

# Estimating pension coverage using different data sets

Author: Geoffrey Sanzenbacher

Persistent link: <http://hdl.handle.net/2345/bc-ir:104448>

This work is posted on [eScholarship@BC](#),  
Boston College University Libraries.

---

Chestnut Hill, Mass.: Center for Retirement Research at Boston College, August 2006

These materials are made available for use in research, teaching and private study, pursuant to U.S. Copyright Law. The user must assume full responsibility for any use of the materials, including but not limited to, infringement of copyright and publication rights of reproduced materials. Any materials used for academic research or otherwise should be fully credited with the source. The publisher or original authors may retain copyright to the materials.

# ESTIMATING PENSION COVERAGE USING DIFFERENT DATA SETS

BY GEOFFREY SANZENBACHER\*

---

## Introduction

Employer-provided pensions are an essential piece of the U.S. retirement income system. Calculating the percent of individuals covered by and participating in such plans as well as analyzing the changing nature of those plans is becoming an increasingly important exercise. A variety of data sets are available to estimate pension coverage and participation. Since deciding which data set to use is often not obvious, this *Issue in Brief* will examine five of them: the U.S. Department of Labor Form 5500 series, the Survey of Income and Program Participation (SIPP), the Current Population Survey (CPS), the Panel Study of Income Dynamics (PSID), and the Survey of Consumer Finances (SCF). The *brief* will describe the strengths and limitations of each data set and compare pension participation rates derived from each set.

of participants for two reasons: 1) the participant count may include non-vested employees and 401(k)-eligible employees not actively participating in their plans; and 2) employers must file one form for each plan they offer, which allows for some employees to be counted twice as individuals covered by both a defined benefit and a defined contribution plan.<sup>1</sup> The calculations in this *brief* adjust for these potential problems using methods similar to those employed by the Department of Labor, although some overestimation of coverage may still remain.<sup>2</sup> The 5500 data set does not allow breakdowns of the employee population by demographic characteristics and it only includes private sector employees. However, as data from establishments — which may be more accurate than data from surveys of individuals — the 5500 data set likely yields reliable numbers on the types of pension plans being offered to employees.

## Establishment Data Set: DOL Form 5500

Every year, private pension plan sponsors are required to file with the U.S. Department of Labor a return, the Form 5500 series, containing detailed information about their plans' finances and participants. However, these numbers may overestimate the number

## Household Data Sets: The SIPP, CPS, PSID, and SCF

Another valuable source of information on pension participation comes in the form of household surveys. Surveys are often advantageous because of the wealth of information they provide on the demographic characteristics of the respondents. Thus, while the

---

\* Geoffrey Sanzenbacher is a Graduate Research Assistant at the Center for Retirement Research at Boston College. The author wishes to thank Marric Buessing, Luke Delorme, Francesca Golub-Sass, and Kelly Haverstick for their help in collecting data from the various sources as well as input on content. Greg Wiles' work on this topic two years ago was particularly helpful.

Form 5500 may offer an excellent estimate of the number of people participating in pension plans, it does not offer much insight into these people beyond their participation. Unfortunately, surveys are open to individual error in a way establishment data sets are not, and the extent to which this affects data analysis should always be considered. The four surveys discussed in detail here are the CPS, SIPP, PSID, and SCF, although a wealth of other surveys can provide information on pension-related issues.<sup>3</sup>

### *SIPP*

Since 1984, the U.S. Census Bureau of Demographic Research has conducted the Survey of Income and Program Participation (SIPP). Each year the SIPP interviews a new panel of individuals, asking them a series of core questions every four months for about two and a half years. While questions on pension participation are not asked every four months, they are asked to each individual in the panel at least once as part of the topical module “Retirement Expectations and Pension Plan Coverage.”

An immediate problem with this design is that it does not allow for the derivation of a continuous series on pension participation, but rather a biennial or triennial series beginning in 1985.

