Health and Retirement Study

Imputations for Pension Wealth
Final Version 2.0
December 2006

Data Description and Usage

1. Overview and Background

The *Imputations for Pension Wealth* (Version 2.0) data release consists of information derived from the Health and Retirement Study (HRS), a national longitudinal study of the economic, health, marital, family status, and public and private support systems of older Americans. The National Institute on Aging provided funding (NIH U01 AGO9740), with supplemental support from the Social Security Administration. The Institute for Social Research (ISR) Survey Research Center (SRC) at the University of Michigan conducted the survey.

By receiving these data, which have been freely provided, you agree to use them for research and statistical purposes only and make no effort to identify the respondents. In addition, you agree to send HRS a copy of any publications you produce based on the data. See Obtaining the Data for additional details.

This file contains estimates for pension wealth data from current jobs based on both the self-reported and the employer data. The pension estimation software estimates the value of Defined Benefit (DB) entitlements very well. However, not all the information required for calculating Defined Contribution (DC) values is available[1]. Moreover, the amount in the DC account is a simple piece of information that may be known to the respondents at the time of the interview. In contrast, a special calculation is required of most respondents in order to report expected DB benefits at the date of expected retirement. Accordingly, DB values are calculated or imputed from the 1993 and 1999[2] employer data which were obtained based on employer contact information provided by the respondent in the 1992 and 1998 core surveys. DC value imputations are based on the self-reported data.

> **What's new**: The *Imputations for Pension Wealth* (Version 2.0) data release is based on information derived from employer-provided summary plan descriptions (SPDs). In 1993 and 1999 SPDs were obtained by directly contacting the employers of HRS respondents and requesting the documents. To increase the number of respondents with matched employer-provided SPDs, employers who promised to send SPDs in 1999 and never did so were re-contacted in 2004 with a supplemental "callback" survey. Alternative means of obtaining SPDs such as Internet searches and use of commercial databases were also exploited.
>
> The issue of respondent-employer matches was revisited and amendments were made, resulting in a better quality of matches as well as an increased number of matched cases. Specifically, in version 1, employer-provided SPDs (and hence the calculated DB wealth) were missing for 70% of respondents who reported at least one DB pension plan in the1998 HRS survey. In version 2, 59% of respondents are missing calculated DB wealth and required imputations. A final change is that

---

[1] To estimate the value of DC accounts one requires the history of employer and employee contributions as well as the history of portfolio compensation and the returns on investments.

[2] The 1999 employer data will be supplemented by the Form 5500 attachment obtained from the Department of Labor.

calculated DB wealth is now obtained by using an enhanced release of the *Pension Estimation Program* which incorporates extra features as well as error fixes

2. Procedures

2a. Introduction

For this version of pension wealth imputations, plan type is determined by the respondent's report. If the respondent reported a DC plan, either the account balance will be the amount reported by the respondent or the imputed account balance if the account balance is missing. If the respondent reported a DB plan, the plan value is either calculated from the employer provided SPD or it is imputed using DB values from other respondents if there was no matched plan. DB values are calculated or imputed for 1992 and 1998; DC values are reported or imputed for each wave.

DB values are calculated and imputed for each respondent at six different ages. Those include the expected age of retirement reported by respondents with DB or combination plans, early age of retirement, normal age of retirement, ages 60, 62, and 65. The calculations use the intermediate future interest rate, the wage growth rate, and the inflation rate forecast by the Social Security Administration.

2b. Imputation Methods

We used three methods of imputation depending on the sample size. These methods include mixed, hot-decking, and replacement. In the mixed method, we use a regression to get the predicted value of the variable to be imputed. Then the data are sorted by the predicted variable. At the end, the missing/DK/RF value is replaced with the observed value of its preceding adjacent observation. This method is used when the sample size is large enough to run a regression with firm size, industry, union, and personal characteristics as explanatory variables to estimate the predicted values.

> **What's new**: In version 2, we added several person characteristics that increased the predictive power of regressions used in imputations. For example, respondent's industry was classified as public or private based on the number of matched public or private SPD for that industry.

In the hot-decking method, we generate a set of random numbers and sort the data by the random values. Then the missing/DK/RF value was replaced with the observed value of its preceding adjacent observation. A variable such as the expected age of retirement for the second and third most important plans is imputed by this method.

The third method, the replacement method, is used when the sample size is too small for either the mixed or hot-decking methods or by the design of the instrument. For example, for the account balances of the fourth plan, if there are only two observations in the sample and one of them had a missing value, the missing/DK/RF value is replaced with the observed value in the sample. The expected age of receiving benefits is also replaced with the actual start date of

pension benefits (reported in the Income section) when a respondent reported already receiving benefits in response to the question of the age he/s expects to start receiving benefits.

