

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Annals of Economic and Social Measurement, Volume 2, number 2

Volume Author/Editor: NBER

Volume Publisher: NBER

Volume URL: <http://www.nber.org/books/aesm73-2>

Publication Date: April 1973

Chapter Title: The CPS Viewed from the Outside

Chapter Author: Sanford V. Berg

Chapter URL: <http://www.nber.org/chapters/c9884>

Chapter pages in book: (p. 99 - 104)

THE CPS VIEWED FROM THE OUTSIDE

BY SANFORD V. BERG

This paper summarizes the presentations at the September 1972 Conference on the Current Population Survey and evaluates the potential of this data base for research.

The expected March 1973 release of the public use CPS file on tape is a significant event for those who have only had access to cross-tabulations of the monthly survey results. The introduction of punchcard technology had facilitated aggregation, making tabular arrays of published information more easy to construct. And now, the analysis of disaggregated information has become possible with larger computer memories, improved input-output devices, and the availability of microdata. So it was appropriate that the NBER sponsor a conference to review the objectives and procedures of the CPS and bring together the producers and users of the data. The papers in this issue of the *Annals*, which were presented at the conference, provide an introduction to the CPS. But more importantly, they extend statistical methodology related to survey data and report on work in progress on the use of previously unavailable CPS files.

The theme paper for the conference was prepared by Marvin M. Thompson and Gary Shapiro, and delivered by Conrad Taeuber, Associate Director for Demographic Fields, Census Bureau. This overview explains the operations involved in sampling, data collection, and processing for the CPS. After briefly describing the history and content of the CPS, the authors explore sample design and rotation and some topics not dealt with in Technical Paper Number 7, *The Current Population Survey—A Report on Methodology*, 1963. Lenore Bixby, Director of the Division of Retirement and Survivor Studies of the Social Security Administration, stressed in her discussion that the CPS "... is probably the most scientifically and 'correctly' selected national sample extant. ... Also, based on Dr. Taeuber's description of the training, supervision, and quality control techniques applied to interviewers, those procedures certainly appear to be the best possible under constraints imposed by time and money."

Bixby went on to raise a fundamental question about the sample design relative to the types of population groups which one may wish to study. She noted that in the past, we were concerned with statistics about the aggregate, and were interested in central tendencies. But now the policy arena recognizes several universes, and we are interested in the relative status of segments of the population: black-white, male-female, and urban-rural, for example. Perhaps only an increase in the sample size can meet the new research needs. Yet there remains the matter of priorities facing an agency which generates data files only as a by-product: periodic reviews of variables and survey experimentation require resources.

The open discussion was perhaps the most lively of the entire Conference, moving from technical problems to more fundamental issues. Examples of the former were questions regarding specific non-response rates and the basis for survey stratification. With respect to the latter, it was suggested that if income

variables were unsatisfactory, why not expand the Internal Revenue Service forms to make more effective use of the tax return, at least on a sample basis. Certainly, if a negative income tax is passed, most of the population will be covered.

Some of the flavor of the discussion is captured in James Smith's article on the content of the CPS. He raises a number of substantive issues regarding the kinds of questions asked and the types of policy needs which might be met by altering or reorganizing some questions. In the discussion, some concern was expressed regarding the use of CPS as a vehicle for obtaining measures of consumer satisfaction or the degree of satisfaction with government output. One does not want to contaminate the employment and earnings data. Bixby concluded that "Perhaps questions could be developed and, as an experiment, used with each expiring CPS rotation group over a number of months."

Arnold Zellner, who was the discussant of the next two papers, noted that both of these statistical papers took into account survey design. Richard Porter's paper shows that the choice of regression technique for estimating coefficients for a population depends on the sample design if individuals have different coefficients. After reviewing some aspects of sample survey theory, he uses the results to form estimators for the random coefficient regression model based on panel data. He also explores the proper weights to be used in the linear model when different strata are sampled at different rates.

Frederick Scheuren uses the log linear model to estimate the probability of being poor, utilizing tabulations from the CPS. First he formulates a simple model relating poverty to race, sex, and age; taking the published tables from 1959 to 1970, he shows that the impact of race on poverty has declined, particularly between 1965 and 1968. He then "combines" published tables by using the published marginals to obtain a *fitted* version of tables (or cross-tabulations) which are not published—thus incorporating the influence of work experience and education. Next he turns to estimating the bias in model coefficients, and suggests ways to estimate the variance. Scheuren concludes that the technique is a powerful tool for describing the population. It may be that the micro-analytic models based on individual observations will depend heavily on the relationships found through "Ransacking CPS Tabulations."

