South Central | 2.1 | 2.1 | 2.1 |
| Mountain | 9.4 | 9.4 | 9.4 |
| Pacific | 11.8 | 11.8 | 11.8 |
| Metropolitan Status | | | |
| Central county of metropolitan area | 7.4 | 6.7 | 6.6 |
| Other metropolitan | –2.0 | –0.3 | –3.9 |
| Nonmetropolitan | 0.5 | 2.0 | 2.8 |
| 1990 Population Size | | | |
| under 7,500 | –4.5 | 2.5 | 4.7 |
| 7,500-14,999 | 0.4 | 5.5 | 6.0 |
| 15,000-24,999 | –0.4 | 2.3 | 2.8 |
| 25,000-49,999 | 0.5 | 1.8 | 1.9 |
| 50,000-99,999 | 1.2 | –0.4 | –0.1 |
| 100,000-249,999 | 3.1 | 0.4 | 1.1 |
| 250,000 or more | 8.4 | 8.3 | 7.9 |
| 1980 to 1990 Population Growth | | | |
| Decrease of more than 10.0% | 3.0 | 5.6 | 9.0 |
| Decrease 0.1-10.0% | 4.3 | 4.4 | 5.9 |
| 0.0-4.9% | 2.0 | 2.0 | 2.5 |
| 5.0-14.9% | 5.0 | 3.8 | 3.8 |
| 15.0-24.9% | 13.1 | 11.1 | 10.9 |
| 25.0% or more | 0.7 | 3.5 | –0.5 |

| Log Rate Under 21 | | Rate Under 21 | | Log Hybrid Rate-Number Under 21 | |
|---|---|---|---|---|---|
| D.4 | D.4a | D.5 | D.5a | D.6 | D.6a |
| 1.9 | 1.9 | 1.9 | 1.9 | 1.9 | 1.9 |
| 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 |
| 4.7 | 4.7 | 4.7 | 4.7 | 4.7 | 4.7 |
| 6.8 | 6.8 | 6.8 | 6.8 | 6.8 | 6.8 |
| 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 |
| 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 |
| 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 |
| 9.4 | 9.4 | 9.4 | 9.4 | 9.4 | 9.4 |
| 11.8 | 11.8 | 11.8 | 11.8 | 11.8 | 11.8 |
| | | | | | |
| 4.8 | 4.5 | 4.8 | 4.5 | 7.7 | 7.4 |
| 10.2 | 7.5 | 9.7 | 7.0 | 2.5 | –0.1 |
| 4.6 | 5.8 | 4.6 | 5.8 | –0.9 | 0.2 |
| | | | | | |
| 3.0 | 4.4 | 5.6 | 7.2 | –6.6 | –5.3 |
| 7.6 | 8.6 | 7.7 | 8.7 | –0.9 | 0.0 |
| 5.3 | 6.4 | 5.2 | 6.3 | –1.5 | –0.4 |
| 5.6 | 6.1 | 5.5 | 6.0 | 0.3 | 0.7 |
| 3.6 | 3.9 | 3.8 | 4.0 | 0.3 | 0.6 |
| 3.0 | 3.1 | 1.7 | 1.8 | 2.1 | 2.2 |
| 5.5 | 5.0 | 5.7 | 5.3 | 9.2 | 8.8 |
| | | | | | |
| 1.3 | 1.9 | 2.4 | 3.0 | 2.4 | 3.0 |
| 2.9 | 3.0 | 3.1 | 3.2 | 3.9 | 4.0 |
| 1.6 | 2.3 | 1.2 | 1.9 | 1.3 | 2.0 |
| 5.2 | 5.1 | 5.6 | 5.6 | 4.2 | 4.2 |
| 10.7 | 9.9 | 10.9 | 10.0 | 12.6 | 11.7 |
| 6.7 | 6.6 | 5.8 | 5.6 | 4.1 | 3.9 |

