SOME RECENT APPLICATIONS OF MODELS TO PROBLEMS OF ESTIMATION AND INFERENCE FROM CENSUS BUREAU SURVEYS

Robert E. Fay III, U.S. Bureau of the Census

This paper describes two applications of models to data from Census Bureau surveys. In one, a modified James-Stein estimator incorporating a linear regression model has been applied to sample data from the 1970 census. The resulting estimates have been employed as base figures for small areas in the Census Bureau's program of estimation for purposes of General Revenue Sharing. (The title of this paper treats the 1970 census 20-percent sample as a survey.) The other application, a linear regression model to estimate the current proportion of children in poverty families by State, represents an important component in the Census Bureau's evaluation of the sample estimates of this characteristic from the 1976 Survey of Income and Education. The application of the latter model, therefore, falls broadly into the category of inference.

One intent of this paper is to illustrate possible directions for future applications of models for purposes of both estimation and inference from survey data. The two applications here may be viewed as approaches to problems lying beyond the scope of more standard survey practice. Thus, each represents an attempt to extend the utility of survey data in a manner that may be considered generally consistent with existing objectives and practice at the Census Bureau.

A second aim of this paper is to raise questions related to the foundations of estimation, and particularly of inference, from sample surveys. In this second respect, this paper footnotes "Statistical Theory and Practice at the U. S. Bureau of the Census," of Nisselson and Isaki (1977). The two authors observe that models and other prior information are employed extensively in the design of Census Bureau surveys but emphasize that the class of survey designs is generally restricted to probability sampling. Nisselson and Isaki also imply that models play essentially no role in estimation or inference from surveys at the Census Bureau. While finding no essential argument with these conclusions, this paper will suggest that the link between demographic survey data and the practice of estimation and inference is almost always, although to varying degrees, mediated by models.

The two applications to be discussed in detail here fail to serve as typical examples of this general application of models in survey practice; rather, they represent extreme cases, illustrating the thesis by exaggeration. Other examples will be offered to indicate the sense in which models play a key role in estimation and inference at the Census Bureau.

Finally, at the risk of some ambiguity, the term "model" will apply to two different concepts in this paper. In a general sense, "model" may signify the representation of a process of scientific inquiry, so that one may speak of the theory of sampling from finite populations as a model for the practice of Census Bureau surveys. The more specific sense in this paper will be to indicate a restriction imposed upon an unknown finite population beyond the restrictions dictated by logic alone. For example, a linear model expressing a relationship between a population characteristic and other measured quantities may be considered a restriction upon the class of possible populations through the imposed linear relationship.

The two applications to be considered here involve models in this second sense, and they will be employed to illustrate that estimation and inference from sample surveys are also typically based upon models that restrict the structure of the unknown population to some extent.

Estimation

Estimation from demographic surveys at the Census Bureau is almost always accomplished through summation of weights. (A significant exception occurs with the composite estimator in the Current Population Survey.) From the perspective of the theory of sampling from finite populations, the weights reflect both the inverse of the probability of selection (the so-called "unbiased" estimate) and stages of ratio estimation. Again, from this perspective, ratio estimation may be construed as an often effective technique to reduce variance at the cost of negligible bias.

In practice, multiple steps of ratio estimation are employed to accomplish far more than the reduction of variance. Adjustments to compensate for noninterviews are followed by adjustments to bring the survey estimates up to independently-derived national estimates by age, race, and sex. Independent estimates for States may also be introduced, as well as rules to force logical consistencies, such as the number of husbands living with their wives matching the number of wives living with their husbands. Clearly, this extended process of estimation is based upon what might be called a model of coverage for the survey, since the process can produce consistent estimates of characteristics such as unemployment or poverty only under certain restrictions of the population. In other words, the weighting procedures specify a series of steps leading from the raw survey data to a set of estimates, but the operations could be equated, at least approximately, to a set of assumed structural relationships between the expected frequencies in the survey data and the population.

The coverage model reflected by the preceding approach has known deficiencies. Contradictory results spring from the variation in weights among family members, and persistently discrepant patterns fall outside of sampling error. A considerable amount of research, both within and outside of the bureau, has been directed toward the development of alternative models of survey coverage. (The paper by Yuskavage, Hirschberg, and Scheuren (1978) provides a thoughful discussion of this problem, as well as references to other work.)

If one adds to the problem of weighting the rules for editing, and particularly the procedures for imputation, it is clear that an extensive array of implied models links the collected survey data to the final estimates. These links have been acknowledged many times before, but their importance seems to be often omitted from discussions of the foundations of survey estimation. This omission, I believe, leads to incomplete consideration of the foundations of actual survey practice.