One of the uses to which CPS microdata files have been put is the simulation of income transfer and tax policies. John Moeller presents some lessons learned in the course of the Urban Institute's development of TRIM, a micro-simulation system. Although the effect of the computer and the new data bases on the policy-making process is apparently not unsubstantial, Moeller stresses how causation also runs in the other direction. The legislative process required over 150 different production runs of alternative grant/tax simulations. From this experience, certain system economies in data manipulation and grant/tax eligibility determination became obvious, as overlaps in cost and coverage estimations occurred. Moeller then describes a generalized model designed to include modular construction and parameterization. Although the revised model was costly to create, the benefits probably outweigh the costs in the long run, not only in terms of analysts' and programmers' time, but in increased flexibility for policy analysis.

The discussants, Robinson Hollister and Edward Gramlich, both questioned the relevance of results obtained from a "passive system" which assumed that

changes in the tax structure did not affect economic behavior. Gramlich noted that "...despite its complexities, TRIM basically does not deal with human behavior, only with the effect of various legal and administrative arrangements in distributing and taxing money. Its policy simulations concern hypothetical tax-transfer programs, and cannot be validated in the usual sense." In response, Moeller stated that those involved in the project were working on the estimation and inclusion of behavioral parameters. But simulations modified in this manner will have to wait until we have a much better understanding of relationships underlying labor market participation.

Discussion at the Conference then turned to the general criteria for adding or deleting questions from the CPS. Hollister asked whether small scale experiments were tried to test for the *analytical* usefulness of different classes of data. The marginal social benefit-cost ratio for different types of information would be a useful item for determining priorities within Census. Although we have some measures of the marginal social returns of reducing the sampling error of crop and livestock statistics reported by the USDA,¹ we have no comparable estimates for portions of the CPS. A similar methodology might be applied to resource allocation within government agencies to relate the impact of various policies to stated goals.

The next article in this issue of the *Annals* is Jodie Allen's "Guide to the 1960-1971 Current Population Survey Files." The author of this *Data Base Review* was also a key participant in the preparation of the data files. But she is not uncritical of the tapes. Rather, her paper is a realistic evaluation of a large effort at the Urban Institute involving documentation, error correction, formatting, and analysis. Even the researcher not particularly interested in the CPS data base will find her discussion of data documentation enlightening. She reminds us that these CPS data files were not collected nor formatted with micro-analytic research in mind. Now, it is possible for economists and other social scientists to use these earlier files, but we should be aware of problems with the data. For example, the weights assigned persons and families raise some analytical issues.

While researchers might be thankful to those within and outside government who made the files available, one is reminded of Thomas Juster's plea for systematic collection of microdata with analysis in mind.² Availability as an afterthought is a weak second best, as Allen's chronology of changes in coverage and formats makes clear. When she turns to the files for March 1968 to the present, her comments are more favorable, or at least contain fewer warnings than for the earlier files.

The following two papers also deal with the Urban Institute's research program in this area. Nelson McClung's note on "Editing Census Microdata Files for Income and Wealth" represents another example of data correction procedures which reduce errors by "adjusting reported income to amounts which are close to those estimated by other and presumably more reliable sources." He matches observations on the basis of particular characteristics, but does not merge data bases. Sims has questioned the usefulness of such synthetic data bases

¹ Yujiro Hayami and Willis Peterson, "Social Returns to Public Information Services: Statistical Reporting of U.S. Farm Commodities," *American Economic Review*, March 1972, pp. 119-130.

² F. Thomas Juster, "Microdata Requirements and Public Policy Decisions," *Annals of Economic and Social Measurement*, January 1972, pp. 7-16.

in another context, and the problem of possible biases introduced by such procedures can be raised again.³ Perhaps a workshop being scheduled for May 5, 1973 on Matching and Merging of Data Sets will resolve some of these issues. (A list of the members of the Executive Committee of the Conference on the Computer in Economic and Social Research is appended to this article.)

Terence Kelly's note is, in effect, another data base review. He shows how the CPS, a cross-section survey, can be the basis for time series analysis. The longitudinal microdata were obtained by an *exact* matching procedure involving identification numbers of the families, although decision rules came into play to correct for peculiarities in the assignment of identification numbers. Kelly recognizes the major limitations of the project: reduced sample size, linkages of only two-years, and sample bias. He notes that the resulting file is a sample of the nonmobile population; it shows an increase in poverty from 1963-1966, when overall CPS data show a decline. Nevertheless, the matched file has potential for some research problems.⁴

The last paper in this issue is "Problems of Access: Some Comments," by Richard Taeuber. After touching upon the privacy issue, which in principle can be solved, he examines the institutional and financial implications of different methods of access. When the costs for files, documentation, programming, and computer manipulation are totaled, one may conclude that "Large Science" has become part of economic research—which has been traditionally an industry with few scale economies.⁵

John Beresford, a discussant, summarized some of Taeuber's ideas and added a few of his own in a list of certain basic activities that are needed to achieve a standard CPS file which is available to and useable by the public:

1. Prepare detailed file descriptions.
2. Insure that codes are compatible with existing statistical software packages, and that the codes scale.
3. Identify operational definitions of universes defined in the documentation.
4. Include totals for the cases identified by each code (and including illegal codes, if any).
5. Provide dictionaries defining each concept in detail.
6. Develop brief, but precise, descriptions of procedural details on file construction, edits, coverage, data processing, and sampling.
7. Identify and distribute information on file errors and problems.
8. Identify, catalogue, and index derived files.
9. Provide data processing capability for those users who do not wish to acquire the files or who wish only a selected universe or summary tapes.
10. Evaluate existing software appropriate for file use; prepare needed user-oriented software; and distribute, install, update, document and maintain such software.

