

A Guide to Merging Jobs Between the SIPP Core Data and the Retirement & Pension Plan Coverage Topical Module

Jeremy O. Skog*

February 12, 2015

Abstract

This paper describes the algorithm that selects the main job for which detailed questions are asked in the Survey of Income and Program Participation Retirement & Pension Plan Coverage Topical Module. It also describes the appropriate method to match that job to the same job in the Core survey data and provides sample SAS code. Procedures are provided for the 2008 panel and the 2014 panel of the Redesigned SIPP.

1 Description of the SIPP and the Retirement & Pension Plan Coverage Topical Module

The Survey of Income and Program Participation (SIPP) provides longitudinal information on a variety of topics including earnings, employment, assets, program usage, and demographics. The Census Bureau describes¹ the structure of the SIPP as:

The Survey of Income and Program Participation (SIPP) is a household-based survey designed as a continuous series of national panels. Each panel features a nationally representative sample interviewed over a multi-year period lasting approximately four years. SIPP is a source of data for a variety of topics and provides for the integration of information for separate topics to form a single, unified database.

*This guide has been created to inform interested parties of methods to better make use of SIPP data. Any views or opinions expressed in the paper are the author's own and do not necessarily reflect the views or opinions of the U.S. Census Bureau. Please direct correspondence to Jeremy Skog, U.S. Census Bureau, SEHSD, HQ-6H144F, 4600 Silver Hill Rd., Washington, DC 20233, or via e-mail at jeremy.o.skog@census.gov. I would like to thank Cindy Easton, Mark Klee, and Lori Reeder for their helpful comments during the development of this guide.

¹The Census Bureau's description of SIPP is located at: <http://www.census.gov/programs-surveys/sipp/about.html>

The SIPP also provides in-depth information on a variety of subjects through its periodic topical modules. The Retirement & Pension Plan Coverage topical module provides, among other information, detailed data on the offerings and features of retirement plans at each adult respondent's main current job. The SIPP User Guide² describes the content and goal of the Pensions topical module as:

Retirement Expectations and Pension Plan Coverage. Obtains information about the respondent's pension plan coverage for the most important current job or business, and information from persons currently receiving retirement benefits from a former job or business. Respondents are asked about their coverage and vesting in pension plans, types of plans, the reasons they are not included by or do not participate in plans, current contributions and amounts of money in their accounts if applicable, and how the money in their own plans is invested. Other questions concern loans from pension accounts and treatment of lump sums received from prior job pension plans. Respondents currently receiving pension income are asked about the types of pension they receive, provisions for cost-of-living adjustments, and health benefits. Respondents are also asked Industry and Occupation data about the job or business from which their pensions are received.

This topical module was conducted in Waves 3 and 11 of the 2008 panel. Whereas the Core questions in the SIPP survey are designed to gather some information on all sources of the respondent's employment the Pensions topical module, since a redesign in 1996, has been designed to gather in-depth information for only a single source of employment, whether that is a job at which the respondent works or a business that the respondent owns. This raises two issues for researchers:

1. Understanding the procedure through which the main job is chosen.
2. Understanding how to correctly link the job that is the subject of the topical module questions with the equivalent job in the Core survey data.

A special supplemental survey, known as the SSA Supplement was conducted for the 2014 panel and contained information from several previous topical modules. These issues of selecting the appropriate job and merging the Core and Supplement data also exist in the SSA Supplement. Following the 2008 SIPP panel, the SIPP was redesigned as an annual survey that made use of an event history calendar. The Pensions topical module was grouped with other topical modules into a supplementary survey sponsored by the Social Security Administration³. The Supplement covers information on workplace-based retirement

²Page 3-13 & 3-14; A downloadable PDF of the SIPP User Guide is available at: http://beta.census.gov/content/dam/Census/programs-surveys/sipp/methodology/SIPP_USERS_Guide_Third_Edition_2001.pdf

³Information on the SSA Supplement Survey can be found at: <http://www.census.gov/programs-surveys/sipp/about/SSA-Supplement.html>

plans, personal retirement plans, marital history, and health and disability status.

