COMPUTER METHOD TO PROCESS MISSING INCOME AND WORK EXPERIENCE INFORMATION IN THE CURRENT POPULATION SURVEY

Emmett F. Spiers and Joseph J. Knott^{*} U.S. Bureau of the Census

Introduction

Missing information in interviews, a common problem found in all sample surveys which depend on public cooperation, can be handled primarily in three ways. One method is to show in detail reported income information only and allow the user to make all adjustments necessary to take account of missing data. The second method is to inflate respondents' data to control totals assuming that the characteristics of respondents and nonrespondents are alike. The third method is to impute missing data for nonrespondents based on information obtained from respondents with social and economic characteristics similar to those of nonrespondents.

Starting with the March 1962 Current Population Survey (covering 1961 income), when a respondent does not answer one or more of the income items in the CPS, an imputed amount is given to the nonrespondent based on the income information reported by a respondent with similar demographic and economic characteristics, such as age, sex, family status, color, number of weeks worked, earnings, and major occupational group. The income amount(s) assigned to a nonrespondent are the income amount(s) stored from the last respondent who otherwise had the same characteristics.

The March CPS Supplement includes eight questions on a person's income during the preceding calendar year.1/ In the February and April CPS supplements, questions are asked on a person's work experience (number of weeks worked last year, class of worker, occupation and industry of longest job, etc.) in the preceding year.2/

Information collected in the February, March, and April CPS supplements provides the basis for a number of Census Bureau publications. The wide interest in and use of data covering income and work experience makes the problem of income and work experience nonresponses of particular interest.

This paper, divided into two parts, outlines the computer methodology involved in this allocation procedure. The first part of the paper describes in summary fashion the computerized "allocation" procedure used to impute missing income and work experience information using the UNIVAC 1107 computer. The second part is a set of flow charts describing the logical flow used in imputing missing income information.

Allocation Procedures

A discussion of the field collection and editing procedures is beyond the scope of this paper, but it is noted that limited editing of the data is done in the field before the information on the questionnaires is converted to magnetic tapes for computer processing. After initial computer editing and format conversion, the February-April work experience data are added to the March income record for further joint processing. The matching of the work experience data with the income data is done on the basis of unique household and person numbers. If a February or April record corresponding to the March record is not found the work experience data are allocated based on the March data including age. sex, earnings, and family relationship. It should be noted that all March data excepting income have been edited and any necessary allocations have been made.

The actual processing of missing income and work experience data may be thought of as five separate steps or functions: (1) Work experience allocation; (2) earnings-work experience consistency edit; (3) earnings allocation; (4) postearnings allocation edit; and (5) "other income" allocation. Each of these steps or functions is discussed below in some detail.

(1) <u>Work Experience Allocation Procedures</u> (Chart A)

Work experience data are allocated in two situations:

1. When there is a March record and no corresponding February or April record can be found (nonmatch).

2. Work experience information had been allocated before the match occurred in a procedure not using earnings data. In order to have all missing work experience data allocated using the same procedure, the matched work experience data, with certain items previously allocated, are rejected and reallocated in this procedure using earnings data.

(2) <u>Earnings-Work Experience Consistency</u> Edit (Chart B)

A primary purpose of the earnings-work experience consistency edit is to eliminate illogical combinations of class of worker and type of income. For example, an employee who worked for a private company last year should have had some wage or salary income. Therefore, a person who worked more than 13 weeks last year (the smallest available interval) is edited in one of two ways if he reports no earnings in the earnings type consistent with his class-of-worker status.2/ In the first way an inconsistent earnings type is exchanged with the "none" in the consistent earnings type. In the second way the consistent earnings type is made NAM/ so that the person will go through the earnings allocation procedure. The effect of these

* Comments by members of the Consumer Income Statistics Branch and Dr. Murray S. Weitzman, Assistant Division Chief, Economic Statistics Programs, Population Division, are gratefully acknowledged. procedures is to force a person to have an earnings amount consistent with his class-of-worker status.