3 Distribution Information

3a. Distribution Set

The Pension Wealth Imputation files are packaged for distribution in a .ZIP file, *ImpPenW.zip*, which contains seven files:

- *ipw.da*   ASCII data file
- *ipw.sas*   SAS data descriptors
- *ipw.sps*   SPSS data descriptors
- *ipw.dct*   Stata data descriptors
- *ipw.do*   Stata "do" file
- *ipw.txt*   codebook information
- *ImpPenW.pdf*   data description (this file)

3b. Data File

The data file is a respondent level file containing one record for every respondent in the HRS and War Baby cohorts who participated in the core survey in 1992 to 2000. The file includes 15879 observations and 188 variables. Identification variables -- HHID, PN, ASUBHH, CSUBHH, ESUBHH, FSUBHH, and GSUBHH – are provided for merges with other HRS data files. The file contains imputed DB values for the 1992 and 1998 survey years, DC account balances for 1992 through 2000, and information necessary for calculation of pension wealth.

3b1. Imputed DB values for 1992 and 1998 survey years

DB values are calculated for all respondents who reported a DB/Combination plan in 1992 and/or 1998 survey year(s). The values are calculated for six alternative ages: expected retirement age reported by respondents with DB or combination plan(s), early and normal retirement ages indicated in the plan formula, and ages 60, 62, and 65. The data are from employer survey data collected in 1993 and 1999.

For respondents with a matched employer-provided DB plan, the DB values are estimated using the formula obtained from the pension providers and the work and earning histories (current earnings projected backward and forward) reported by the respondents who are covered by each plan. Respondents without a matched plan have imputed DB values.

The imputations are performed for all six alternative ages in 1992 and 1998 survey data. Imputations were performed using a nearest-neighbor form of hot deck. Regression models were estimated for the cases with matched data, and used to predict predicted values for matched cases and for unmatched cases. Each unmatched case was paired with a matched case with the nearest predicted value, and the actual value for the matched case was assigned to the unmatched case. Each of the six versions was imputed independently with no constraint on the relationship among the different versions.

3b1-1. Naming Conventions for DB Variables

Values	DB60_98x	Imputed DB values at age 60
	DB62_98x	Imputed DB values at age 62
	DB65_98x	Imputed DB values at age 65
	DBer_98x	Imputed DB values at early retirement age
	DBnr_98x	Imputed DB values at normal retirement age
	DBxp_98x	Imputed DB values at expected retirement age
Imputation flag	DB60_98t	Flag indicating that DB value at age 60 is imputed
	DB62_98t	Flag indicating that DB value at age 62 is imputed
	DB65_98t	Flag indicating that DB value at age 65 is imputed
	DBer_98t	Flag indicating that DB value at early retirement age is imputed
	DBnr_98t	Flag indicating that DB value at normal retirement age is imputed
	DBxp_98t	Flag indicating that DB value at expected retirement age is imputed

Note that these values represent the present discounted value of future benefits at the indicated target age, assuming the person works up to that target age. They are not discounted to the actual survey year, and they include future accruals.

As a rough approximation to the present value at the survey year without future accruals, in order to allow comparison between DB and DC amounts, additional measures have been constructed in which the value at the expected age of retirement is prorated and discounted to the survey year (1992 and 1998). The values are prorated based on respondents' work to date (1992 or 1998) as a share of work from date of hire until the expected age of retirement. They are then discounted from the expected age at retirement to the survey date by a nominal interest rate of 5.8%

> PVxp_92 – DB value at expected retirement age prorated and discounted to 1992
> PVxp_98 – DB value at expected retirement age prorated and discounted to 1998

3b2. DC Account Balances for 1992 to 2000 Surveys

These variables include the observed and imputed values if missing. The values are those reported by respondents with a DC (s) and/or combination plans in each survey year. The DC amounts are account balances as of 1992, 1994, 1996, 1998, and 2000. They do not include future contributions.

3b2-1. Naming Conventions for DC Variables

In each survey year, there are 13 questions about DC account balances. The imputations are performed for each account separately. In the imputation file, the names of the original HRS variables are preserved (with the letter "X" added to the variable name to indicate that the missing values are imputed). For example, variables *F3365_1X, F3365_2X and F3365_3X* indicate amounts in DC accounts reported by respondents in 1998 survey with missing values imputed. The variables *F3365_1T, F3365_2T and F3365_3T* indicate type of imputations performed for the missing values.

Summary variables that indicate the values from all DC or combination account reported are included as well. The naming conventions for these variables are:

DC_92	Reported and imputed DC values from all DC and/or combination accounts from respondent's present job in 1992.
DC_94	Reported and imputed DC values from all DC and/or combination accounts from respondent's present job in 1994
DC_96	Reported and imputed DC values from all DC and/or combination accounts from respondent's present job in 1996.
DC_98	Reported and imputed DC values from all DC and/or combination accounts from respondent's present job in 1998.
DC_00	Reported and imputed DC values from all DC and/or combination accounts from respondent's present job in 2000

3bc. Calculating Total Pension Wealth

In order to compute total pension wealth, a user needs to add DB values at expected age of retirement, prorated and discounted to 1992 or 1998 (variables PVxp_92 and PVxp_98 respectively) and the values of all DC/Combination accounts for the corresponding year (variables DC_92 and DC_98)

3c. Directory Structure

Although a particular setup is not required for using HRS files, we have traditionally suggested a directory structure. By using this directory structure, you will not have to change the path name in your data descriptor files. If you use a different structure, just change the directory references in the program statement files.

```
Directory              Contents
==================     ==================================
c:\ipenw               File downloaded from Web site
c:\ipenw\codebook      codebook and documentation files
c:\ipenw\data          data files
c:\ipenw\sas           SAS command files
c:\ipenw\spss          SPSS command files
c:\ipenw\stata         Stata command files
```

Decompress the .zip file into the appropriate subdirectories. You will need approximately 15 MB of free space on your storage device to store these files.