Each individual is asked whether he is offered a plan, whether he participates in a plan, and the type of plan for the primary and secondary plans. Since the questions are asked of individuals, responses can be inaccurate, which some studies have shown lead to underestimation of participation in defined contribution plans if not corrected in some way.<sup>4</sup> However, asking the questions of individuals has the advantage of allowing breakdowns by demographic characteristics as well as by sector of employment and size of employer, factors shown to have significant effects on coverage and participation.

### *CPS*

Like the SIPP, the Current Population Survey (CPS) also allows for extensive demographic breakdowns. The Bureau of Labor Statistics has been conducting the CPS on a monthly basis for more than fifty years. The survey provides annual information on labor force participation and earnings. Since 1980, the Annual March Social and Economic Supplement of the CPS has provided data on pension coverage. An obvious advantage of the CPS over the SIPP is that it can be used to derive a continuous yearly series from 1980 on.

However, the CPS does not provide information about the type of pension plan a person is in for all years. Such information is available for the years 1983, 1988, and 1993 from the CPS Employee Benefit Supplement.<sup>5</sup> Relative to the SIPP and Form 5500, the CPS yields somewhat lower estimation of participation in pension plans for the set of all private workers.<sup>6</sup>

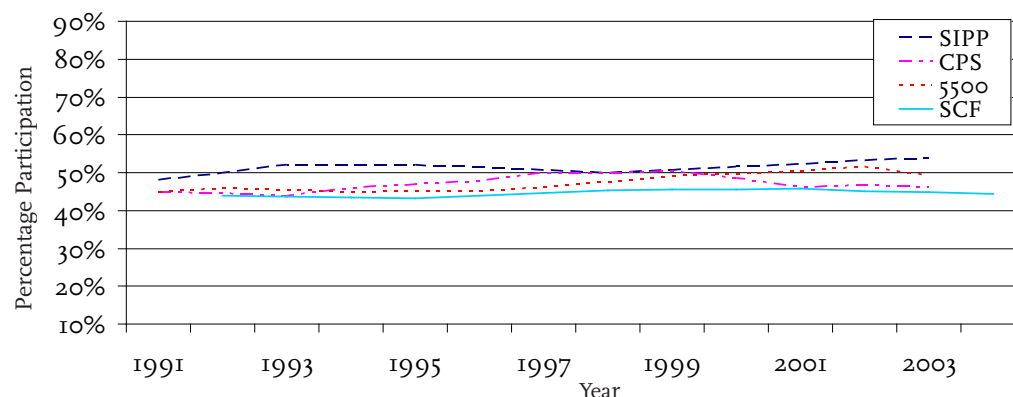
### *PSID*

Another survey that includes information on pension coverage is the Panel Study on Income Dynamics (PSID). The PSID is a longitudinal survey conducted by the Survey Research Center at the Institute for Social Research at the University of Michigan. Beginning in 1969 the PSID has followed the same set of households (with minor exceptions) with low sample attrition rates, making it an excellent source of data for a variety of research issues. However, until 1999, pension data from the PSID were somewhat limited, making the derivation of a consistent measurement of pension participation difficult. In 1999, a new section added to the core questionnaire introduced questions on pension participation as well as the type of pension plan. This new section makes it possible to derive a biennial series on pension participation using a variety of demographic characteristics and the intergenerational information unique to the PSID from 1999 on.

### *SCF*

The final survey considered here is the Survey of Consumer Finances (SCF). The SCF is conducted triennially by the Board of Governors of the Federal Reserve System. The survey is designed to be nationally representative and in 2004 sampled 4,500 households. While these data are generally considered to be of very high quality, one minor drawback is that they do not distinguish between private and public sector employees. Since other data sets show that around 72 percent of public sector workers participate in a pension plan at any given time, many researchers restrict their studies of participation to private sector workers because their participation tends to be less certain. However, the SCF is an excellent source of information on the type of plans people are involved in as well as the levels of investment households have in these plans. Like the other household surveys, the SCF allows breakdowns by demographic characteristics.