*continued on next page*

TABLE D-4 Continued

| Category | Model: Log Number Under 21 D.1 | Log Number Under 18 D.2 | Log Number Under 21, Fixed State Effects D.3 |
|---|---|---|---|
| Percent Poor School-Age Children, 1980 | | | |
| Less than 9.4% | 0.8 | 0.2 | –1.0 |
| 9.4-11.6% | 4.4 | 3.9 | 3.3 |
| 11.7-14.1% | 8.8 | 7.3 | 7.0 |
| 14.2-17.2% | 5.8 | 6.2 | 5.2 |
| 17.3-22.3% | 6.8 | 6.7 | 8.5 |
| 22.4-53.0% | 2.6 | 5.7 | 7.7 |
| Percent Hispanic, 1990 | | | |
| 0.0-0.9% | 1.4 | 1.4 | 2.3 |
| 1.0-4.9% | 5.5 | 5.0 | 4.7 |
| 5.0-9.9% | 3.5 | 4.3 | 3.3 |
| 10.0-24.9% | 7.3 | 6.8 | 7.4 |
| 25.0-98.0% | 9.0 | 9.8 | 8.5 |
| Percent Black, 1990 | | | |
| 0.0-0.9% | 3.6 | 5.2 | 5.3 |
| 1.0-4.9% | 4.2 | 2.8 | 2.9 |
| 5.0-9.9% | 1.9 | 2.4 | 1.5 |
| 10.0-24.9% | 7.0 | 6.2 | 5.7 |
| 25.0-87.0% | 6.0 | 6.7 | 7.9 |
| Persistent Rural Poverty, 1960-1990[b] | | | |
| Rural, not poor | 0.8 | 1.0 | 1.4 |
| Rural, poor | –0.3 | 2.7 | 5.2 |
| Not classified | 6.7 | 6.2 | 5.8 |
| Economic Type, Rural Counties[b] | | | |
| Farming | –0.8 | 2.4 | 7.0 |
| Mining | –6.3 | –0.4 | –4.0 |
| Manufacturing | –1.6 | –1.2 | 0.4 |
| Government | 7.2 | 3.6 | 8.7 |
| Services | 0.8 | 1.8 | 1.1 |
| Nonspecialized | 1.0 | 3.9 | 3.4 |
| Not classified | 6.7 | 6.2 | 5.8 |

| Log Rate Under 21 | | Rate Under 21 | | Log Hybrid Rate-Number Under 21 | |
|---|---|---|---|---|---|
| D.4 | D.4a | D.5 | D.5a | D.6 | D.6a |
| | | | | | |
| 4.9 | 1.7 | 5.6 | 2.3 | 5.6 | 2.3 |
| 3.2 | 3.0 | 4.4 | 4.2 | 5.6 | 5.3 |
| 6.8 | 6.9 | 6.2 | 6.4 | 7.5 | 7.6 |
| 3.7 | 6.7 | 2.8 | 5.7 | 2.7 | 5.7 |
| 5.3 | 5.8 | 4.4 | 4.8 | 5.0 | 5.6 |
| 6.3 | 6.8 | 6.2 | 6.7 | 2.1 | 2.7 |
| | | | | | |
| 3.3 | 3.1 | 3.1 | 3.0 | 1.6 | 1.4 |
| 5.4 | 5.1 | 5.1 | 4.8 | 5.6 | 5.3 |
| 3.8 | 3.4 | 5.0 | 4.7 | 4.4 | 3.9 |
| 5.7 | 5.1 | 7.2 | 6.4 | 8.2 | 7.6 |
| 7.2 | 8.9 | 5.9 | 7.7 | 6.9 | 8.6 |
| | | | | | |
| 9.0 | 9.1 | 8.6 | 8.7 | 4.3 | 4.3 |
| 6.3 | 5.6 | 6.9 | 6.1 | 4.8 | 4.0 |
| 4.2 | 3.6 | 4.1 | 3.6 | 4.3 | 3.6 |
| 3.9 | 3.8 | 3.9 | 3.8 | 6.5 | 6.3 |
| 3.1 | 4.2 | 2.9 | 4.1 | 4.2 | 5.5 |
| | | | | | |
| 3.6 | 5.4 | 3.6 | 5.3 | –1.1 | 0.5 |
| 5.7 | 5.9 | 5.7 | 5.8 | –1.4 | –1.2 |
| 5.2 | 4.8 | 5.2 | 4.8 | 7.2 | 6.9 |
| | | | | | |
| 3.3 | 5.2 | 5.0 | 6.9 | –3.9 | –2.1 |
| –1.7 | 1.5 | –1.4 | 1.8 | –6.0 | –3.1 |
| 3.2 | 3.1 | 3.4 | 3.3 | –1.5 | –1.7 |
| 11.6 | 11.7 | 9.7 | 9.7 | 1.9 | 2.0 |
| 3.1 | 4.8 | 3.1 | 4.8 | –0.5 | 1.2 |
| 4.8 | 6.8 | 4.4 | 6.3 | –0.4 | 1.4 |
| 5.2 | 4.8 | 5.3 | 4.8 | 7.3 | 6.9 |

*continued on next page*

TABLE D-4 Continued

| | Model | | |
|---|---|---|---|
| Category | Log Number Under 21 D.1 | Log Number Under 18 D.2 | Log Number Under 21, Fixed State Effects D.3 |
| Percent Group Quarters Residents, 1990 | | | |
| Less than 1.0% | –2.1 | 2.1 | –0.5 |
| 1.0-4.9% | 5.2 | 5.7 | 5.4 |
| 5.0-9.9% | 7.4 | 0.3 | 5.0 |
| 10.0-41.0% | 19.9 | 1.6 | 11.9 |
| Status in CPS, 1989-1991 | | | |
| In CPS sample | 6.4 | 5.9 | 5.8 |
| In CPS, no poor children 5-17 | 2.2 | 3.0 | 0.8 |
| Not in CPS sample | 0.6 | 2.0 | 2.8 |

NOTES: See text for definitions of models and measures. 3,141 counties are assigned to a category for most characteristics; 3,135 counties are assigned to a category for 1980-1990 population growth and 1980 percent poor school-age children; 3,133 counties are assigned to a category for 1980-1990 percent change in poverty rate for school-age children; see Table D-1 for number of counties in each category.