³ See Benjamin Okner, "Constructing a New Data Base from Existing Microdata Sets: the 1966 Merge File," *Annals of Economic and Social Measurement*, July 1972, along with the Comments and Rejoinder by Christopher Sims.

⁴ For a description of the Urban Institute's modelling effort, see Harold W. Guthrie, *et al.*, "Micro-analytic Simulation of Household Behavior," *Annals of Economic and Social Measurement*, April 1972.

⁵ See Charles C. Holt, "A System of Information Centers for Research and Decision Making," *American Economic Review*, May 1970.

11. Hold training sessions on CPS files and software for users.
12. Collect information on results of useage and report these on a regular basis via newsletter.
13. Stimulate the formation of problem-oriented user groups for the exchange of information, applications, and new techniques.

The Urban Institute has certainly taken the lead in providing services for making *past* CPS data files available, but a research organization may not be the best institution for providing such services for the *public use sample*. However, there was a consensus that if Census cannot, or will not, facilitate access in a support mode, some organization must.

The activities described by Beresford are clearly based on those performed by DUALabs (National Data Use and Access Laboratories) in connection with the 1970 Census.⁶ That organization provides an alternative model for a Data Support Center. DUALabs is a private nonprofit institution, established (with Ford funding) to assist a consortium of universities in the acquisition and use of Census material—particularly, in providing derivative tapes and packing them more efficiently. With NSF aid and together with the Center for Research Libraries, it has created a Clearinghouse and Laboratory for Census Data (CLCD). The CLCD performs many systems management functions, including consulting, group training, and publications (*Data Access News* and *Technical Bulletins*). In addition, it can serve in an advisory capacity to Census and as the key node of an information system on users and uses of census tapes. The sharing of resources through the consortium has made possible research which otherwise could never have occurred.

University-based Data Centers are a third type of support institution, evolving from research efforts which have focused on particular data bases. Substantial scale economies have been achieved, and the maintenance of the data base has been linked to substantive research at the university, facilitating interaction between producers and users. For example, the Survey Research Center at the University of Michigan has a databank from its Economic Behavior Program, whose content ranges from detailed financial information collected by the annual Survey of Consumer Finances to quarterly economic attitude surveys and special studies. The OSIRIS system is used for the analysis of these data. The University of Wisconsin has also been engaged in the development of software for social science applications. In particular, SEOSYS was developed for the retrieval of information from the Survey of Economic Opportunity, and a more general Social Science Information Management System (SIMS) is under development.⁷ Work at the Brookings Institution using Internal Revenue Service tax files is another example of a large scale research effort at a nonprofit institution.

Even with the various institutions specializing in the use of particular data bases, there is competition for improved models, software, and research results. Thus, there are still incentives for the efficient production of research. The conference concluded on the following note: What institutional form is appropriate

⁶ John Beresford, "The 1970 Census Start Community," *Annals of Economic and Social Measurement*, January 1972.

⁷ Max E. Ellis, "Social Science Computing at the University of Wisconsin: SIMS and SEOSYS," *Annals of Economic and Social Measurement*, April 1972.

for facilitating and encouraging use of the new public use CPS files? The articles in this issue describe methodological and organizational problems faced in the past; perhaps the key to the future will be found in these experiences.

*University of Florida
National Bureau of Economic Research*

APPENDIX

Executive Committee of the Conference on the Computer in Economic and Social Research 1972-1973

<i>Members</i>	<i>Affiliation</i>	<i>Term (years)</i>
Ivan Fellegi	Statistics Canada	3
Harold Watts	University of Wisconsin	3
Arnold Zellner	University of Chicago	3
Gregory Chow	Princeton University	2
Joel Popkin	Bureau of Labor Statistics	2
James Smith	Pennsylvania State University	2
Charlotte Boschan	National Bureau of Economic Research	1
Charles Holt	Urban Institute	1
Richard Ruggles (Chairman)	Yale University	1
 <i>Ex Officio</i>		
M. I. Nadiri	National Bureau of Economic Research, and New York University	
Sandy Morton	Executive Secretary, Conference National Bureau of Economic Research	
 <i>Observers</i>		
David Kresge	<i>Latin American Conference,</i> National Bureau of Economic Research, and New York University	
Sanford Berg	Editor, <i>Annals of Economic and Social Measurement,</i> and University of Florida	
Donald Farrar	National Bureau of Economic Research	