

## DISCUSSION

Herman P. Miller, Bureau of the Census

Mr. Hearn is one of a small but valiant group of men who have spent many years pushing the use of administrative records as a source of general statistical information. As far back as 1947 Saul and others like Jacob Pearlman, Ben Mandel and Irwin Wolkstein have been displaying the statistical wares of the Social Security Administration and urging other government agencies to partake of them. Much has already been accomplished as a result of their efforts, but I am sure that what we have seen is only the beginning. Social Security records are now being used to produce a variety of general purpose statistics such as County Business Patterns. Efforts are also being intensified to increase the use of Internal Revenue Service tax records to produce general statistics. I have no doubt that about 10 years from now, when the IRS files are fully automated, we will see statistics produced each year based on an amalgamation of data from the files of the Census Bureau, the Social Security Administration, and the Internal Revenue Service. At that time, men like Saul Hearn will be able to say with some satisfaction, "I told you so."

The paper that Mr. Hearn has given us today is in the tradition of those given on the same subject in the past. It contains a straight forward factual description of the way in which the QASI work history sample is selected and a very brief discussion of some of the limitations of the data based on that sample. The difficulty that I find with this kind of paper is that there is not much to criticize in it, nor is there much to praise. The story Hearn tells is a good one, but I have heard it before and no longer find it very exciting. This does not mean that it is not worthwhile. I am sure that when the day comes that I want to use this sample, I will be very happy to have the description that was presented today. My chief criticism is that it contains too much description and not enough analysis.

Near the end of his report, Hearn mentions some of the limitations of the data based on the work history sample. Although everything he says is correct, I feel that he passes too lightly over some defects that turn out to be major obstacles to analysis. For example, he states that the limited coverage of QASI in the early years is a problem, but he reassures us by stating that over 90 percent of the workers are now covered. Farber's analysis suggests, I think, that this overall figure may be misleading when the data are used to examine trends for the low-income groups. I shall say more on this point later. The relatively low upper limit of taxable wages is also stated as a limitation by Hearn. Here again it is only when the data are put to use, as they are in the Farber study, that we can see how very serious this shortcoming can be. Incidentally, here is point where I think Hearn could have

made a real contribution with relatively little effort. There is no need in analytical studies involving the use of QASI data to be saddled with the limitation imposed by the failure to count wages above the taxable limit. There are methods for extrapolating taxable wages above the upper limit to represent full year totals. My understanding is that these estimates are fairly good. If they are not good, we should be told about it; but if they are good I would like to know why they are not being used. It is not very meaningful to talk about high-wage workers whose earnings start below the median.

I believe that Hearn's paper would have been a much more interesting and useful document if he had taken the time to explore somewhat more fully the implications of the procedures he has used instead of confining himself as he did to factual description. Let me cite one more case in point before turning to the Farber paper. In the last three sentences of his paper, Hearn mentions the possibility of matching QASI data with IRS records. In less than a decade this possibility should become a reality. A matching study of this type would open up large areas for analysis and would eliminate some of the defects that now exist in the wage records. I am sorry that he did not say more on this subject, particularly since I am now working on the problem of using administrative records at the Bureau of the Census.

Dave Farber deserves much credit for presenting in great detail one kind of use that can be made of the work history sample. I am sure that a vast amount of work has gone into this paper as well as considerable statistical dexterity and imagination. He has created a set of statistics which, I am sure, will receive considerable use in years to come. Even more important are the concepts and procedures he has developed.

Much as I credit Farber for the work he has done, I cannot say that I agree with his interpretation of the results. I found a disturbing naivete in his analysis and a tendency to accept things too much at their face value. The results are often treated as though they represent the entire universe instead of a segment of it and insufficient attention is paid to the restrictions that are imposed on the data because of the limited nature of the universe. There is also a tendency to push the figures too hard and to make generalizations that are based on small differences. In my discussion of several specific aspects of this paper, I shall confine my attention to only a few points that I consider fundamental. The fact that I am first at the trough does not give me the right to drain it. I am sure the other discussants will stress other points and that they will have more to say about some of the things that I will only touch upon.

Farber's major purpose in writing this paper was to examine changes in the level of wages for different groups of workers during the years 1951-57. Since this period includes the 1954 recession, he pays particular attention to changes during that period. Although the analysis was made for a dozen different age groups, age does not appear to be his chief concern, as the title suggests. After all, workers don't age very much in a seven-year period. I think age is really used in this study more as a way of identifying cohorts than as a method for analyzing changes over time. But, this criticism is perhaps only incidental.