FIGURE 1. COMPARISON OF PARTICIPATION RATES FROM DIFFERENT DATASETS, PRIVATE WAGE AND SALARY WORKERS, AGED 25-64, 1991-2004



Sources: U.S. Department of Labor (2004) and author's calculations from U.S. Department of Labor (2001-06); U.S. Bureau of the Census (1992, 1994, 1996, 1999, and 2004); U.S. Bureau of the Census (1992-2004); and Board of Governors of the Federal Reserve System (1994, 1997, 2000, 2003, and 2006).

## Comparison of Participation Rates

Figure 1 compares the four data sets for which an extended series of data could be derived of overall participation rates in pension plans for private wage and salary workers. (The PSID, which has pension data only since 1999, showed levels slightly below that of the SCF for comparable years).

While pension participation has risen and fallen over short intervals, all four data sets show stable participation rates over the period 1991-2004. Each series estimates that at any given time over the past decade 45 to 52 percent of private sector employees are participating in some kind of pension plan. Thus, the choice of data set does not significantly affect the overall levels or long-term trends of the estimated pension participation rate. Of course, there are differences in short-term trends between the various data sets as well as slight differences in the levels suggested by each set.

The SIPP and SCF appear to provide an upper and lower bound on the estimation of pension coverage.<sup>7</sup> In between these two series are the Form 5500 and CPS, which show similar levels and trends, with one notable exception. For the period from 1999 to 2002, the Form 5500 series trends upward, with the participation rate increasing from 49 percent to 52 percent, while the CPS rate trends downward,

decreasing from 51 percent to 47 percent. The SIPP and SCF do not offer any clarity to the issue, as from 1998 to 2003 the SIPP estimates show an increase in participation among the sample, from 50 percent to 54 percent, while the SCF shows a slight decrease from 45 percent to 44 percent. Thus, it is not immediately clear from the data whether pension coverage has increased or decreased in recent years. This issue is obviously important not only for researchers but also for policymakers interested in promoting more widespread pension coverage.

## Conclusion

The choice of a data set is a difficult, but important, task. While establishment data sets are certainly reliable, they lack the ability to identify individuals and determine which of their characteristics led them to participate in a pension plan. Survey data sets on the other hand easily allow for such distinctions, but are more susceptible to inconsistent responses. Furthermore, all five data sets yield slightly different estimates of pension participation. However, by recognizing these differences and understanding why they occur and their effects on estimation, a researcher can choose the most appropriate data set for his project.

---

# APPENDICES

---

## Appendix A: Variables and Definitions Used to Derive Participation Rates

### SIPP

In the SIPP a person is said to be participating in a pension plan if he says his employer offers some sort of pension plan and that he participates in that plan. Individuals who say they were not offered a pension plan but also say they were offered a tax-deferred plan are said to be participating in a plan if they said they participated in that tax-deferred plan. The variables needed to derive participation by the above definition can be obtained from the topical module of the SIPP entitled “Retirement Expectations and Pension Plans.” The series used here is derived from wave 4 1990, wave 4 1992, wave 9 1993, wave 7 1996, and wave 7 2001. The data are said to represent participation for the year the interview took place so that, for example, wave 7 2001 was used to derive the 2003 number. The variables used can be seen in Table A1. Although there was a fundamental change in how variables were labeled between 1993 and 1996, the underlying variables needed to measure participation remained unchanged. In an attempt to make the data comparable with the other sources described here, only private wage and salary workers between the ages of 25 and 64 who worked for others are included in the sample. All calculations are weighted using the final person weight provided by the SIPP.

### CPS

In the CPS a person is said to participate in a plan if he answered “yes” to the question regarding inclusion in a pension plan at work. This question has been available in the March Supplement to the CPS every year since 1980. The variable name for the inclusion variable is *pensincl* and it has not changed since 1988. Again, only private wage and salary workers aged 25-64 who worked for others are included in the sample.