[a]Census division states:
New England: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
Middle Atlantic: New York, New Jersey, Pennsylvania
East North Central: Ohio, Indiana, Illinois, Michigan, Wisconsin
West North Central: Missouri, Minnesota, Iowa, North Dakota, South Dakota, Nebraska, Kansas

| Log Rate Under 21 | | Rate Under 21 | | Log Hybrid Rate-Number Under 21 | |
|---|---|---|---|---|---|
| D.4 | D.4a | D.5 | D.5a | D.6 | D.6a |
| 7.0 | 4.9 | 8.9 | 6.7 | 2.7 | 0.6 |
| 4.6 | 4.6 | 4.7 | 4.7 | 5.7 | 5.7 |
| 5.5 | 7.3 | 4.0 | 5.8 | 0.3 | 2.0 |
| 12.7 | 17.4 | 5.0 | 9.3 | 0.6 | 4.7 |
| 4.7 | 4.4 | 5.3 | 5.0 | 6.8 | 6.6 |
| 12.6 | 10.2 | –1.0 | –2.9 | 5.3 | 3.0 |
| 4.8 | 6.2 | 5.0 | 6.3 | –1.4 | –0.1 |

South Atlantic: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida
East South Central: Kentucky, Tennessee, Alabama, Mississippi
West South Central: Arkansas, Louisiana, Oklahoma, Texas
Mountain: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada
Pacific: Washington, Oregon, California, Alaska, Hawaii

[b]The Economic Research Service, U.S. Department of Agriculture, classifies rural counties by 1960-1990 poverty status and economic type. Counties not classified are urban counties and rural counties for which a classification could not be made.

SOURCE: Data from the Bureau of the Census.

TABLE D-5 Comparison of First-Round Model Estimates with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989: Average Proportional Algebraic Difference for Counties in Each Category (in percent)

| | Model | | |
|---|---|---|---|
| Category | Log Number Under 21 D.1 | Log Number Under 18 D.2 | Log Number Under 21, Fixed State Effects D.3 |
| Census Division | | | |
| New England | 9.3 | 9.7 | 8.1 |
| Middle Atlantic | –1.2 | –3.9 | –3.5 |
| East North Central | 1.2 | 1.8 | 2.2 |
| West North Central | 1.7 | 4.4 | 7.4 |
| South Atlantic | 6.2 | 7.6 | 8.1 |
| East South Central | 0.1 | 1.8 | 0.9 |
| West South Central | –3.0 | 0.1 | 0.2 |
| Mountain | 5.6 | 10.6 | 12.2 |
| Pacific | 15.6 | 19.2 | 19.2 |
| Metropolitan Status | | | |
| Central county of metropolitan area | 5.6 | 2.9 | 3.5 |
| Other metropolitan | 1.1 | 4.1 | –0.1 |
| Nonmetropolitan | 2.2 | 5.1 | 6.5 |
| 1990 Population Size | | | |
| under 7,500 | –1.3 | 6.6 | 9.9 |
| 7,500-14,999 | 3.9 | 8.1 | 9.3 |
| 15,000-24,999 | 1.6 | 3.0 | 4.2 |
| 25,000-49,999 | 3.4 | 4.2 | 3.7 |
| 50,000-99,999 | 3.4 | 1.5 | 1.0 |
| 100,000-249,999 | 4.2 | 1.4 | 1.4 |
| 250,000 or more | 5.9 | 5.4 | 5.0 |
| 1980 to 1990 Population Growth | | | |
| Decrease of more than 10.0% | –0.5 | 3.9 | 10.5 |
| Decrease 0.1-10.0% | 1.5 | 2.6 | 5.5 |
| 0.0-4.9% | 3.6 | 5.3 | 5.1 |
| 5.0-14.9% | 4.2 | 4.9 | 4.1 |
| 15.0-24.9% | 9.2 | 9.0 | 7.5 |
| 25.0% or more | 0.7 | 7.3 | –0.3 |