The application of modeling to be presented here differs considerably from the general models just alluded to. As required by law, the Census Bureau transmits to the Department of the Treasury current estimates of per capita income for the approximately 39,500 units of local government participating in the General Revenue Sharing Program. In short, the current estimates represent an updating of base year figures for income year 1969. For all larger places, 1970 census sample estimates provide the base year figure, whereas for places of population under 500 persons, the respective county values of per capita income were employed as base figures for 1969 in forming the set of estimates for 1972. Allowing county figures to stand for these small places represents a model of the simplest sort. Through the use of other models, it subsequently became possible to replace the substituted county figures for small places with a James-Stein estimator incorporating the census sample estimates, the county values, and other auxiliary information. This application has been presented elsewhere (Fay and Herriot, 1977; Fay, 1978b), but a description will be given here to emphasize the extent to which models form an integral part of this estimation.

The situation was well-suited for modeling: sample estimates of varying reliability were available for the places in question, and there was a presumed but unassessed relationship between the true per capita income for a given place and that for its county. In addition, auxiliary information was available for many of the places in question: the value of owner-occupied housing was collected in the 1970 census on a complete-count basis for all non-farm units; and the Internal Revenue Service tax return file provided for most places a value of the adjusted gross income per exemption. In short, the strategy selected for the estimation related the auxiliary information to the sample values of per capita income; differences between the sample estimates and the derived regression equation were assessed relative to the sampling error of the census values in order to derive a measure of average fit; and this measure of fit was used to determine a weighting of the sample estimates and the original regression or model estimates. The James-Stein estimator formed the basis of this strategy. (Papers of Efron and Morris (1972, 1973, and 1975) provide most of the theoretical basis for the following analysis.)

Although it is possible to present the James-Stein estimator in a manner that centers upon its frequentist properties, its description as an empirical Bayes estimator more fully highlights the modeling aspect of the estimator. Suppose that Y_i are given sample estimates of θ_i with variance (conditional upon θ_i) D_i. The Bayesian formulation treats θ_i as a random variable; in this application θ_i was modeled as a linear combination of auxiliary variables X_{ij} plus random error. In matrix notation, the model became

$$\theta \sim N(X\beta, AI) \qquad (1)$$

$$Y|\theta \sim N(\theta, D) \qquad (2)$$

with A representing the underlying super-population variance of θ about the predicted values $X\beta$, and D taken to be the diagonal matrix of the sampling variances D_i. The technique developed to estimate the unknown parameters β and A involved the joint solution of a weighted linear regression to determine $\hat{\beta}$

$$\hat{\beta} = (X^T(D + \hat{A}I)^{-1}X)^{-1} X^T(D + \hat{A}I)^{-1}Y \qquad (3)$$

and adaptation of $\hat{A} \geq 0$ to satisfy the relationship among the residuals of the regression

$$n\text{-}k = (Y - X\hat{\beta})^T(D + \hat{A}I)^{-1}(Y - X\hat{\beta}) \qquad (4)$$

for n-k as the residual degrees of freedom of the regression. Equations (3) and (4) represent an extension of the original James-Stein estimator to this problem. The original weighted combination of the regression and sample values

$$\delta = X + \hat{A}(D + \hat{A}I)^{-1}(Y - X\hat{\beta}) \qquad (5)$$

was restricted to lie within one standard error of the sample estimates by

$$\delta_i' = Y_i + \sqrt{D_i} \quad \text{if } \delta_i > Y_i + \sqrt{D_i} \qquad (6)$$

$$= Y_i - \sqrt{D_i} \quad \text{if } \delta_i < Y_i - \sqrt{D_i} \qquad (7)$$

$$= \delta_i \quad \text{otherwise} \qquad (8)$$

This last modification prevents any model-based estimate from lying excessively far from the sample estimate, and limits the expected error that occurs to any particular place.

In this application, Y_i was taken to be the sample (natural) logarithm of per capita income and θ_i the (natural) logarithm of the true per capita income. D_i was given approximately as 9.0 times the inverse of the sample estimate of total persons; for a place of 100 persons, $D_i = .09$, equivalent to a coefficient of variation of about 30 percent for the sample estimate of per capita income ($.09 = .30^2$). (The variance of a natural logarithm can be equated approximately to the square of the coefficient of variation.) Similarly, all independent variables were expressed in logarithmic form. Reduced equations were fitted for places without acceptable values for value of housing or IRS adjusted gross income per exemption based on all places with adequate data for that equation.

