

Computerized and Hand-editing Guidelines for Weeks, Hours and Wages

README

1994-2001 Hours of Work and Wage Files

Yong-seong Kim, Tecla Loup, and Frank P. Stafford

December 5, 2002

I. Introduction and Overview

Average annual hourly earnings of the head (and wife) are the result of some of the most complex processing in the PSID. (Note: There is one global question, B16 ('What is your hourly wage rate for your regular work time?'), which is asked of those currently paid hourly.) This readme is about a calculated annual hourly wage for everybody who was employed positive hours and is far more encompassing.

Historically (prior to the 1993 survey), the hourly wage calculation was accomplished substantially by preprocess editing of paper questionnaires - making case by case judgment easier. Here we calculated such variables by extensive programming code and then, ex post, applied judgmental hand editing to the remaining 'problem' cases. This approach was mandated by the resources available.

A. Computerized Processing Strategy

To get hours of work per year we start with weeks of work per year on the 'main job' and then turn to 'extra jobs'. To begin to process weeks of various activities (work for pay, unemployment, ...) and annual work hours, we first applied SAS statements to create a processing program. Then, we turned to hand editing of the cases which are individually rare and/or difficult to cover with pre-specified rules. First we processed annual weeks on 'main jobs'. Simple rules processed annual weeks missed due to the illness of others (B60-62), weeks missed due to own illness (B63 -65), vacation weeks (B66-68), weeks of strike (B69-71), weeks of unemployment (B72-B74a), weeks out of the labor force (B75 - 77a), and actual work weeks (B78) into a 52 week total. Respondents also reported, on the average, hours per week on the main job (B79) and annual overtime hours (if not included in (B79)) (B81).

The product of workweeks times reported annual average hours per weeks plus annual overtime hours represents our annual work hours measure. In the case where none of the needed input variables is anomalous and there are no 'extra jobs', this is simple. First, anomalies arise. The respondents are asked to reconcile their various components of weeks into a 52 week year at the end of B78. This is often impossible in the context of a telephone interview, and to maintain response rates, the interviewer moves to the remaining questions - with the anomalies left for post-field judgments. Second, there are multiple main jobs possible and extra jobs can be held during the year (B82-B93 and B94). These extra jobs may fully, or partially overlap with the main jobs - or, in some cases be the only job held for some time period.

To provide a better picture of the year's events, the PSID includes month strings on time out of the labor force, time unemployed, and months during which an extra job (or jobs) was held. By using these month strings, the start and end date of employment (B48 - B55), and by knowing the likely inability of respondents to easily distinguish between being out of the labor force versus unemployed (active search), versus illness (on leave ill from an employer and just not out of the labor force because of illness), we developed software to reconcile most of the cases. We were still left with hundreds of cases needing tender, loving judgment. This hand-editing process is described below.

Question references for head, parallel questions in the D section apply to the wife.

B. Guidelines for Judgmental Hand Editing

1. Month String and Week Gaps

If there was a significant difference in work weeks between month strings and respondent's report of work weeks, we needed to make some reconciliation. If the number of work weeks in the month strings (number of months marked as work multiplied by 4.333) and the number of work weeks reported (B78) significantly differed (in most cases, the number of work weeks from month strings exceeded the number of weeks reported), we looked at work hours per week (B79) in order to obtain a rough idea of the job type. If B79 was smaller than expected in most normal jobs, the job was often judged to be temporary or at least not employment on a full time basis. In this case, the number of work weeks based on the month strings was often judged to exaggerate the actual work weeks. That suggested the adoption of the B78 value as a measure of work weeks. If B79 was approximately equal to what is expected in most normal jobs (30-60 hours per week), the job was judged to be a 'real' one. In this case, the number of work weeks based on the average of month strings and B78 was assumed to reflect actual number of works quite well.

2. Student (BI = 7)

For a student with few reported work weeks, and no report of other weeks of activity, we assumed that annual weeks missed due to the illness of others (B60- 62) = weeks missed due to own illness (B63 -65) = vacation weeks (B66-68) = weeks of strike (B69-71) = weeks of unemployment (B72-B74a) = 0. That is, weeks out of the labor force (not actively looking - (B75 - 77a), and actual work weeks (B78) combine into a 52 week. For a student with considerable work weeks per year, and missing information on the weeks out of the labor force, we assumed that weeks out of the labor force (not actively looking - (B75 - 77a) = 0 (so he is actively in the labor force). Aside from the out of the labor force assumption, for this person, total weeks sum up to a 52 week total just as for others and were apportioned according to the general rules in A above.