The remainder of this paper addresses these two issues. First, the algorithm that was used to systematically choose the main job is explained. Then, a procedure for linking the main job in the topical module with the equivalent job in the Core survey data is explained. Sample code is provided, to illustrate how a researcher could implement the matching procedure.

2 Procedure for Selecting the “Main Job”

Whereas the Core SIPP questionnaire includes a series of questions for all sources of the respondent’s employment, the pension topical module asks a series of detailed questions about the retirement plans offered on a person’s main job. To ensure consistency, the survey designers created an algorithm to select the respondent’s main job and/or main business. If both exist, the algorithm then selects whether the main source of employment is the main job or main business. This source of employment is then the reference job or reference business for the questions on employment-based retirement programs.

2.1 Algorithm for the 2008 Panel, Waves 3 and 11

In the 2008 panel, the main job was chosen in a manner similar to the choice of the first job in the Core. However, if the first job is not ongoing at the end of the reference period, the second job is selected as the main job if it both exists and is ongoing at the end of the reference period. Therefore, the reference job for the topical module may be the second job if the first job is not current. To select the correct employment, researchers who want to use information from both the topical module and the Core can use the procedure described in this paper to match data from the topical module to the appropriate job in the Core data. This same rule is followed for businesses, so if the first business is ongoing at the end of the reference period, it would be the reference business for the topical module.

2.1.1 Jobs

The main job in the topical module is chosen according to the same method as the main job in the labor force section of the Core survey. Jobs are ordered according to the following algorithm:

1. The survey instrument first attempts to select the longest-held job with the most number of weeks worked during the reference period as the main job. The reference period for the topical module is the same four-month period as the current wave of the survey. This step is imposed by the survey instrument based on previous information that the respondent provided.
2. If the number of weeks between multiple potential jobs was equal, then the job with the highest number of hours worked per week was selected as

the main job. The number of hours is determined by a question asked of the respondent: for which job did they work the most hours. Therefore, this step can be determined by the respondent, separately from other information they provided on the survey.

3. If main job still could not be determined, then the first job listed is designated as the main job.

2.1.2 Businesses

The main business in the topical module is set as the first business listed in the Core survey, unless that business ends during the reference period. If this business has ended, then the second business is set as the reference business for the topical module. The first business in the Core survey is chosen according to these rules:

1. The survey instrument first attempts to select the business at which the respondent worked the most weeks in the reference period.
2. If the respondent worked an equal number of weeks at two or more businesses, the main business is set as the one that had the highest earnings before expenses. This choice is determined by asking the respondent.

2.1.3 Main Source of Employment

If the respondent had both a main job and a main business then the respondent was asked which one was the primary source of work-related income. This question determines whether the reference employment in the topical module was the respondent's job or their business. In the topical module data main source of employment is identified by the RMNJBBS variable.

2.2 Algorithm for the SSA Supplement

The algorithm for choosing the main job or business in the SSA supplement differs from the algorithm used in the topical module. The order of jobs and businesses is not predetermined in the Core. The survey instrument checks all of the jobs reported in the Core survey which are still current as well as any newly-reported jobs to determine which is the main job (or business). Rather than the length of employment during a particular survey-wave reference period, the instrument now checks which job the respondent first began working at, including time outside of the reference period. The goal of the algorithm is not to determine which job provides the most retirement benefits. Rather, it is to provide a consistent method for selecting the reference source of employment across respondents and to minimize the potential for a biased measurement of retirement plan offerings.