Table 1 shows the values of $\hat{A}$ obtained for a number of States. When county values alone are used in the regression, the values of A tend to exceed .04, which is equivalent to an intrinsic error of prediction of 20 percent ($.04 = .20^2$). These results imply that in spite of the advantage of the county values in terms of sampling error, (5) weights the sample estimates more heavily than the county estimates down to a population of about 225, which corresponds to $D_i = .04$ also. The other columns of table 1 show that the auxiliary information achieves a significant reduction in the average error over the use of the county values alone.

This application may be seen to involve two sorts of models. The James-Stein estimator itself may be motivated through an appeal to an infinite population underlying the given finite population. Furthermore, the use of regression techniques to exploit basic relationships with auxiliary data represented a restriction or model of the infinite population. The modeling approach constituted an effective way to reduce the average risk, and its use could be justified in this instance by the large number of estimates required and the inability to fulfill the requirements through sample data alone on a timely basis. (The 1980 census tentatively will include a 50-percent sample to collect income and other sample statistics for small areas.)

Inference

The question of inference has held its place among the most controversial in all of statistics. Bayesian, frequentist, and fiducial schools of inference, with sometimes substantial differences within each interpretation, have offered us conflicting solutions. The respective limitations of these theories of inference have been recited often before by their critics. In spite of the extensive literature on these questions, aspects remain to be fully explored, as Kiefer's (1977) recent work has reminded us.

Relatively recently, these same questions have been raised with respect to the particular problems of survey sampling. Many of the resulting debates have recognized their parallels to the broader area of general statistics, while others have seemingly proceeded without this awareness.

The Census Bureau has essentially endorsed the Neyman-Wald frequentist approach to inference from its surveys. (U. S. Bureau of the Census, 1974.) In practice, however, inference has actually occupied a decidedly secondary position relative to estimation. For example, practically all variances published by the Census Bureau for demographic surveys are based upon models. The variances given may have been derived in designing the survey or represent generalizations of variance estimates from the survey data for a selection of items; although assumptions about components of variance rather than complete characterizations of the basic population are involved, they represent a restriction or model of the population nonetheless. Without explicit recognition of the underlying models in an inferential formalism, the confidence intervals constructed from the published estimates are potentially too short for possible finite populations. The frequentist interpretations given these intervals therefore assume more than is stated.

The example to be discussed here illustrates the potential and actual importance of models in survey inference, although in an unusual manner. The model in this case both enabled statements to be made about a set of survey estimates, thus serving the purpose of inference, while simultaneously leading to results that question the adequacy of the foundations of standard survey inference to address this situation.

The Congress, under Title I of the Elementary and Secondary Education Act of 1965, has allocated funds, currently about $2 billion annually, to school districts through use of a formula that includes the number of school-age children living in poverty families by county. To date, the most recent statistics on poor children in the allocation formula have come from the 1970 census. In the Educational Amendments of 1974, however, Congress mandated both a sample survey adequate to produce current State estimates of children living in poverty families (later to become the 1976 Survey of Income and Education (SIE)) and also a report evaluating the survey estimates (submitted as "Assessment of the Accuracy of the Survey of Income and Education"). The principal effort in the evaluation centered upon a reinterview of approximately 6,000 interviewed households in the SIE and 2,000 in the CPS (Current Population Survey). The reinterview attempted to create a standard for comparison through use of more intensive interviewing techniques than the original survey. (The reinterview has been described in U.S. Bureau of the Census, 1978; and in Fay, 1978a.)

The reinterview data supported the basic accuracy of the SIE estimates of the number of children in poverty families by State on important points. The returns from the SIE had seemed initially to require explanation in one important regard: the SIE national estimate of children in poverty families was about 12 percent below the comparable figure for the CPS. The reinterview data for both the SIE and CPS affirmed the SIE national result, however, implying that more intensive interviewing procedures would not have substantially altered the SIE national result. The SIE reinterview data also detected no statistically significant bias in the SIE estimates at the level of Census region or division.