### PSID

In the PSID a person is said to participate in a pension plan if his employer offers a defined benefit plan or if his employer offers a defined contribution plan and either the worker contributes to that plan or the employer contributes to that plan. If a person was identified as a head or wife, the person’s participation was determined from variables describing the head or wife respectively. The variables needed to derive participation were available in 1999, 2001, and 2003 from Section P of the core questionnaire. It should be noted that Section P was available in previous years but covered subjects other than pension coverage. The variables used can be seen in Table A2. The sample is limited to only include heads and wives between the ages of 25 and 64 working for private employers other than themselves.

TABLE A1. VARIABLES USED TO DERIVE PARTICIPATION FROM SIPP

Label	W4 1990	W4 1992	W9 1993	W7 1996	W7 2001
Offered Pension Plan	TM8342	TM8342	TM6018	EPENSNYN	EPENSNYN
Participated in this Plan	TM8346	TM8346	TM6026	EINCPENS	EINCPENS
Availabilty Tax Deferred Plan	TM8438	TM8434	TM6110	E3TAXDEF	E3TAXDEF
Participated in Tax Deferred Plan	TM8442	TM8436	TM6114	E3PARTIC	E3PARTIC

Source: Author’s documentation from U.S. Bureau of the Census (1992, 1994, 1995, 1998, 2003).

TABLE A2. VARIABLES USED TO DERIVE PARTICIPATION FROM PSID

Label	Name by Year		
	1999	2001	2003
WTR Pension at Current Job (Head)	ER15156	ER19327	ER22722
Type of Pension at Current Job (Head)	ER15175	ER19343	ER22738
If DC Does Individual Contribute (Head)	ER15166	ER19334	ER22729
If DC Does Employer Contribute (Head)	ER15176	ER19344	ER22739
WTR Pension at Current Job (Wife)	ER15302	ER19470	ER22866
Type of Pension at Current Job (Wife)	ER15321	ER19486	ER22882
If DC Does Individual Contribute (Wife)	ER15312	ER19477	ER22873
If DC Does Employer Contribute (Wife)	ER15322	ER19487	ER22883

Source: Author's documentation from University of Michigan (2000, 2002, and 2004).

## SCF

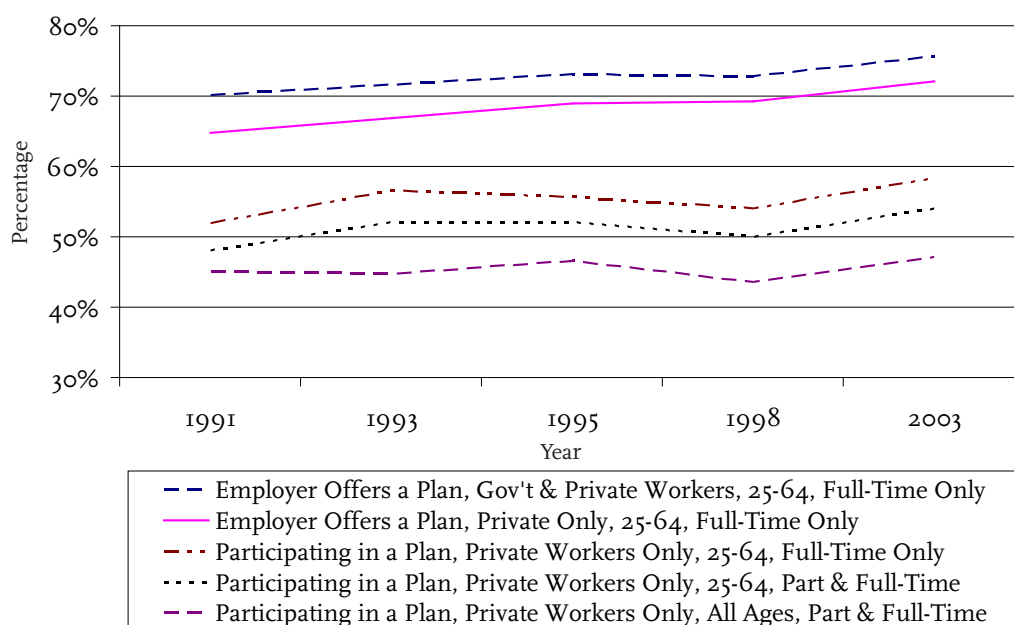
In the SCF an individual is considered to be participating in a plan if he said he is actively participating in a defined benefit plan, a defined contribution plan, or both. The variables used for this derivation were x4203, x4303, x4403, x4803, x4903, and x5003. These variables are the type of plan participated in for the first, second, and third jobs of the head and the plan participated in for the first, second, and third jobs of the spouse respectively.