2.2.1 Jobs

The main job is chosen from all jobs at which the respondent was working at the time of the interview. This includes jobs reported during the Core survey, as well as new jobs that the respondent began since the Core interview took place. From this set of potential jobs, the main job is chosen according to the following algorithm:

1. The job selected as the main job is the longest-held job which was begun in the earliest year (checked by the survey instrument).
2. If the number of years are equal for two or more of the longest-held jobs, then the main job is one with the highest number of months worked (checked by the survey instrument).
3. If the number of months is equal for two or more of the longest-held jobs, then the main job is one with the highest number of hours worked per week (checked by the survey instrument).
4. If main job still can not be determined by the survey instrument, then the respondent chooses their main job, after being asked which job they consider their main job.

2.2.2 Businesses

The choice of business in the SSA supplement is not determined by the length of time that the respondent worked at the business. Instead, the main business is the one that generates the highest earnings. The algorithm is simply:

1. The main business is the one that had the highest earnings before expenses (as reported by the respondent).

2.2.3 Main Source of Employment

As for the topical module, a respondent who had both a job and a business in the SSA Supplement was asked which one was the primary source of work-related income. This question determines whether the reference employment in the Supplement was the respondent's job or their business. The main source of employment is again identified by the RMNJBBS variable.

3 Variables for Merging

As the logic of choosing the main job indicates, the reference job may not always be in the same "position" in the Core survey. To enable a data-user to match the data from the SSA Supplement or topical module to the Core survey, there are several variables included in the data. In the 2008 panel topical module, the variables RMNJBBS, RTMEENO, and RTMEBNO are used. In the SSA Supplement, a new variable (EMAINJOBID) was created to match the reference

employment in the Supplement to the redesigned method for storing jobs in the Core data. The variables RMNJBBS, RMJB, and RMBS will still identify whether there is a main job or business, but do not directly match the job identifier in the Core data. Although variables with these same names are included in other panels of the survey, including 1996, 2001, and 2004, they require different matching procedures due to peculiarities of the data. The data dictionaries include similar definitions for each variable, but unique circumstances of the data mean that, although the variables are similar in concept in that they are used in merging the data, they are not directly comparable across SIPP panels. Specifically, the 1996 documentation describes RMJB and RMBS in the same way as the 2008 documentation describes RMJB and RMBS. The documentation for the 2001 and 2004 SIPP panels describes RMJB and RMBS with the same wording that the 2008 panel describes RTMEENO and RTMEBNO. For the 2001 panel, the variables RMJB and RMBS can be used in the same way that RTMEENO and RTMEBNO are used to match data the 2008 panel. The Census Bureau has published user notes for the 1996⁴ and 2004⁵ panels describing the peculiarities of the data in these panels and difficulties in matching them to data in the SIPP Core.

4 Matching Data to the Core Survey

To help clarify the process of matching the reference job in the topical module or SSA Supplement to the core I provide example code which a data-user can use to conduct the match. This code is written in the SAS statistical programming language, but the matching logic would be similar in other languages.

4.1 2008 Topical Modules

In the 2008 panel the variables RTMEENO and RTMEBNO replace the job-line variables RMJB and RMBS. These new variables should be used to match the job in the wave 3 or wave 11 topical module to the appropriate job in the Core. Previously, in the 2004, 2001, and 1996 supplements, the RMJB and RMBS variables were used to help identify the correct job line in the core survey, but they are not applicable to the 2008 panel public-use data file.

The code below shows two potential ways to identify the job or business to match and the correct matching variable. The first method simply uses checks and blocks of code and requires some repetition of programming steps. However, it clearly portrays the logic of the matching process, without requiring in-depth knowledge of the SAS language. The second method is more efficient, but makes use of efficiencies that are built-in to SAS and may not exist in the same form in other programming language. Specifically, the code employs SAS macro

⁴See: http://www.census.gov/programs-surveys/sipp/tech-documentation/user-notes/1996_W7_TM_DF_DD.html , Also: http://www.nber.org/sipp/1996/t7_rmnjbbs.sas

⁵See: <http://www.census.gov/programs-surveys/sipp/tech-documentation/user-notes/2004w3CoreLaborForceFeedback.html>

variables which can vary in value. This method is similar to the code written for the proposed method to work with the SSA supplement, where there can be up to seven potential jobs in the Core survey. That is, whereas in the 2008 panel, a block of code to match variables would only need to be written twice, in the SSA Supplement it would need to be replicated seven times. This repetition increases the benefits from employing a more efficient matching procedure.