Other questions remained, however, which could only be addressed adequately through appeal to a model. The model providing these answers was originally developed in 1975 by Gordon Green and myself (the model is described in the report to Congress, Fay 1978a and 1978b). Incorporating suggestions from previous work of Ericksen (1972, 1973) and others, a regression model was developed to estimate the proportion of poor children by State. The model employs sample estimates of the current proportions as the dependent variable. The census proportions by State constitute one independent variable, while two variables are formed from BEA estimates of current per capita personal income by State by finding the median PCI_m of the 51 State (and D.C.) figures and computing

$$X_{i2} = \ln (PCI_i/PCI_m) \quad \text{if } PCI_i > PCI_m \qquad (9)$$

$$= 0 \quad \text{otherwise} \qquad (10)$$

$$X_{i3} = 0 \quad \text{if } PCI_i > PCI_m \qquad (11)$$

$$= \ln (PCI_i/PCI_m) \quad \text{otherwise} \qquad (12)$$

Similarly, two additional independent variables are formed from the values of BEA income in 1969. (A typical outcome in fitting the model is to obtain coefficients that are negative for the current income year and positive for the census year, with the implication that States experiencing a more rapid than average increase in income are likely to have a concomitant decline in poverty.)

The model was developed in 1975 through attempts to predict the 1970 census proportions of families in poverty on the basis of BEA data and the corresponding proportions from the 1960 census. Although SIE and CPS data have since agreed relatively favorably with subsequent results, a full, independent evaluation of the model will have to await the findings of the 1980 census.

The model has already served to illuminate important aspects of the SIE data, however. For example, a concern was initially expressed regarding the comparability of the SIE and census procedures; the difference in national levels between the SIE and CPS is compounded by a difference between the CPS and census in 1970, and, altogether, it may be argued that there is an intrinsic difference of around 20 percent between the national level measured by the census and SIE procedures. Some (e.g., Ginsberg and Grob (1978)) suggested initially this substantial difference might also imply a potential inconsistency between the SIE and census in the measurement of the distribution of poverty among States, the characteristic critically important for the allocation.

The regression estimates derived from the CPS data provide the most direct evidence on this question, since they link 1970 to 1976 by an annual series obtained from a fixed methodology. Figures 1 to 4 were presented in the report to Congress in support of the supposition that the changes in the distribution of poverty among States since 1970 as indicated by the SIE reflected predominately an actual change in the true distribution and not consequences of procedural changes.

The regression was also fitted directly to the SIE State estimates, giving the results in table 2. The average relative difference between the two sets of estimates for 1975 is 14 percent (root mean square), somewhat beyond the average sampling error of 10 percent for the SIE estimates.

The comparison of the reinterview and the regression estimates of table 2 yields a remarkable outcome. When States are classified by the directions of difference for the model and the reinterview from the SIE estimate, table 3a results. There is a suggestion in this first table of a statistical tendency for the reinterview and model to differ in the same way: for 30 States there is agreement versus disagreement for 20. A relation between the two is difficult to rationalize as a statistical artifact: the reinterview results represent differences for a matched subsample of the SIE sample and are therefore unrelated in principle to any unrepresentativeness of the SIE sample itself. For purposes of further analysis, a covariance adjustment was applied to the SIE-to-model comparisons using aggregate data on AFDC income. The basic nature of the adjustment was to reduce the SIE to model ratios for States with comparatively high reported AFDC income relative to aggregate administrative controls, because these States probably had an over-representation of poor families in the SIE sample. Conversely, the adjustment moved the SIE to model ratio in the opposite direction for States with low reported AFDC income relative to the independent controls. (A more detailed description of the rationale and mechanics of this covariance adjustment is given in the report to Congress.) Table 3b compares the differences between reinterview and original results, still based on a matched subsample of the original sample, and the adjusted relationships between the model and SIE estimates. Here, the evidence of association is incontrovertible.

A parametric interpretation of the reinterview data and model results implied the existence of a component of nonsampling error in the SIE State estimates sufficient to increase the average sampling error of 10 percent to a total average mean square error of about 12 percent. The result is far from precise, however, since the sampling error in the reinterview estimates leads to a 95 percent confidence interval for the total error ranging from just above 10 percent to 14 percent.

What inferences can be made about the SIE State estimates of children in poverty families? The report to Congress stressed the essential fact central to the evaluation: "The limitations of the survey estimates, both in terms of sampling reliability and other possible survey errors, are found to be small relative to the changes in poverty since the 1970 census. The SIE estimates, therefore, more accurately reflect the current distribution of poverty among States than the 1970 census values." Precision beyond this is difficult. For example, what frequentist interpretation can be given an interval two standard deviations around a State estimate? Or for that matter, a similar interval based upon an imprecisely estimated total error instead? Meaningful answers to these questions will require further advances in the theory of survey inference, and, I surmise, models will occupy a decidedly key role in the solution.

References

Ericksen, Eugene P. (1973), "A Method of Combining Sample Survey Data and Symptomatic Indicators to Obtain Population Estimates for Local Areas," Demography, 10, 137-60.