Secondly, this edit is used to check a person's response to the number-of-weeks-worked-lastyear question against his earnings data to see if the two are consistent. If a respondent reported he did not work last year, logically he should have had no earnings. In the present method a respondent must have less than \$300 in earnings if he reports he did not work last year. If a respondent reports he did not work last year but also reports \$300 or more in earnings, his work experience and earnings are inconsistent and either his earnings or work experience should be changed. The earnings data are assumed to be "correct" at this stage of processing and the work experience data are rejected and allocated from the last respondent with similar characteristics.

Thirdly, the income-work experience edit identifies and treats persons who have no work experience data and also have incomplete earnings data (three "NA's" or a combination of "NA's" and "Nones"). The allocation procedure is designed to use known information to allocate missing or unknown information. Very little economic information is available for a person whose work experience information is missing or unusable and whose earnings data are missing or incomplete. These cases are assigned a temporary earnings amount or "none" in the first earnings type which is NA. Wages or salaries earnings are checked first, nonfarm self-employed earnings next, and then farm self-employed earnings. Fortunately, this combination of missing earnings and work experience data does not occur frequently.

(3) Earnings Allocation Procedure (Chart C)

Missing earnings data are allocated from the reported earnings data of the last person with similar socioeconomic characteristics. The socioeconomic characteristics used in the earnings allocation procedure are:

- (a) Number of weeks worked last year;6/
- (b) occupation of longest job last year;
- (c) family relationship;
- (d) sex;
- (e) age;
- (f) color (white or nonwhite); and
- (g) class of worker.6/

These six characteristics are combined in various ways to divide the records into 226 mutually exclusive classes. Each time a person answering all three earnings items is encountered, his three earnings items are stored temporarily in the appropriate class until another fully reported person having the same characteristics is encountered. When the next fully reported person belonging to the same class is encountered, his earnings data are stored replacing the previously stored data. This process continues throughout the processing. If at any time a person having any missing earnings data is encountered, he is assigned the earnings value(s) of the last person stored in his class. The division of the records into these 226 classes is done with two objectives in mind. First, each class should be fairly homogeneous, i.e., the variables used in stratifying persons should be correlated with earnings. Second, the number of persons passing through each cell should be large enough to avoid having many persons' earnings allocated from the same fully reported person. The danger is that an atypical fully reported person could bias the allocated group if more than a few NA's are allocated using the fully reported person's atypical earnings data.

(4) <u>Post Earnings Allocation Consistency</u> Edit (Chart D)

The purpose of the post earnings allocation consistency edit is very much like the earningswork experience consistency edit. However, only persons who were allocated one or more types of earnings are affected by this edit (an allocation may be either a dollar amount or a None).

Inconsistencies are caused by not controlling on class of worker uniformly throughout the allocation matrix. For example, if the last fully reported person in a class is a wage and salary worker, a self-employed worker with an NA in nonfarm self-employed earnings is likely to be allocated a None in nonfarm self-employed earning.

(5) Other Income Allocation Procedure (Chart E)

NA's in the five types of "other income" are treated by a separate allocation procedure essentially similar to the earnings allocation procedure. Earnings is used as a characteristic in the other income allocation matrix since it is considered to be correlated with other income types. For example, a person with high earnings is more likely to have dividend income and no public assistance than a person with low earnings. Other characteristics covered are: Family relationship, sex, worker-nonworker status, age, and color (white or nonwhite). These five characteristics are grouped to form 286 mutually exclusive classes. Like the earnings allocation procedure, the five other income items of the last fully reported person are stored in his appropriate allocation matrix cell.