3. Retirement (BI = 4)

Unmarked months in the month strings after the job start date (B24 month and year) but before retirement (C5 month and C5 year) were treated as vacation weeks. Unmarked months in the month strings after retirement were treated as out of the labor force weeks. If there is no unmarked months in the month strings before retirement but with vacation reported as positive weeks and no weeks out of the labor force reported, then vacation weeks were assumed 0 and weeks out of the labor force was set to 52- the weeks of work reported in B78.

4. Few Reported Weeks of Work

Consider a respondent who is not a retiree nor a student and with a relatively small number of work weeks and more than 40 weeks or 9 months of time-off in the continuous month strings and for whom one cannot see a reason for this time off. (For example unemployment weeks are reported to = 0 and weeks out of the labor force are reported to = 0.) In this case we simply split the unaccounted weeks equally to weeks out of the labor force and weeks of unemployment.

Note: For those cases in Wife/"Wife's" (D/E) section, we applied the same rules. However, the response indicated a housewife (D1a=6) then we left time-off in the out of the labor force category.

5. Those who did not work during the last year at all

If a respondent did not work at all during the last year, all weeks were assumed to either unemployment or out of the labor force. If the number of unemployment weeks were reported, we assigned the remaining weeks to out of the labor force. In cases where neither unemployment nor out of the labor force was reported, we simply split the entire weeks into weeks out of the labor force (=26) and weeks of unemployment (+26).

C. Hand-editing Guidelines in Comparing Extra and Main Jobs

1. Basic Idea

An extra job should be extra. In other words, when a respondent reports an extra job, this job must be concurrent with a main job (or jobs). If a respondent has an extra job while he does no reported corresponding main job, the extra job was treated as a main job for that time interval. To identify the concurrence, the month strings of extra and main jobs were compared. If there was no concurrence between extra and main jobs, we further checked whether the extra work occurred in the unemployment or out-of-labor force periods. Finally, we edited the number of weeks for

unemployment, out-of-labor force, or both, in some cases. In the process of editing, the original reports by respondents were fully respected. This means that facing suspected cases we first considered many possible situations under which a respondent reported values in such way (For example, see below).

2. Guidelines

- i) The complete overlapping of month strings between an extra job and unemployment (or out of the labor force):

We assumed that a respondent mistakenly took an extra job out of consideration in counting unemployment periods. Although a main job might not have existed during this period, an extra job now should be treated as a main job. That is, the respondent was not unemployed but rather under-employed. We took all these weeks from unemployment and added them to work weeks.

- ii) The partial overlapping of month strings:

In the case of overlapping month strings between an extra job and unemployment (or out of the labor force), we first checked for any reason why a respondent would have reported in this way. In a few cases, the overlapping can be reasonably rationalized. Example 1: A respondent stopped an extra job in the middle of 'x' month and he was in a string of unemployment months since then, and the 'x' month is marked in both the extra job month string and the unemployment month string. Here, there was judged to be no need for editing.

- iii) Continuous month strings for extra job while the same months are marked as unemployment or out of the labor force:

When the extra job month strings are continuous before and after 'x' month(s), the 'x' month, should be thought of as a month in which the respondent worked throughout.

Example: A respondent had an extra job throughout the year (from January through December) and he reported unemployment in June. In fact, the respondent quit the previous job and started the current job in June. In terms of main jobs, he was unemployed in June. However, it is clear that he had an extra job even during that time. Consequently, the respondent should not have counted June as an unemployment month.

- iv) Splitting of month equally between an extra job and unemployment (or out of the labor force):

In the case where the beginning or the end of an extra job month string overlaps those of unemployment (or out of the labor force) month strings, we split the corresponding month(s) equally between an extra job and unemployment (or out of the labor force)

Example: A respondent marked September, October, November, and December as his extra job months. The respondent reported that he worked until August and started a new job in October. Finally, he reported unemployment in September. From this, it seems clear that the respondent quit a job sometime in early September and then started an extra job in the same month. Because it is very difficult to know exactly when he started the extra job, we split September equally between unemployment and extra job. Note that the extra job during the second half of September was not concurrent with a main job. Hence, the extra job during this period should be another main job.

II. Annual Average Hourly Earnings (AAHE) of the Head (and Wife)

In order to get AAHE, first total hours of work should be calculated. Total hours of work is the sum of work hours on a main job(s), on an extra job(s), and overtime. There are some cases where a respondent has positive work hours (B78>0) but no hours per week (B79). Based on the job duration, we looked at work hours per week in the previous year in order to get B79. If the job duration and at work hours per week in the previous year can not support this procedure, we use 35 hours per week as an approximate value of annual average hours per week.

Given the annual work hours as of the head (or wife) constructed above, the next step is to simply divide the annual labor income of the head (or wife) [see the income documentation for details of the construction] by the annual work hours. However, even if the annual labor income value is 'plausible' in its own right and the annual work hours seemed to be 'plausible' in their own right, there is the possibility that the resulting ratio, AAHE, is not plausible. This is part of a longstanding problem with hourly wage measures and has appeared in the literature under the name of 'division' bias when applied to the estimation of labor supply elasticities (Borjas, 1978).

