

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 3

Volume Author/Editor: NBER

Volume Publisher:

Volume URL: <http://www.nber.org/books/aesm73-3>

Publication Date: July 1973

Chapter Title: Conference Notes: Conference on Research and the Public Use Samples

Chapter Author: Cynthia Taeuber

Chapter URL: <http://www.nber.org/chapters/c9912>

Chapter pages in book: (p. 352 - 357)

## CONFERENCE NOTES

### CONFERENCE ON RESEARCH AND THE PUBLIC USE SAMPLES

BY CYNTHIA M. TAEUBER

A conference on Research And The Public Use Samples, co-sponsored by the National Bureau of Economic Research and the Southern Regional Demographic Group, was held in Atlanta, Georgia, March 23-24, 1973 (a program is appended). The objective of the conference was to explore the potential uses and problems of the census Public Use Samples (P.U.S.) for a wide variety of users. Paul Zeisset, Bureau of the Census, opened the conference with an overview of the Public Use Samples. In this he gave an historical report of the Census Bureau's development of the samples, discussed services available to users of the P.U.S., and announced the forthcoming availability of other Public Use Samples from the 1970 Census of Puerto Rico, the 1970 Employment Survey, and the Current Population Survey (1968-1971). Jack Beresford, DUALabs, commented that the 1970 Census Public Use Samples should receive wide distribution and use; and he observed that social science work has entered a new stage in which the use of public data will become a part of the common experience of all social scientists.

A panel discussion by discipline-oriented researchers highlighted the previous and potential uses of the Public Use Samples. Jim Sweet, University of Wisconsin, underlined the rich opportunities offered by the P.U.S. on a number of topics in the area of the family, including labor force participation of women, family composition and living arrangements, and marital disruption. Researchers will be able to investigate these processes in detail as they occur in specific sub-populations such as ethnic groups, the affluent, and the poor. The authors of the paper on aging and mortality, Beth Soldo and George Myers of Duke University, noted a number of special features of the Public Use Samples that are of value in studying this topic; for example there is very little published data on the aged, cross tabulations are minimal, and there are not detailed breakdowns by the older ages. The P.U.S. also allows the researcher to collate household information for the non-institutionalized aged, a factor which has been absent in studies of this group.<sup>1</sup> Turning to the study of migration, Larry Long of the Census Bureau, stated that with the Public Use Samples, the freedom of a researcher is increased because he need not be bound by the printed reports of the Census Bureau which are prepared without the benefit of prior analyses. Information on "mobility status during the 1965-1970 interval" and "year moved into present dwelling unit" seemed to Dr. Long to be particularly useful for research into the ways in which the family structure influences migration decisions. Charles Nam, Florida State University, reviewed the content of the 1960 and 1970 Public Use Samples related to socioeconomic analysis. He surveyed previous research uses of these data and suggested types of analyses that could be made of the information.

<sup>1</sup> See "The Public Use Samples and Research in Aging and Mortality," by George C. Myers and Beth Soldo, *Review of Public Data Use*, Volume 1, Number 2, April 1973.

The Friday afternoon session examined analytical strategies for use with the P.U.S. Richard Rockwell of the University of North Carolina discussed the matching of the 1970 P.U.S. with other data files and, as an example of this method, he matched the P.U.S. with the Survey of Economic Opportunity (SEO) data to study the effects of different types of behavior on fertility. Subjects in the two files were matched by various socioeconomic variables to create a hybrid data base that allowed new questions to be studied at a low cost. Richard Ruggles, of the NBER, was discussant; he emphasized that Rockwell's work was analytically very interesting and that such experimentation holds much hope for future progress in this area. With regard to the specific problem chosen, Mr. Rockwell could have increased his sample size from 5,000 to 50,000 by matching the P.U.S. with itself; i.e. by matching one age-specific P.U.S. group with an "aged-back" group also from the P.U.S., rather than by matching the SEO file with the P.U.S.

In a jointly authored paper by Martin Levin and William W. Pendleton, the perspective of structural effects is proposed as a useful model for the analysis of demographic processes with the neighborhood Public Use Samples. In particular, it is argued that the structural effect model provides both a logical framework for such research and an interpretative mechanism to further understanding.