Unlike earnings nonresponses, other income nonresponses are divided into two types. The first type occurs when a person being interviewed fails to indicate not only the amount of a specified type of other income but also fails to indicate whether he received that type. The second type occurs when a person indicates that he received a particular type of other income but fails to report the amount. To accommodate these two types of other income nonresponses, two distinct allocation matrices of equal size were developed, designated as matrix A and B, respectively. In matrix A, each type of other income is stored, either dollar values or "nones." On the other hand only dollar values are stored in matrix B. Therefore, when a nonrespondent indicates that he had received a specified source of other income but does not report the amount, he is allocated

from matrix B, containing only dollar values. If, on the other hand a person refused to give any information (whether received or not and the amount) on a particular type of other income, that type is allocated from matrix A.

Footnotes

1/ In the March 1969 CPS the eight sources were: Earnings-(1) Money wages or salary; (2) net income from nonfarm self-employment; (3) net income from farm self-employment; Other income--(4) Social Security; (5) dividends, interest (on savings or bonds), income from estates or trusts or net rental income; (6) public assistance or welfare payments; (7) unemployment compensation, government employee pensions, or veterans' payments; (8) private pensions, annuities, alimony, regular contributions from persons not living in this household, royalties, and other periodic income.

2/ Because of the CPS sample design it is necessary to ask the work experience questions for one-quarter of the sample in the April CPS sample in order to have work experience data for all the households in the March CPS survey.

3/ The basic assumption in this edit is that a person who reports he worked more than 13 weeks last year should have some earnings of the type consistent with his class-or-worker status. Employees of incorporated private companies or any governmental unit must have a dollar value in wages or salaries to have his type of earnings and class of worker consistent. Self-employed farmers must have a dollar value in farm selfemployment earnings to be consistent.

 $\underline{4}$ "NA" means the information was not reported during the interview or was not available for some reason.

5/ The socioeconomic characteristics used in the work experience allocation procedure are: (1) Type of earnings; (2) amount of earnings; (3) age; (4) color (white or nonwhite); (5) sex; and (6) family relationship.

6/ Number of weeks worked and class of worker are not used uniformly throughout the matrix.

odic income

"Other Income" Other income data--Sources A, B,

C, D, and E

Representation

Private pensions, annuities, ali-

Earnings data---W/S, NFSE, and FSE

mony, regular contributions from

persons not living in the house-

hold, royalties, and other peri-

FLOWCHARTS OF THE INCOME ALLOCATION PROCEDURE

Notation

Source E

Earnings

The following flowcharts outline the general methodology used to allocate missing income data collected in March of 1969 as a supplement to the Current Population Survey (CPS). The flow charts cover the six functional steps, each on a separate page. The first chart shows the overall flow of operations necessary for processing each income record. The other five charts (charts A to E) show the following: Chart A---Work Experience Allocation Procedures, Chart B---Work Experience-Earnings Consistency Edit, Chart C--Earnings Allocation Procedures, Chart D--Post Earnings Allocation Edit, and Chart E---"Other Income" Allocation Procedures.

Notations used in the flow charts are outlined below:

<u>Notation</u> W/S	<u>Representation</u> Money wages and/or salaries	NA	Not available—any item of infor- mation which is not available during processing
NFSE	Net income from nonfarm self- employment	W/E	Work experience datanumber of weeks worked, class of worker,
FSE	Net income from farm self-		occupation of longest job, etc.
	employment	Matrix A	"Other Income" allocation matrix A
Source A	Social Security payments	contains the five "other income" items reported by the last pre- ceding respondent in each class.	
Source B	Dividends, interest (on savings or bonds), income from estates or		class.
	trusts or net rental income	Matrix B	"Other Income" allocation matrix B
Source C	Public assistance or welfare payments		contains, for each type of in- come within each class, the dollar value reported by the last preceding respondent hav- ing a dollar value in that par- ticular type of income.
Source D	Unemployment compensation, govern- ment employee pensions, or veterans' payments		





 \underline{l} This path is a small group of people who did not report work experience and have NA's in earnings or a combination of NA's and none's.