__________ (1974), "A Regression Method for Estimating Population Changes for Local Areas," Journal of the American Statistical Association, 69, 867-75.

Efron, Bradley, and Morris, Carl (1972), "Limiting the Risk of Bayes and Empirical Bayes Estimators -- Part II: The Empirical Bayes Case," Journal of the American Statistical Association, 67, 130-9.

__________, and Morris, Carl (1973), "Stein's Estimation Rule and Its Competitors -- an Empirical Bayes Approach," Journal of the American Statistical Association, 68, 117-30.

__________, and Morris, Carl (1975), "Data Analysis Using Stein's Estimator and Its Generalizations," Journal of the American Statistical Association, 70, 311-9.

Fay, Robert E., and Herriot, Roger (1977), "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," submitted to the Journal of the American Statistical Association.

__________, (1978a), "Problems of Nonsampling Error in the Survey of Income and Education: Content Analysis," Proceedings of the Social Statistics Section, 1977, American Statistical Association, Washington, D.C., 418-21.

__________, (1978b), "Some Recent Census Bureau Applications of Regression Techniques to Estimation," presented at the NIDA/NCHS Workshop on Synthetic Estimates, Princeton, N.J., April 13-14, 1978.

Ginsberg, Alan, and Grob, George (1978), "Uses of Data from the Survey of Income and Education," Proceeedings of the Social Statistics Section, 1977, American Statistical Association, Washington, D.C., 409-13.

Kiefer, J. (1977), "Conditional Confidence Statements and Confidence Estimators," Journal of the American Statistical Association, 72, 789-808.

Nisselson, Harold, and Isaki, Cary T. (1977), "Statistical Theory and Practice at the U.S. Bureau of the Census," presented at the 41st Session of the International Statistical Institute, New Delhi, India, December 5-15, 1977.

U.S. Bureau of the Census (1974), Standards for Discussion and Presentation of Errors in Data, Technical Paper No. 32, U.S. Government Printing Office, Washington, D.C.

__________ (1978), "Assessment of the Accuracy of the Survey of Income and Education," submitted to Congress on April 25, 1978.

Yuskavage, Robert; Hirschberg, David; and Scheuren, Frederick J. (1978), "The Impact on Personal and Family Income of Adjusting the Current Population Survey for Undercoverage," Proceedings of the Social Statistics Section, 1977, American Statistical Association, Washington, D.C., 70-80.

Table 1. Estimated $\hat{A}$ for Places with 20-Percent Sample Estimate of Population Less Than 500

States	Regression Equation			
	County	County and Tax	County and Housing	County, Tax, and Housing
a. States with More Than 500 Places in Class				
Illinois	.036	.032	.019	.017
Iowa	.029	.011	.017	.000
Kansas	.064	.048	.016	.020
Minnesota	.063	.055	.014	.019
Missouri	.061	.033	.034	.017
Nebraska	.065	.041	.019	.000
North Dakota	.072	.081	.020	.004
South Dakota	.138	.138	.014	-
Wisconsin	.042	.025	.025	.004
b. States with 200-500 Places in Class				
Arkansas	.074	.036	.039	.018
Georgia	.056	.081	.067	.114
Indiana	.040	.012	.003	.000
Maine	.052	.015	-	-
Michigan	.040	.032	.028	.023
Ohio	.034	.015	.004	.004
Oklahoma	.063	.027	.049	.036
Pennsylvania	.020	.018	.016	.011
Texas	.092	.048	.056	.040

Note: A dash (-) indicates that the regression was not fitted because of too few observations.

Fig. 1 Model Estimates Based on CPS of the Percent of Total Poor Children in the Northeast Region, by Income Year and Division (1970 Census and 1976 SIE Estimates Shown as Points)