| Log Rate Under 21 | | Rate Under 21 | | Log Hybrid Rate-Number Under 21 | |
|---|---|---|---|---|---|
| D.4 | D.4a | D.5 | D.5a | D.6 | D.6a |
| 11.9 | 13.1 | 10.9 | 12.2 | 8.3 | 9.4 |
| 5.7 | 4.1 | 4.2 | 2.8 | –1.2 | –2.6 |
| 7.5 | 8.5 | 6.4 | 7.4 | –0.1 | 0.7 |
| 5.4 | 7.3 | 6.1 | 8.0 | –0.2 | 1.6 |
| 14.3 | 12.6 | 14.5 | 12.8 | 7.7 | 6.1 |
| 5.4 | 4.8 | 5.3 | 4.6 | 0.7 | 0.0 |
| 0.7 | 3.3 | 1.8 | 4.3 | –6.7 | –4.4 |
| 12.5 | 14.6 | 17.0 | 19.3 | 3.9 | 5.7 |
| 23.7 | 23.8 | 25.6 | 25.8 | 15.6 | 15.7 |
| | | | | | |
| 6.0 | 4.9 | 5.0 | 4.0 | 6.1 | 4.9 |
| 17.1 | 13.3 | 16.1 | 12.4 | 6.8 | 3.4 |
| 7.9 | 9.4 | 9.0 | 10.5 | 0.3 | 1.6 |
| | | | | | |
| 7.7 | 9.2 | 12.7 | 14.2 | –3.5 | –2.3 |
| 10.9 | 12.3 | 11.5 | 12.8 | 2.2 | 3.4 |
| 7.2 | 8.2 | 6.9 | 8.0 | 0.1 | 1.1 |
| 9.3 | 10.1 | 8.8 | 9.6 | 2.8 | 3.5 |
| 7.5 | 7.3 | 7.3 | 7.0 | 3.1 | 2.8 |
| 6.6 | 6.0 | 3.3 | 2.9 | 4.0 | 3.4 |
| 6.3 | 4.4 | 7.3 | 5.5 | 8.7 | 6.9 |
| | | | | | |
| 3.7 | 3.9 | 7.9 | 8.0 | –1.7 | –1.5 |
| 5.0 | 6.4 | 5.4 | 6.8 | –0.7 | 0.6 |
| 9.2 | 9.9 | 8.2 | 8.9 | 2.7 | 3.4 |
| 9.9 | 10.2 | 9.6 | 10.0 | 3.4 | 3.6 |
| 16.0 | 15.6 | 15.6 | 15.2 | 8.0 | 7.6 |
| 15.2 | 15.7 | 16.2 | 16.9 | 3.9 | 4.1 |

*continued on next page*

TABLE D-5 Continued

| | Model | | |
|---|---|---|---|
| Category | Log Number Under 21 D.1 | Log Number Under 18 D.2 | Log Number Under 21, Fixed State Effects D.3 |
| Percent Poor School-Age Children, 1980 | | | |
| Less than 9.4% | 0.6 | 1.8 | –1.3 |
| 9.4-11.6% | 3.2 | 4.8 | 3.5 |
| 11.7-14.1% | 2.9 | 3.6 | 4.3 |
| 14.2-17.2% | 4.6 | 5.8 | 8.1 |
| 17.3-22.3% | 2.2 | 3.7 | 6.9 |
| 22.4-53.0% | 2.5 | 8.3 | 11.2 |
| Percent Hispanic, 1990 | | | |
| 0.0-0.9% | 1.6 | 3.5 | 5.1 |
| 1.0-4.9% | 6.0 | 8.2 | 6.7 |
| 5.0-9.9% | 4.3 | 5.7 | 6.4 |
| 10.0-24.9% | –1.1 | 1.8 | 3.1 |
| 25.0-98.0% | –1.5 | 1.5 | 4.7 |
| Percent Black, 1990 | | | |
| 0.0-0.9% | 2.4 | 6.5 | 7.3 |
| 1.0-4.9% | 3.5 | 2.8 | 3.5 |
| 5.0-9.9% | 2.4 | 2.4 | 1.8 |
| 10.0-24.9% | 4.2 | 5.6 | 4.5 |
| 25.0-87.0% | 0.9 | 2.1 | 5.6 |
| Persistent Rural Poverty, 1960-1990 | | | |
| Rural, not poor | 2.2 | 4.9 | 6.1 |
| Rural, poor | 1.0 | 5.3 | 7.7 |
| Not classified | 4.5 | 3.8 | 2.9 |
| Economic Type, Rural Counties | | | |
| Farming | –0.5 | 5.3 | 9.9 |
| Mining | –4.1 | 3.7 | 0.7 |
| Manufacturing | 1.0 | 2.7 | 3.5 |
| Government | 11.0 | 10.3 | 13.2 |
| Services | 2.7 | 4.5 | 4.3 |
| Nonspecialized | 2.0 | 4.9 | 4.8 |
| Not classified | 4.8 | 4.1 | 3.3 |