Table 1. Total cases of annual work hours and labor income

Year	Head	Wife
1994	8659	4638
1995	8570	4621
1996	8517	4649
1997	6747	3854
1999	6997	3987
2001	6010	3174

When the resulting AAHE was unusually high (over \$100 per hour) or unusually low (under \$2 per hour) we referred to the reported hourly wage (B16) for those providing an hourly wage on their main job. The latter took precedence over AAHE in these cases. For the remaining 'implausibly' high and low AAHE values, we simply looked at other job features and information to reach a 'judgmental' AAHE.

In some cases, total annual hours of work were positive while labor income was not available. In other cases, the opposite happened. Besides simple misreporting, there can be reasons for these cases. Some labor income arises from a farm or business. Under this situation, it is sometimes possible that hours of work are positive but no labor income generated. Another reason for these cases is the lag of timing between work and pay. One could work in 1993 but not get paid until 1994.

III. Data

There are six data files, one for each of the 1994, 1995, 1996, 1997, 1999 and 2001 Hours of Work and Wage Files. Each file contains information about Total Annual Hours of Work, which is the sum of hours from a main job(s), overtime work, and extra job(s) if any.

The following variables of the Head/Wife appear: Work weeks, Average work hours per week, Overtime, Work hours of extra job(s), Total work hours, Wage rate, Weeks missed due to the illness of others, Weeks missed due to own illness, Vacation weeks, Weeks of strike, Weeks of unemployment, and Weeks out of labor force. The Total labor income variables is located with the other generated income calculated variables.

These 1994 - 2001 Hours of Work and Wage Files contain one record for each family interviewed in 1994 - 2001. For each year, notably numerous in 1994, the file includes a special sample of recontacted respondents, as part of a large methodology study.

The special Latino sample, interviewed in 1994 and 1995, are not included in these files for the corresponding waves. The case count of families in the 1994 Hours of Work is 8659. For 1994 the case count of families that have a non-

zero family panel weight (see the Public Release I weights files for 1994-1996 released 9/98) was 7747. The difference is the consequence of the recontact families. They can be used for some analysis purposes, but simply have a zero family weight. Parallel differences of this sort exist for other years. Users wishing to apply FAMILY WEIGHTS in their analysis will need to visit the weight section of the data center (PSID Data Files, 1993-1999 Public Release I).

The 1997, 1999 and 2001 Public Release I weights are complicated by sample suspension and the addition of a refresher sample of post-1968 immigrants, but they are now available and can be applied to these 1997, 1999 and 2001 family income and hours of work variables.

These files are based on the Public Release I versions of the 1994-2001 waves. The 2001 data, as well as that for 1994-1999, may be subject to relatively minor changes once the Public Release II versions of the 1994, 1995, 1996, 1997, 1999 and 2001 family files become available. The data are in raw ASCII form. Refer to the data definition statements - - SAS and SPSS -- for record format layout information, variable names, variable labels, and missing data codes.

File Attributes and Variables for Data Files			
File Name	Records	LRECL	# of Variables
WRKHRS94.DAT	8,659	109	25
WRKHRS95.DAT	8,570	109	25
WRKHRS96.DAT	8,517	107	25
WRKHRS97.DAT	6,747*	109	25
WRKHRS99.DAT	6,997*	141	25
WRKHRS01	7,406	107	25

*In these two years, the sample of post-1968 immigrants was first added. They are included in the numbers for all subsequent years.

IV. SAS and SPSS Data Definition Statements

These files contain SAS and SPSS data definition statements providing information about the variables in the data files. Two files, one of each type, SAS and SPSS, corresponding to each data file, are provided. The naming conventions are the same as for the data files, e.g., WRKHRS94.SAS contains SAS statements for the 1994 Hours of Work and Wage data file, and WRKHRS.SPS contains SPSS statements for the 1994 Hours of Work and Wage data file. Similar files also exist for 1995 -2001.

The data definition statements provide variable names, variable labels, locations. These processed files have no `missing' data.

The SAS and SPSS data definition statements are NOT intended to represent completed and full programs for the respective statistical program packages to run extracts, analysis, etc. You must provide all other SAS or SPSS statements needed to complete a program. Users wishing to migrate to other formats may use a commercial software for such purposes, such as STAT/TRANSFER or go to the PSID data-center.

V. Documentation

The following machine-readable documentation files are provided for the designated data files.

Title	File Name	Pages
Computerized and Hand-editing Guidelines	README.TXT	7
Codebook for 1994-2001 Hours of Work and Wage	WRKHRS.TXT	(Varies)