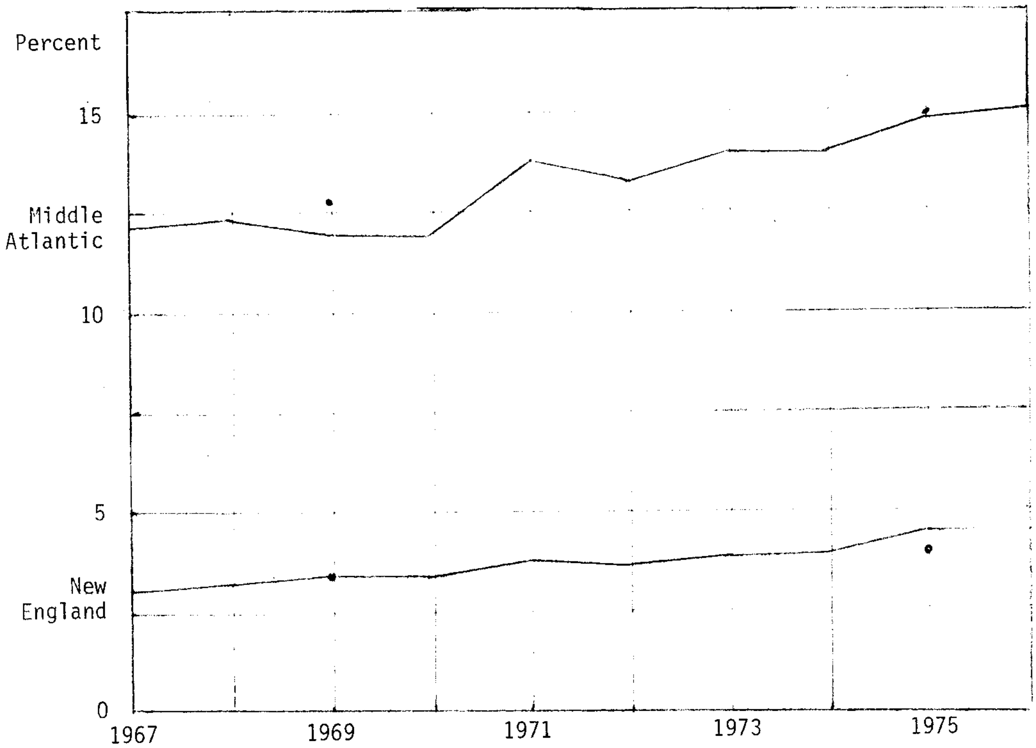

Fig. 2 Model Estimates Based on CPS of the Percent of Total Poor Children in the North Central Region, by Income Year and Division (1970 Census and 1976 SIE Estimates Shown as Points)

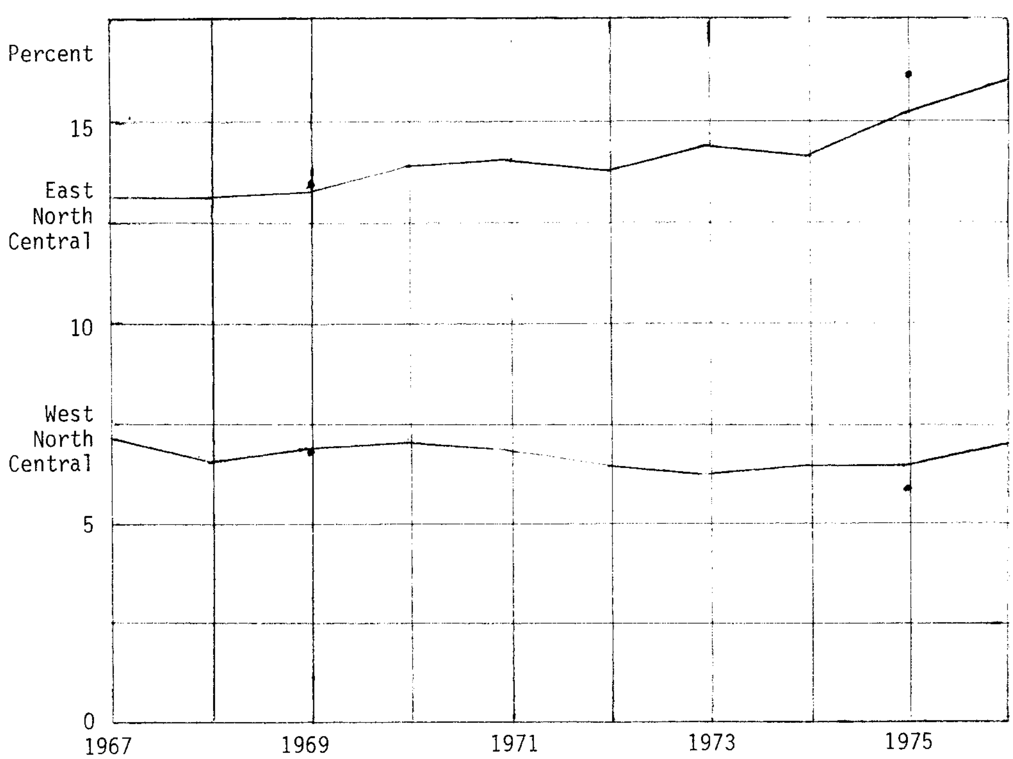

Fig. 3 Model Estimates Based on CPS of the Percent of Total Poor Children in the South Region, By Income Year and Division (1970 Census and 1976 SIE Estimates Shown as Points)