| Log Rate Under 21 | | Rate Under 21 | | Log Hybrid Rate-Number Under 21 | |
|---|---|---|---|---|---|
| D.4 | D.4a | D.5 | D.5a | D.6 | D.6a |
| | | | | | |
| 8.9 | 7.5 | 7.2 | 5.9 | 3.8 | 2.5 |
| 7.5 | 9.0 | 8.5 | 10.2 | 2.3 | 3.7 |
| 6.4 | 7.7 | 7.3 | 8.6 | 0.9 | 2.2 |
| 9.1 | 11.1 | 10.2 | 12.3 | 1.4 | 3.2 |
| 7.0 | 8.1 | 7.5 | 8.6 | –0.6 | 0.4 |
| 11.5 | 10.6 | 13.0 | 11.0 | 2.4 | 1.5 |
| | | | | | |
| 7.7 | 7.6 | 8.4 | 8.2 | 1.5 | 1.4 |
| 12.4 | 12.9 | 12.1 | 12.7 | 5.6 | 6.1 |
| 7.2 | 10.0 | 9.3 | 12.4 | 0.6 | 3.0 |
| 1.9 | 5.4 | 3.3 | 6.9 | –6.0 | –2.9 |
| 2.6 | 7.4 | 3.8 | 8.7 | –7.7 | –3.5 |
| | | | | | |
| 9.2 | 10.4 | 10.4 | 11.7 | 1.5 | 2.6 |
| 8.2 | 8.7 | 8.0 | 8.7 | 2.0 | 2.4 |
| 7.7 | 6.6 | 7.9 | 6.9 | 3.1 | 2.1 |
| 9.9 | 9.8 | 9.0 | 9.0 | 3.9 | 3.7 |
| 5.0 | 5.4 | 5.9 | 6.2 | –1.0 | –0.5 |
| | | | | | |
| 7.3 | 9.4 | 8.7 | 10.8 | 0.1 | 1.9 |
| 8.6 | 8.3 | 8.4 | 8.1 | 0.0 | –0.2 |
| 10.3 | 8.7 | 9.7 | 8.2 | 6.0 | 4.5 |
| | | | | | |
| 5.3 | 7.6 | 9.3 | 11.6 | –3.5 | –1.3 |
| 3.1 | 8.6 | 4.5 | 10.3 | –6.7 | –2.0 |
| 7.6 | 7.6 | 7.5 | 7.4 | 1.2 | 1.1 |
| 17.3 | 17.2 | 15.0 | 14.8 | 7.2 | 7.0 |
| 6.6 | 8.2 | 7.9 | 9.6 | 0.8 | 2.3 |
| 7.0 | 8.7 | 6.9 | 8.6 | 0.3 | 2.0 |
| 10.6 | 9.0 | 10.1 | 8.6 | 6.3 | 4.8 |

*continued on next page*

TABLE D-5 Continued

| Category | Model: Log Number Under 21 D.1 | Log Number Under 18 D.2 | Log Number Under 21, Fixed State Effects D.3 |
|---|---|---|---|
| Percent Group Quarters Residents, 1990 | | | |
| Less than 1.0% | –1.1 | 7.5 | 3.8 |
| 1.0-4.9% | 1.7 | 4.3 | 5.0 |
| 5.0-9.9% | 10.4 | 4.3 | 9.5 |
| 10.0-41.0% | 19.4 | –0.3 | 12.4 |
| Status in CPS, 1989-1991 | | | |
| In CPS sample | 4.0 | 3.6 | 3.9 |
| In CPS, no poor children 5-17 | 3.6 | 6.0 | 3.1 |
| Not in CPS sample | 1.8 | 5.1 | 6.7 |

NOTE: See notes to Table D-4.
SOURCE: Data from the Bureau of the Census.

| Log Rate Under 21 | | Rate Under 21 | | Log Hybrid Rate-Number Under 21 | |
|---|---|---|---|---|---|
| D.4 | D.4a | D.5 | D.5a | D.6 | D.6a |
| 11.3 | 10.7 | 13.7 | 13.2 | 1.8 | 1.1 |
| 6.7 | 7.4 | 7.4 | 8.1 | 1.5 | 2.1 |
| 11.9 | 14.2 | 11.3 | 13.5 | 3.3 | 5.3 |
| 17.0 | 19.0 | 9.9 | 11.8 | 2.0 | 3.8 |
| 7.0 | 6.6 | 8.9 | 8.6 | 4.3 | 3.9 |
| 15.4 | 13.9 | 4.6 | 3.5 | 5.8 | 4.5 |
| 8.2 | 9.7 | 9.5 | 11.0 | –0.3 | 1.1 |

# References

Abelson, R.P., and J.W. Tukey
1963 Efficient utilization of nonnumerical information in quantitative analysis: General theory and the case of simple order. *Annals of Mathematical Statistics* 34:1347-1369.

Alexander, C.H., S. Dahl, and L. Weidman
1997 Making Estimates from the American Community Survey. Paper prepared for presentation at the Joint Statistical Meetings, Anaheim, Calif. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (August)

Bell, W.R.
1997a A County CPS Model with "Census Residuals". Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (July)
1997b Regression Diagnostics for Models for County Poverty Estimates—Complete. Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (September)

Belsley, D.A., E. Kuh, and R.E. Welsch
1980 *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Bureau of the Census
1987 *State Population and Household Estimates, with Age, Sex, and Components of Change: 1981-86*. Current Population Reports, Series P-25, No. 1010. Washington, D.C.: U.S. Department of Commerce.
1993 *Poverty in the United States: 1992*. Current Population Reports, Series P60-185. Washington, D.C.: U.S. Department of Commerce.
1995 *Subnational Estimates of Total Population by the Tax Return Method.* Population Estimates Branch, Population Division. Washington, D.C.: U.S. Department of Commerce.