Simulation and modeling uses of the Public Use Samples were discussed by Guy Orcutt, Yale University, and Bob Michielutte, Bowman Gray School of Medicine in Winston-Salem, North Carolina. Orcutt discussed the need for publicly available microdata sets in order to develop microanalytic models of social systems. An example of this type of model is the Urban Institute Poverty and Inequality Modeling Project which, when given a sample representation of the population at a particular moment imputes events to individuals and families over a period of time. The creation of the P.U.S. has contributed immeasurably to the development and policy application of microanalytic models.

Michielutte focussed on the relationship between the P.U.S. and microanalytic modeling and simulation, particularly the development of causal models and the use of microanalytic simulation. With respect to the development of causal models, a number of assumptions must be made including standard errors, random measurement errors in the sample itself, and assumptions about the modeling procedure. Careful attention must be paid to the type of simulation to be used for analytic purposes.

Charles Laidlaw of the Baltimore Regional Planning Commission, explained his uses of the P.U.S. for regional planning purposes. Laidlaw said that the problems of using the P.U.S. were inherent in the sample itself rather than in the geographic area being studied. He uses the P.U.S. to study characteristics of the Baltimore metropolitan area for example, sources of in-migration, special characteristics of special populations, and household size pattern. Mr. Laidlaw cautioned that before using the P.U.S. one should check to see if the data needed are not already available in 4th and 6th counts from the Census and one should also be certain that the county group chosen from the P.U.S. for study matches the metropolitan area.

Richard C. Taeuber announced that there was a possibility that financial support could be found for the 1940 and 1950 Public Use Samples if the research community could justify the expense. Those wishing to support this effort were

asked to write letters explicating the need to Dr. Taeuber at the Oak Ridge National Laboratory in Oak Ridge, Tennessee.

The final phase of the conference was concerned with the technical problems of handling the data base. Bill Downs discussed how the Census Bureau used allocations to resolve the problem of missing information and suggested some considerations the P.U.S. user should be aware of when using data with allocations. Peter Bounpane of the Statistical Division of the Census Bureau briefly described the P.U.S. selection method and discussed the rationale and assumptions in the standard error tables of the P.U.S. documentation, comparing the efficiency of this method to other methods of estimating standard errors.

The last session was a panel which discussed software, large vs. small computers, and prospects for innovative approaches. Martin Levin and William Pendleton, of Emory University, discussed a data processing system for handling the 1970 P.U.S. In addition to a standard cross-tabulation capability, the system includes sophisticated statistical procedures, a data compression feature to reduce the physical size of the data set, and English language-type input instructions which require very low training costs. Moreover, the system maintains the integrity of published documentation. James Sakoda, Brown University, described his statistical package written in FORTRAN IV for use with small computers, one which provides many of the features of the larger packages such as SPSS or DATATEXT. These include data conversions and recoding, alphabetic table headings, six-way cross-tabulations, summary statistics, one-way AOV and *t*-tests, correlation coefficient and test of linearity. Joan Haworth of Florida State University noted that the approach used at that institution was an ad-hoc one. Gary Hill of DUALabs discussed two English-language computer systems being developed by DUALabs to make the 1960 and 1970 Public Use Samples more accessible. Public Use Sample Helper (PUSH) enables a user to create subsamples and restructured files which can then be processed by existing analytical software packages such as SPSS; and CENTS-AID/CENTS is a "hyper-speed" approach to creating cross-tabulations and machine-readable summary data files from the original Public Use Samples.<sup>2</sup>

*Southern Regional Demographic Group  
P.O. Box 117  
Oak Ridge, Tennessee 37830*

<sup>2</sup> These computer systems are described and contrasted with SPSS, DATATEXT, and other systems in "Maximizing Access to the Public Use Samples," by Gary L. Hill, Lawrence L. Brown III, and Kisun Han, *Review of Public Data Use*, Volume 1, Number 1, December 1972.