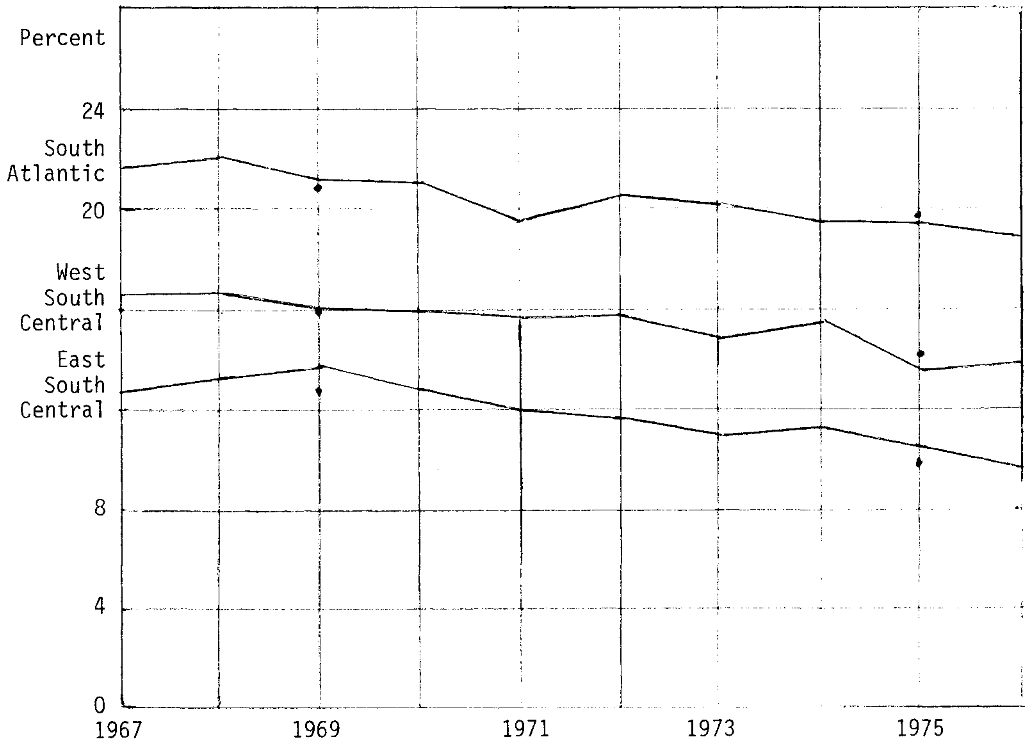

Fig. 4 Model Estimates Based on CPS of the Percent of Total Poor Children in the West Region, by Income Year and Division (1970 Census and 1976 SIE Estimates Shown as Points)