Coder, J.F., R.C. Fisher, and P.M. Siegel
1996 Making Model-Based Estimates of the Number of Related Persons Age 5-17 in Poverty for all U.S. Counties, 1994. Housing and Household Economic Statistics Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (November)

Davis, S.
1994 *Evaluation of Post Censal County Estimates for the 1980's*. Population Division Working Paper Series, Number 5. Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Fay, R.E.
1996 Methodology for Poverty Estimation for 1993. Section 2. Demographic Statistical Methods Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (November)
1997a Comparisons of CPS and Census Poverty Levels. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (October)
1997b Notes on Direct Variance Estimates for Counties. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (September)

Fay, R.E., and Herriot, R.A.
1979 Estimates of income for small places: An empirical Bayes application of James-Stein procedures to census data. *Journal of the American Statistical Association* 78:269-277.

Fay, R.E., C.T. Nelson, and L. Litow
1993 Estimation of median income for 4-person families by state. Pp. 9-1 to 9-17 in *Indirect Estimators in Federal Programs*. Statistical Policy Working Paper 21, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget. Washington, D.C.: U.S. Government Printing Office.

Fay, R.E., and G.F. Train
1995 Aspects of survey and model-based postcensal estimation of income and poverty characteristics for states and counties. *American Statistical Association Proceedings of the Section on Government Statistics*. Alexandria, Va.: American Statistical Association.
1997 Small domain methodology for estimating income and poverty characteristics for states in 1993. *American Statistical Association Proceedings of the Social Statistics Section*. Alexandria, Vir.: American Statistical Association.

Fisher, R., and P. Siegel
1997 Methods Used for Small Area Poverty and Income Estimation. *American Statistical Association Proceedings of the Social Statistics Section*. Alexandria, Va.: American Statistical Association.

Ghosh, M., and J.N.K. Rao
1994 Small-area estimation: An appraisal. *Statistical Science* 8 (1):55-93.

Long, J.
1993 *Post Censal Population Estimates: States, Counties, and Places*. Population Division Working Paper Series, Number 3. Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Moskowitz, J., S. Stullich, and B. Deng
1993 *Targeting, Formula, and Resource Allocation Issues: Focusing Federal Funds Where the Needs are Greatest*. Washington, D.C.: U.S. Department of Education.

National Research Council
1995 *Measuring Poverty: A New Approach*. Constance F. Citro and Robert T. Michael, eds. Panel on Poverty and Family Assistance: Concepts, Information Needs, and Measurement Methods, Committee on National Statistics. Washington, D.C.: National Academy Press.
1997 *Small-Area Estimates of Children in Poverty, Interim Report 1, Evaluation of 1993 County Estimates for Title I Allocations*. Constance F. Citro, Michael L. Cohen, Graham Kalton, and Kirsten K. West, eds. Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics. Washington, D.C.: National Academy Press.

Otto, M.C., and W.R. Bell

1995 Sampling error modeling of poverty and income statistics for states. *American Statistical Association Proceedings of the Section on Government Statistics*. Alexandria, Va.: American Statistical Association.

1997 Bayesian Inference About Poverty and Income for States. Paper prepared for presentation at the Joint Statistical Meetings, Anaheim, Calif. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (August)

Platek, R., J.N.K. Rao, C.E. Särndal, and M.P. Singh, eds.

1987 *Small Area Statistics*. New York: John Wiley and Sons.

Prasad, N.G.N., and J.N.K. Rao

1990 The estimation of mean squared errors of small area estimators. *Journal of the American Statistical Association* 78:47-59.

Reed, J.M.

1996 Puerto Rico Estimates and Projections Based on Demographic Component Method: 1990-1995. International Programs Center, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (November)

Robinson, J.G., B. Ahmed, P. Das Gupta, and K. Woodrow

1993 Estimation of population coverage in the 1990 U.S. census based on demographic analysis. *Journal of the American Statistical Association* 88(423):1061-1079.

Siegel, P.

1997 Background on Poverty Estimates for School Districts [Tables]. Housing and Household Economic Statistics Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (May).

Sink, L.

1996 *Estimates of the Population of Counties by Age, Sex, Race, and Hispanic Origin: 1990-1994*. Release PE-48 (Methodology). Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Spencer, B.D., and Che-Fu Lee

1980 Postcensal population estimation methods of the Census Bureau. Pp. 131-187 in *Estimating Population and Income of Small Areas*. Panel on Small-Area Estimates of Population and Income, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.

Thiel, H.

1971 *Principles of Econometrics*. New York: John Wiley and Sons.

Voss, P.R., R. Gibson, and K. Morgen

1997 Assessment of the Initial 1993 Estimates of School Age Children in Poverty. Report prepared for 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.