APPENDIX: PROGRAM

*Conference on  
Research and the Public Use Samples  
March 23-24, 1973  
Emory-Sheraton Inn, Atlanta, Georgia  
Co-sponsored by the Southern Regional Demographic Group and  
the National Bureau of Economic Research*

1. Overview of the P.U.S.  
History, Perspectives and Structure  
Speaker: Paul Zeisset, Census Bureau  
Discussant: Jack Beresford, DuaLabs
2. Research and the P.U.S.  
A Panel Discussion by discipline-oriented researchers of previous and potential uses of the Public Use Samples
  - Family Jim Sweet, University of Wisconsin
  - Aging and Mortality George Myers and Beth Soldo, Duke University
  - Migration Larry Long, Census Bureau
  - Socio-Economic Characteristics Charles Nam, Florida State University
3. Analytical Strategies for the P.U.S.
  - A. Matching the 1970 P.U.S. With Other Data Files  
Richard Rockwell, University of North Carolina  
Discussion with questions from the floor  
Richard Ruggles, National Bureau of Economic Research
  - B. Structural Effects Analysis for Demographic Research with the Census P.U.S.  
Martin Levin and W. W. Pendleton, Emory University
  - C. Simulation and Modeling Uses  
The Affinity of Public Use Samples and Microanalytic Models  
Guy Orcutt, Yale University and Urban Institute  
Discussion with questions from the floor  
Bob Michielutte, Bowman Gray School of Medicine, Winston-Salem, North Carolina  
The P.U.S. for Regional Planning Purposes  
Charles Laidlaw, Baltimore Regional Planning
4. Handling the Data Base
  - A. Problems with the Data Base  
Bill Downs, Housing Division, Bureau of the Census
  - B. Sampling Problems and Error Rates in the P.U.S.  
Peter Bounpane, Statistical Methods Division, Bureau of the Census
  - C. Processing: Software and Documentation—A Panel  
Joan Haworth, Florida State University  
Martin Levin, Emory University  
Gary Hill, DuaLabs  
James Sakoda, Brown University  
Software: SPSS, CENTS, other packages

Large vs. small computers  
Report generation vs. statistical analysis  
Prospects for innovative approaches  
The need for generalized extraction programs

## CONFERENCE ON ECONOMETRICS AND MATHEMATICAL ECONOMICS

The Conference, which was created by a grant from NSF in 1970, has the purpose of stimulating research on recent topics in mathematical economics and econometrics. During 1972 the ten existing seminars of the Conference met fifteen times at universities through-out the country, and two new seminars were formed. The seminars and their leaders are:

- General Equilibrium Models  
Kenneth J. Arrow, Harvard
- Evaluation of Econometric Models  
Saul Hymans and Harold T. Shapiro, Michigan
- Comparison of Econometric Models  
Lawrence R. Klein, Pennsylvania
- Decision Rules and Uncertainty  
Daniel L. McFadden, Berkeley
- Decentralized Economic Planning and Programming  
Roy Radner, Berkeley
- Distributed Lags and Time Series Analysis  
Christopher Sims, Minnesota
- Optimal Economic Growth  
Joseph E. Stiglitz, Yale
- Bayesian Inference in Econometrics  
Arnold Zellner, Chicago
- Quantitative Studies in Industrial Organization  
George J. Stigler and Lester G. Telser, Chicago
- Monetary and Fiscal Analysis  
William C. Brainard, Yale
- Franco Modigliani, MIT
- Analysis of Panel Micro-Data  
James N. Morgan, Michigan
- Public Economics and Nonmarket Decisions  
Martin McGuire and Mancur Olson, Maryland  
[formerly Studies in the Micro Public Sector, Lester Thurow, MIT]

Conference participants have been pleased to have an opportunity to meet in seminar with other economists at work on related problems. Seminar sessions are focused on specific issues, and papers are frequently circulated in advance. More than fifty working papers have now been presented in the seminars, and a number of these have subsequently been published or presented at professional meetings.\* Also, graduate students at the host university are often invited to attend seminar sessions. The Conference thus provides a new and apparently quite successful forum for research.

\* A list of papers currently on file may be obtained by writing to the Secretary, Conference on Econometrics and Mathematical Economics, 155 Whitney Avenue, New Haven, Connecticut 06510.