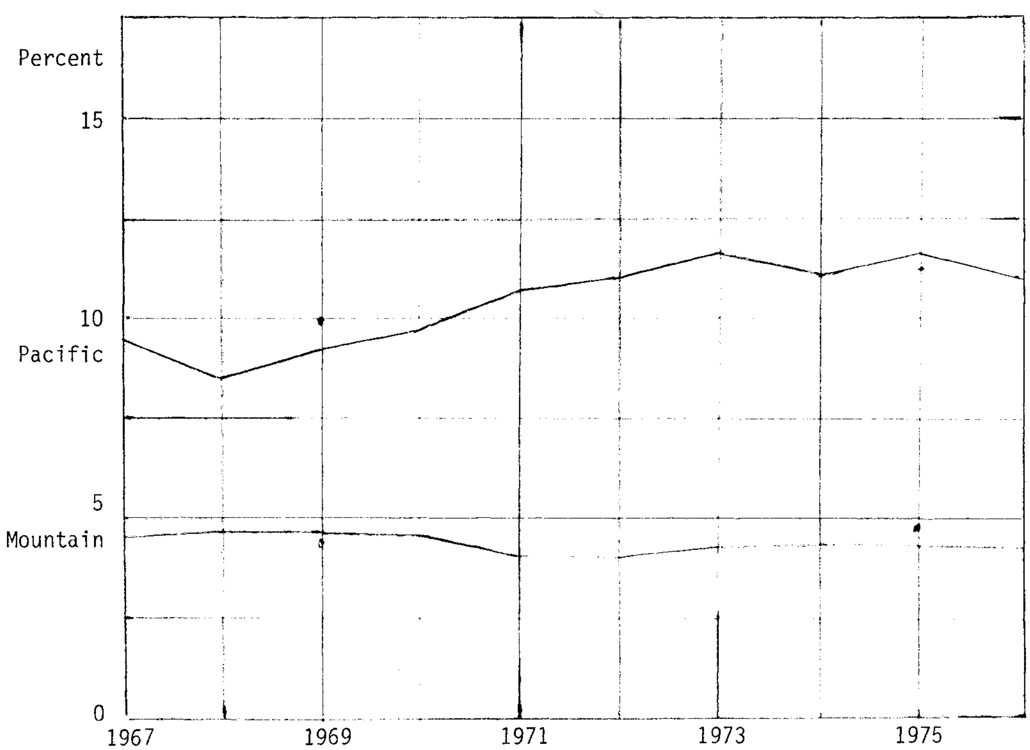

Table 2. Percent of Children 5-17 Years Old in Poverty Families by State, According to the 1970 Census, 1976 SIE, and Regression Model Fitted to the SIE Estimates

States by Region	1969 Estimates	1975 Estimates	
	1970 Census	1976 SIE	Regression Model Fitted to SIE
Northeast			
Maine	14.2	15.3	14.2
New Hampshire	7.7	10.3	10.5
Vermont	11.4	17.8	11.9
Massachusetts	8.4	9.3	10.6
Rhode Island	11.0	10.5	11.8
Connecticut	7.2	8.4	9.6
New York	12.2	13.1	13.8
New Jersey	8.7	11.6	10.2
Pennsylvania	10.6	12.6	10.9
North Central			
Ohio	9.8	11.6	11.8
Indiana	9.0	9.6	10.8
Illinois	10.7	15.1	10.8
Michigan	9.1	11.3	11.2
Wisconsin	8.7	9.4	9.6
Minnesota	9.5	9.1	9.7
Iowa	9.8	7.9	8.2
Missouri	14.8	14.7	14.8
North Dakota	15.7	11.5	10.4
South Dakota	18.3	13.1	15.3
Nebraska	12.0	10.1	10.3
Kansas	11.5	8.6	10.2
South			
Delaware	12.0	10.4	12.3
Maryland	11.5	10.7	11.2
District of Columbia	23.2	15.7	17.8
Virginia	18.2	13.7	15.0
West Virginia	24.3	18.9	18.2
North Carolina	24.0	17.8	20.2
South Carolina	29.1	23.9	23.4
Georgia	24.4	21.3	20.9
Florida	18.9	21.6	16.6
Kentucky	25.1	21.4	20.2
Tennessee	24.8	20.5	20.2
Alabama	29.5	15.9	23.1
Mississippi	41.5	32.6	32.2
Arkansas	31.6	21.4	23.8
Louisiana	30.1	22.9	23.8
Oklahoma	19.5	14.6	16.2
Texas	21.5	20.5	17.7
West			
Montana	12.9	12.5	10.8
Idaho	12.0	11.0	10.5
Wyoming	11.2	8.6	8.2
Colorado	12.3	10.7	10.7
New Mexico	26.3	26.0	21.2
Arizona	17.5	16.8	16.1
Utah	10.0	8.0	9.4
Nevada	8.8	11.0	9.8
Washington	9.3	10.0	10.2
Oregon	10.3	8.4	10.2
California	12.1	13.8	12.5
Alaska	14.6	6.4	6.9
Hawaii	9.7	9.6	9.8

Table 3a. Comparison of Reinterview, Regression Model Fitted to SIE, and SIE Estimates of Children 5-17 Years Old in Poverty Families by State

(See Text for Explanation)

	Comparison of Model to SIE	
Comparison of Reinterview to SIE	States with Model Estimate Less than SIE	States with Model Estimate Greater than SIE
States with re-interview less than SIE	12	10
States with re-interview greater than SIE	10	18

Note: One State is omitted because of an estimate of no change in reinterview.

Table 3b. Comparison of Reinterview, Regression Model Fitted to SIE, and Adjusted SIE Estimates of Children 5-17 Years Old in Poverty Families by State

(See Text for Explanation)

	Comparison of Model to Adjusted SIE	
Comparison of Reinterview to SIE	States with Model Estimate Less than Adjusted SIE	States with Model Estimate Greater than Adjusted SIE
States with re-interview less than SIE	15	7
States with re-interview greater than SIE	8	19

Note: Two States are omitted: one with an estimate of no change in reinterview, and the other with an estimate of no difference (within 0.5 percent) between the model and adjusted SIE estimates.