4.1.1 Code Style 1: Recode blocks

First, I present a matching method that shows the logic behind the merger. This program first identifies if the respondent has a main job or a main business by checking the RMNJBBS variable. (RMNJBBS = 1 if the respondent's job is the main source of employment and RMNJBBS = 2 if the respondent's business is the main source of employment. This variable identifies a main job or main business, even if the respondent has only a job or only a business.) Then, the appropriate "job number" variables are compared to determine which Core job contains the correct number. That is, if the respondent has a main job, then the value of RTMEENO is checked against the value of EENO1 to see if they are the same. If they are, then the first job is the reference job for the topical module. If not, then the value of RTMEENO is checked against EENO2, which should provide a match. Once the correct job has been identified the needed variables can be populated with their values from the corresponding variables in the SIPP Core. A similar procedure takes place if RMNJBBS=2 and the respondent has a main business.

```

if RMNJBBS = 1 then do;
    if (EENO1 = RTMEENO) then do;
        FirmSize = TEMPALL1;
        [etc.]
    end;
    else if (EENO2 = RTMEENO) then do;
        FirmSize = TEMPALL2;
        [etc.]
    end;
end;
else if RMNJBBS = 2 then do;
    if (EBNO1 = RTMEBNO) then do;
        BusSize = TEMPB1;
        [etc.]
    end;
    else if (EBNO2 = RTMEBNO) then do;
        BusSize = TEMPB2;
        [etc.]
    end;
end;
end;

```

4.1.2 Code Style 2: Macros

Using SAS macro variables minimizes the code that a programmer must write and makes it faster to update code while also making it less likely that errors are introduced. The code which matches the respondent's employment using SAS Macros is shown below:

```
%if RMNJBBBS = 1 %then %do
    %if (RIMEENO = EENO1) %then %let jobnum = 1;
    %else %if (RIMEENO = EENO2) %then %let jobnum = 2;
    FirmSize = TEMPALL&jjobnum.;
    [etc.]
%end;
%else %if RMNJBBBS = 2 %then %do;
    %if (RIMEBNO = EBNO1) %then %let busnum = 1;
    %else %if (RIMEBNO = EBNO2) %then %let busnum = 2;
    BusSize = TEMP&busnum.;
    [etc.]
%end;
```

4.2 SSA Supplement

The new SSA Supplement, conducted in 2014, matches the main job in the Supplement to the core using the variable EMAINJOBID. This variable can take the values 101–107. To identify the correct job, the matching number should be found in the Core variable EJB(n)_JOBID, where n identifies one of the seven job variables. Jobs and businesses are not separately identified in the redesigned survey, so only one loop is necessary to identify the correct source of earned income. The code below gives an example of how to loop through these variables to identify the correct job.

```
%do i = 1 %to 7:
    if EMAINJOBID = EJB(&i.)_JOBID then jobnum = &i.;
%end;
[etc.]
```

If EMAINJOBID has a value of 200 then it indicates a new job that it not listed in the Wave 1 survey. It may be possible to match this job to a job that the respondent describes in Wave 2 of the 2014 panel, but this linking is not done automatically and must be done by the researcher.

5 Conclusion

Using the techniques described in this paper, a researcher can make use of the information on employment sources contained within both the Retirement & Pension Plan Coverage Topical Module, the SSA Supplement and the SIPP Core itself. When jobs exist for more than one wave, it is possible to track

employers across survey waves to gain more information by tracking them in the SIPP Core using the employer identifier number. Although this paper does not cover those exact techniques, they are similar to the merging procedures described here.