# Biographical Sketches, Panel Members and Staff

GRAHAM KALTON (*Chair*) is a senior statistician and senior vice president of Westat. He is also a research professor in the Joint Program in Survey Methodology at the University of Maryland. Previously he was a research scientist in the Survey Research Center and a professor of biostatistics and statistics at the University of Michigan, professor of social statistics at the University of Southampton, and reader in social statistics at the London School of Economics. His research interests are in survey sampling and general survey methodology. He is a fellow of the American Statistical Association and of the American Association for the Advancement of Science. He has served as president of the International Association of Survey Statisticians and is president of the Washington Statistical Society. He is a past member of the Committee on National Statistics and has served as chair or a member of several of its panels. He received a B.Sc. in economics and an M.Sc. in statistics from the University of London and a Ph.D. in survey methodology from the University of Southampton.

DAVID M. BETSON is an associate professor of economics at the University of Notre Dame and a visiting scholar at the Joint Center for Poverty Research of the University of Chicago and Northwestern University. His previous positions have been as a research associate at the Institute for Research on Poverty at the University of Wisconsin and an economist in the Office of the Assistant Secretary for Planning and Evaluation in the U.S. Department of Health, Education, and Welfare. His research examines the effects of governments on the distribution of economic well-being with special reference to the measurement of poverty and the analysis of child support policy. He received a Ph.D. degree in economics from the University of Wisconsin–Madison.

CONSTANCE F. CITRO is a senior program officer for the Committee on National Statistics. She is a former vice president and deputy director of Mathematica Policy Research, Inc., and was an American Statistical Association/National Science Foundation research fellow at the Bureau of the Census. For the committee, she has served as study director for numerous panels, including the Panel on Poverty and Family Assistance, the Panel to Evaluate the Survey of Income and Program Participation, the Panel to Evaluate Microsimulation Models for Social Welfare Programs, and the Panel on Decennial Census Methodology. Her research has focused on the quality and accessibility of large, complex microdata files, as well as analysis related to income and poverty measurement. She is a fellow of the American Statistical Association. She received a B.A. degree from the University of Rochester and M.A. and Ph.D. degrees in political science from Yale University.

MICHAEL L. COHEN is a senior program officer for the Committee on National Statistics, currently serving as study director for the Panel on Statistical Methods for Testing and Evaluating Defense Systems. Previously, he was a mathematical statistician at the Energy Information Administration, an assistant professor in the School of Public Affairs at the University of Maryland, a research associate at the Committee on National Statistics, and a visiting lecturer at the Department of Statistics at Princeton University. His general area of research is the use of statistics in public policy, with particular interest in census undercount and model validation, and in robust estimation. He received a B.S. degree in mathematics from the University of Michigan and M.S. and Ph.D. degrees in statistics from Stanford University.

NANCY E. DUNTON is a principal social scientist at the Midwest Research Institute. Formerly, she was a senior research scientist with the New York State Department of Social Services and the New York State Council on Children and Families. Her work focuses on outcome indicators and social demography, with a special emphasis on children's policy issues. She received a Ph.D. degree in sociology from the University of Wisconsin–Madison.

WAYNE A. FULLER is a distinguished professor in the Department of Statistics and Economics at Iowa State University. He is a fellow of the American Statistical Association, the Econometric Society, and the Institute of Mathematical Statistics and is the author of *Introduction to Statistical Time Series* and *Measurement Error Models*. He also has an active research program in survey sampling. He has held offices in national and international statistical organizations and has previously served on National Research Council panels.

THOMAS B. JABINE is a statistical consultant who specializes in the areas of sampling, survey research methods and statistical policy. He was formerly a statistical policy expert for the Energy Information Administration, chief mathematical statistician for the Social Security Administration, and chief of the Statistical Research Division of the Bureau of the Census. He is a fellow of the American Statistical Association and a member of the International Statistical Institute. He has a B.S. degree in mathematics and an M.S. degree in economics and science from the Massachusetts Institute of Technology.

SYLVIA T. JOHNSON is professor of research methodology and statistics in education at Howard University, where she is also editor-in-chief of the *Journal of Negro Education*. She has served on the faculties of Roosevelt, Trenton (NJ) State, and Western Illinois universities and Augustana and Chicago City Colleges and as a visiting scholar at the Educational Testing Service. She is currently a principal investigator at the Center for Research on the Education of Students Placed at-Risk (CRESPAR), a joint activity of Howard University and Johns Hopkins University, and of the Design and Analysis Committee of the National Assessment of Educational Progress (NAEP). Formerly, she was a member of the Technical Advisory Committee for the National Adult Literacy Survey and the Graduate Record Examination and served as a trustee of the College Board. She is a fellow of the American Psychological Association and chairs its Division 15 Committee on Members and Fellows. Dr. Johnson received a Ph.D. degree in educational measurement and statistics from the University of Iowa.