## Appendix B: The Importance of Definitions

Munnell, Sundén, and Lidstone (2001) noted the importance of definitions in determining participation rates using the Current Population Survey. Their paper contains a figure showing how the levels of estimated pension coverage change for different definitions of the sub-sample and of coverage. Figure A1 includes similar information derived from the SIPP. First, the figure shows that by loosely defining “coverage” as an individual working in either the private or public sector having been offered a plan, the data show that the coverage rate ranges from 70 to 75 percent for the aggregate work force. Limiting the sample to private workers, the rate becomes 65 to 72 percent.

If instead, using the more common definition of “coverage” as an individual participating in some way (in this case as defined above for the SIPP) the numbers are significantly lower. This result occurs because, over the time period of this figure, SIPP data indicate that between 20 to 25 percent of individuals offered a plan did not participate. Furthermore, the inclusion of young workers or part-time workers decreases the participation number further. Thus, the researcher should consider how his choice of sample will influence the numbers produced.

FIGURE A1. LEVELS OF PARTICIPATION FOR VARIOUS DEFINITIONS FROM THE SIPP, 1991-2003



Source: Author's calculations from U.S. Bureau of the Census (1992, 1994, 1996, 1999, and 2004).



## Endnotes

1 For example, a firm that offers both a defined benefit and defined contribution plan to 1,000 of its employees could generate a participant count of 2,000, even if the firm has only 1,000 workers.

2 See the technical appendix in Buessing and Soto (2006).

3 For example, the Health and Retirement Study is not discussed in detail here because it focuses on a specific age cohort that is not conducive to calculating aggregate participation numbers. An excellent source of information on all relevant survey data sets is provided by Citro and Hanushek (1997).

4 For example, Turner, Muller, and Verma (2003) found that, by using SIPP data only, the estimated defined contribution participation rate was 19.7 percent, while correcting inconsistencies by linking the SIPP and detailed earnings reports yielded an estimated participation rate of 30 percent. In this brief, SIPP has not been used to derive what type of plans people are participating in. Purcell (2001/2002, 2005) and Copeland (2002) have done so for various years and provided clear explanations of their methods.

5 This supplement was discontinued after 1993.

6 The difference in the levels of the surveys is somewhat puzzling. Since the take-up rates between the surveys are similar, the difference in levels of participation must be coming from a difference in how the surveys measure eligibility. One possible source of this difference could be that the SCF and CPS interview one reference person and ask that person questions about all members of the household, while the SIPP attempts to interview each person in the household. For example, in 1991 60.5 percent of people actually interviewed who were not the head of household or their spouse said they were offered a pension plan while 55.6 percent of the same group interviewed by a proxy said they were offered a plan. The SCF and CPS would probably interview this entire group in proxy, while the SIPP only interviews 40 percent in proxy. Other issues could be differences in question wording and differences in the line of questioning. Finally, Iams (1995) noted that the difference between the CPS and the SIPP becomes smaller as the analysis is restricted to include only full-time workers and only workers aged 25-64.

7 Where possible, the same groups of people were examined for each data set. For the Survey of Consumer Finances, SIPP data were used to estimate the percent of the SCF sample that were public sector workers. This information was then used with the estimated participation rate of public sector workers from the SIPP to estimate the SCF series on private sector pension participation. Also, while the set of 18-24 year olds is in the Form 5500 series, it is possible that this offsets the double counting issue as younger workers are less likely to participate in pension plans.

