

DISCUSSION

Gary M. Shapiro, U. S. Bureau of the Census

This was an overall good set of three papers. I will comment on each paper separately.

1. KASPRZYK-LININGER PAPER

This is an excellent overall review of the 1982 Survey of Income and Program Participation. The paper was clear and thorough. The intended purpose of the paper was achieved, and it served as a good introduction to this session.

2. RODGERS-DeVOL PAPER

This is a very interesting paper and potentially one of the most important at this convention. The theoretical portion of the paper is rather controversial in that the authors seriously question the value of statistical matching. It is, in fact, suggested that it may be almost never, worthwhile to perform statistical matching. The authors' reason for this is that the Y and Z variables may be correlated, violating an inherent assumption of statistical matching. I was convinced by the logic of the authors' argument though I am sure many practitioners of statistical matching are not so easily convinced as I.

The main question now is whether the empirical study confirms the theoretical conclusion. The weakness of this paper, of course, is that more empirical results weren't ready in time for inclusion in the paper. At this time, an overall conclusion from the empirical study is not possible. I eagerly await a sequel to this paper in which more general conclusions can be made.

3. HEERINGA PAPER

This paper suffered even more than the Rodgers-DeVol paper from the unfortunate shortage of empirical results available prior to the convention, and I look forward to the full results. The paper provides a very good review of alternative methods for making small area estimates, which I personally found educational and refreshingly easy to understand. I have only one minor comment: It is implied that the National Center for Health Statistics (NCHS) originated synthetic estimation in the late sixties, and that following this the Census Bureau worked on synthetic estimation. Actually, NCHS originated the term "synthetic estimation," but the general methods had been used much earlier by several people. For example, there was a 1945 Radio Listening Survey conducted by the Census Bureau which employed the basic ideas of synthetic estimation.

Finally, I have one off-hand observation on the preliminary empirical results. Very large correlations between the auxiliary characteristics and the characteristics of interest were achieved--the range was .84 to .995. With such large correlations, I would have expected very good synthetic estimates, and am puzzled as to why the estimates are as bad as they are. It would be useful if the author could throw some light on this.