THOMAS A. LOUIS is professor and head of the Division of Biostatistics at the University of Minnesota School of Public Health and professor of statistics in the School of Statistics. His research interests include Bayes and empirical Bayes methods, research synthesis, risk assessment, analysis of longitudinal and spatial data, and sequential design of experiments. He is codirector of the statistical center for Community Programs for Clinical Research on AIDS and for 6 years directed the statistical center for the school's Cancer Prevention Research Unit. He is an associate editor of *Statistical Science* and on the editorial board of *Statistical Neerlandica*. He is a member of the International Statistics Institute, a fellow of the American Statistical Association and the American Association for the Advancement of Science, and a trustee of the National Institute for Statistical Sciences. He received a B.A. from Dartmouth College and a Ph.D. in mathematical statistics from Columbia University.

SALLY C. MORTON is head of the Statistics Group at RAND in Santa Monica, California. She is on the faculty of the RAND Graduate School of Public Policy Studies and is a lecturer in the School of Public Health at the University of California–Los Angeles. She serves as an associate editor for the *Journal of the*

*American Statistical Association* and *Statistical Science*, is chair of the association's Section on Statistical Graphics, and is a member of the Caucus for Women in Statistics. Her health policy research concerns homelessness, severe mental illness, and outcomes research and quality of care in the areas of childbirth and AIDS. Her methodological research concentrates on meta-analysis, nonparametric regression, and the sampling of vulnerable populations. She received a Ph.D. in statistics from Stanford University.

JEFFREY S. PASSEL is a principal research associate at the Urban Institute and the director of the institute's Program for Research on Immigration Policy. Previously, he was assistant division chief for estimates and projections in the Population Division of the Bureau of the Census, and he also directed the research on demographic methods for measuring census undercount. His research interests include the demography of immigration, particularly the measurement of illegal immigration; the effects and integration of immigrants into American society; and measuring and defining racial and ethnic groups in the United States. He is a member of a number of professional societies and has served in various capacities in the Population Association of America, the American Statistical Association, and the American Association for the Advancement of Science. He is a fellow of the American Association for the Advancement of Science. He received a B.S. in mathematics from the Massachusetts Institute of Technology, an M.A. in sociology from the University of Texas at Austin, and a Ph.D. in social relations from the Johns Hopkins University.

J.N.K. RAO is professor of statistics at Carleton University in Ottawa, Canada, and a consultant to Statistics Canada. Formerly, he was a professor at the University of Manitoba and Texas A&M University. His research interests include survey sampling theory and methods, particularly small area estimation. He is a fellow of the Royal Society of Canada, the American Statistical Association, and the Institute of Mathematical Statistics. He received a Ph.D. degree in statistics from Iowa State University.

ALLEN L. SCHIRM is a senior researcher at Mathematica Policy Research, Inc. Formerly, he was Andrew W. Mellon assistant research scientist and assistant professor at the University of Michigan. His principal research interests include small-area estimation and sample and evaluation design, with application to welfare, food and nutrition, and education policy. He is an associate editor of *Evaluation Review* and a member of the American Statistical Association, the American Economic Association, and the Population Association of America. He received an A.B. in statistics from Princeton University and a Ph.D. in economics from the University of Pennsylvania.

PAUL R. VOSS is professor of rural sociology and chair of the Department of Rural Sociology at the University of Wisconsin–Madison. For the past 21 years he has been affiliated with the Wisconsin Applied Population Laboratory and is currently its director. He also is affiliated with the Wisconsin Institute for Research on Poverty. His research involves modeling small-area population change for purposes of population estimation and projection, and he also has studied and written about the demographic composition of small-area migration streams. He is a member of the U.S. Department of Commerce's Advisory Committee for the 2000 Census as well as the Census Bureau's Advisory Committee of Professional Associations. He received a Ph.D. degree in sociology (demography) from the University of Michigan.

JAMES H. WYCKOFF is associate professor of public administration and policy at the University at Albany, of the State University of New York. His research involves applied public economics and public policy, with particular focus on the economics of education. He was an American Statistical Association Fellow at the Census Bureau. He received a Ph.D. degree from the University of North Carolina at Chapel Hill.

ALAN M. ZASLAVSKY is associate professor of statistics in the Department of Health Care Policy at Harvard Medical School. He was formerly on the faculty of the Department of Statistics at Harvard. His research interests include measurement of quality in health care, census methodology, estimation and correction of census undercount, small-area estimation, microsimulation, design and analysis of surveys, and Bayesian methods. He has served on two other panels of the Committee on National Statistics concerned with planning for the 2000 census and is a fellow of the American Statistical Association. He received a Ph.D. degree in applied mathematics from the Massachusetts Institute of Technology.

MEYER ZITTER is an independent demographic consultant. Formerly, he was chief of the Census Bureau's Population Division and also served as assistant director for international programs. He is a fellow of the American Statistical Association and a member of the International Statistical Institute and the International Union for the Scientific Study of Population. He has a B.B.A. degree from City College of New York.