## References

- Board of Governors of the Federal Reserve System. 1994, 1997, 2000, 2003, 2006. *Survey of Consumer Finances*. [Available at: <http://www.federalreserve.gov/Pubs/oss/oss2/scfindex.html>].
- Buessing, Marris and Mauricio Soto. 2006. "The State of Private Pension Plans: Current 5500 Data." *Issue in Brief 42*. Chestnut Hill, MA: Center for Retirement Research at Boston College.
- Citro, Constance F. and Eric A. Hanushek, Eds. 1997. *Assessing Policies for Retirement Income. Needs for Data, Research, and Models*. Washington DC: National Research Council. National Academy Press.
- Copeland, Craig. 2002. "An Analysis of the Retirement and Pension Plan Coverage Topical Module of SIPP." *Issue Brief 245*. Washington, DC: Employee Benefit Research Institute.
- Iams, Howard M. 1995. "The 1993 SIPP and CPS Pension Surveys." *Social Security Bulletin*: 58 (4), 125-130.
- University of Michigan, Institute for Social Research. 2000, 2002, 2004. *Panel Study on Income Dynamics*. [Available at: <http://psidonline.isr.umich.edu>].
- Munnell, Alicia H., Annika Sundén, and Elizabeth Lidstone. 2001. "Private Pensions: Coverage and Benefit Trends." Prepared for Conversation on Coverage, Washington, DC.
- Purcell, Patrick, J. 2005. "Retirement Plan Participation and Contributions: Trends from 1998 to 2003." *CRS Report for Congress, Order Code RL33116*. Washington, DC: Library of Congress, Congressional Research Service.
- Purcell, Patrick, J. 2001/2002. "Pension Sponsorship and Participation: Trends and Policy Issues." *Social Security Bulletin* 64(2): 92-102.
- Turner, John, Leslie Muller, and Satyendra K. Verma. 2003. "Defining Participation in Defined Contribution Pension Plans." *Monthly Labor Review*: 126(8): 36-43.
- U.S. Bureau of the Census. 1992-2004. *Current Population Survey*. [Available at: <http://www.bls.census.gov/cps/cpsmain.htm>].
- U.S. Bureau of the Census. 1992, 1994, 1996, 1999, 2004. *Survey of Income and Program Participation*. [Available at: <http://www.sipp.census.gov/sipp>].
- U.S. Department of Labor, Employee Benefits Security Administration, Office of Participant Assistance. 2001-2006. *Annual Return/Report Form 5500 Series for Plan Year 1999-2004*. Washington, DC: U.S. Government Printing Office.
- U.S. Department of Labor, Pension and Welfare Benefits Administration, Office of Policy and Research. 2004. "Abstract of 1999 Form 5500 Annual Reports," *Private Pension Plan Bulletin* (12). [Available at" <http://www.dol.gov/ebsa/PDF/1999pensionplanbulletin.PDF>].

---

## About the Center

The Center for Retirement Research at Boston College was established in 1998 through a grant from the Social Security Administration. The Center's mission is to produce first-class research and forge a strong link between the academic community and decision makers in the public and private sectors around an issue of critical importance to the nation's future.

To achieve this mission, the Center sponsors a wide variety of research projects, transmits new findings to a broad audience, trains new scholars, and broadens access to valuable data sources. Since its inception, the Center has established a reputation as an authoritative source of information on all major aspects of the retirement income debate.

## Affiliated Institutions

American Enterprise Institute  
The Brookings Institution  
Center for Strategic and International Studies  
Massachusetts Institute of Technology  
Syracuse University  
Urban Institute

## Contact Information

Center for Retirement Research  
Boston College  
Fulton Hall 550  
Chestnut Hill, MA 02467-3808  
Phone: (617) 552-1762  
Fax: (617) 552-0191  
E-mail: [crr@bc.edu](mailto:crr@bc.edu)  
Website: <http://www.bc.edu/crr>

*The Center for Retirement Research thanks AARP, AIM Investments, AXA Financial, Citi Street, Fidelity Investments, John Hancock, Nationwide Mutual Insurance Company, Prudential Financial, Standard & Poor's and TIAA-CREF Institute for support of this project.*