The universe used in the Farber study are those workers who had some covered employment in 1957 and who had only wage credits in earlier years. He traces the work histories in covered employment of these individuals back to 1951 and divides them into four groups based on covered wages per year employed: under \$1,200 (low-paid); \$1,200 to \$2,400 (intermediate low-paid); \$2,400 to \$3,600 (intermediate high-paid); and \$3,600 or more (high-paid). If we stop here and think for a moment about what he has done, we can detect several of the basic defects in this study.

First of all note that the study is restricted to persons in covered employment in 1957 and their work histories in covered employment in earlier years. Because of this limitation we cannot measure changes in the level of covered employment based on this sample nor can we say anything about changes in total employment even for this panel of workers. Yet, Farber attempts to make generalizations on both of these phenomena on the basis of his sample. For example, he converts the number of workers in each cohort with covered employment in each year into an employment index and makes generalizations based on that index. This does not seem to me to be a valid procedure. Let us consider, for example, his findings for the low-paid cohort of men aged 35-39 years in 1957. There were a total of 419 men in the sample. All of them had covered employment in 1957 but only 241 (58 percent) had covered employment in 1951. On this basis Farber says the employment index rose from 100 in 1951 to 174 in 1957. It did no such thing. All he knows is that 40 percent of the men in this cohort who worked in 1957 had no covered employment in 1951. I find it hard to believe that so large a proportion of the men in this age group really did no paid work at all during the year. After all, they were on the average about 31 years old at the time. I suspect that in the low-paid group there is a considerable amount of uncovered employment that does not get reflected in the figures. There may also be reporting or processing errors of one type or another. Whatever, the reason, a good deal of qualification is needed for Farber's conclusion that there is a great deal of intermittency of employment in the low-paid group. I am sure he is right, but I don't think these figures prove the case.

The limitation of taxable wages to \$3,600 for 1951-54 and to \$4,200 thereafter also creates serious problems of interpretation and, I believe, has led Farber to some erroneous conclusions. He finds, for example, that "for the high-paid male cohort, the average annual wage credits .... remained stable throughout the 1951-57 period." Little wonder that they did. The minimum needed to qualify for the high-paid category was an annual average of \$3,600 in covered wages; but no count was made of wages above \$4,200. Therefore, there was no way in which the average could rise substantially because wages above \$4,200 were not included. There could have been significant increases in the total wages of this group that would not be reflected in the statistics. This error in interpretation leads Farber to the erroneous conclusion that there was a narrowing of wage differentials between the highest paid and the two intermediate cohorts. All that happened here, in my opinion, is that the average for the two lower groups rose, as might be expected during a period of relatively full employment like 1951-57. But, because of the income limitation, the income of the top group did not rise. As a result there was an apparent narrowing of wage differentials. As I mentioned earlier, there is a way to correct this defect in the data and it seems to me that it should receive serious consideration in the future and perhaps be carried back to the earlier years if possible. Procedures have been developed for extrapolating covered wages to represent annual totals. Assumptions are involved in preparing such estimates but the end product in my opinion would be more useful than the unadjusted totals. Such figures would really permit us to make some generalizations about wage trends for the higher-paid workers which cannot, in my opinion, be made on the basis of the present data.

Before concluding, I should like to say that it would be interesting to speculate some time about what we would do with the work history data, even if they had no limitations, that we cannot now do with the cross-section data. This is not intended as a criticism of Farber since he limited his discussion to a consideration of wage movements for a relatively short period. I have thought about this matter from time to time as I worked with the cross-sectional census results, but I have never come up with a clear-cut answer. I have a feeling that some concentrated thinking along these lines might pay handsome dividends.

In conclusion, I should like to say that I have tried to be fair in my criticism of these papers; but I have not worked very closely with this set of numbers and I could be mistaken in some of my points. If so, I am sure Farber or Hearn will let me know about it. I also hope that I have not created the impression that I did not enjoy these papers or that I do not think that the work history sample or Mr. Farber's use of it are unimportant. I think they are important and useful and they should

be continued. I must confess, however, that as I prepared these comments, I asked myself if I would undertake a study of these data, were I requested to do so. The answer, I am sorry to say, was "no" because I could see too many shortcomings. I suspect that I am not alone in this view. This attitude will, I hope, provide some food for thought for those whose respon-

sibility it is to keep this project going. If the limitations of the data are so great as to frighten away prospective analysts, it may be worthwhile to give some thought to improvements that may be made. I see little hope for such improvements for the lowest-paid workers, but perhaps more can be done for the top income groups.