4. Identification Variables

Identification variables for the Employer Pension Tracker File are stored in character format. The file contains HHID, PN, and xSUBHH (ASUBHH, CSUBHH, ESUBHH, FSUBHH, and GSUBHH) identifiers, which can be used for merging with other HRS data files.

4a. HHID – Household Identifier

In the initial wave of data collection (in 1992 for the HRS sub-sample, in 1993 for the AHEAD sub-sample, and in 1998 for the WB and CODA sub-samples) each sample household was assigned a Household Identifier. HHID is stable across waves of data collection and uniquely identifies the original household and any households derived from that household in subsequent waves of data collection. HHID has six digits.

4b. PN – Person Number

In combination with HHID, PN uniquely identifies a respondent or respondent's spouse or partner. PNs are unique within an original household (HHID). The PN assigned to a particular respondent does not change across waves. PN has three digits.

4c. xSUBHH – Sub-household Identifier

In combination with HHID, xSUBHH uniquely identifies a household for a given wage. Sub-household identifiers can be different at each wave. xSUBHH has one digit..

5. Program Statements

The Employer Pension Tracker File comes with associated SPSS, SAS, or Stata program statements to read the data. Files containing SPSS statements are named with a .SPS extension, those with SAS statements with a .SAS extension, and those with Stata statements with .DO and .DCT extensions.

5a. Using the Files with SAS

To create a SAS system file for a particular data set, two file types must be present for that data set -- .SAS program statement files and .DA data files. To create a SAS system file, load the *.SAS file into the SAS Program Editor.

If the *.SAS file is located in "c:\ipenw\sas" and the data file is located in "c:\ipenw\data", you can run the file as is. A SAS system file (*.SD2 or *.SAS7BDAT) will be saved to directory "c:\ipenw\sas".

If the files are not located in the specified directories, you will need to edit the *.SAS file to reflect the proper path names prior to running the file.

5b. Using the Files with SPSS

To create an SPSS system file for a particular data set, two file types must be present for that data set -- .SPS program statement files and .DA data files. To create an SPSS system file, you must first open the *.SPS file in SPSS as an SPSS Syntax File.

If the *.SPS file is located in "c:\ipenw\spss" and the data file is located in "c:\ipenw\data", you can run the file as is. An SPSS system file (*.SAV) will be saved to directory "c:\ipenw\spss".

If the files are not located in the specified directories, you will need to edit the *.SPS file to reflect the proper path names prior to running the file.

5c. Using the Files with Stata

To use Stata with a particular data set, the following three file types must be present for that data set -- .DCT files, .DO files, and .DA data files.

Files with the suffix .DA contain the raw data for Stata to read. Files with the suffix .DCT are Stata dictionaries used by Stata to describe the data. Files with the suffix .DO are short Stata

programs ("do files") which you may use to read in the data. Load the .DO file into Stata and then submit it.

If the *.DO and *.DCT files are located in "c:\ipenw\stata" and the data file is located in "c:\ipenw\data", you can run the .DO file as is.

If the files are not located in these directories, you must edit the *.DO and *.DCT files to reflect the proper path names before you run the files.

Note that the variable names provided in the .DCT files are uppercase. If you prefer lowercase variable names, you may wish to convert the .DCT files to lowercase prior to use. You may do this by reading the .DCT file into a text or word processing program and changing the case. For instance in Microsoft Word, Edit, Select All, Format, Change Case, lowercase.

6. Obtaining the Data

6a. Registration and Downloading the Data

HRS data are available for free to researchers and analysts at the HRS Web site. In order to obtain public release data, you must first register at our Web site. Once you have completed the registration process, your username and password will be sent to you via e-mail. Your username and password are required to download any data files. By registering all users, we are able to document for our sponsors the size and diversity of our user community allowing us to continue to collect these important data. Registered users receive user support, information related to errors in the data, future releases, workshops, and publication lists. The information you provide will not be used for any commercial use, and will not be redistributed to third parties.

6b. Conditions of Use

By registering, you agree to the Conditions of Use governing access to Health and Retirement public release data. You must agree:
- not to attempt to identify respondents
- not to transfer data to third parties except as specified
- not to share your username and password
- to include specified citations in work based on HRS data
- to provide information to us about publications based on HRS data
- to report apparent errors in the HRS data or documentation files
- to notify us (via our Web site) of changes in your contact information

For more information concerning privacy issues and conditions of use, please read *Conditions of Use for Public Data Files* and *Privacy and Security Notice* in the Public File Download Area of the HRS Web site.

6c. Publications Based on Data

As part of the data registration process, you agree to include specified citations and to inform HRS of any papers, publications, or presentations based on HRS data. Please send a copy of any

publications you produce based on HRS data, with a bibliographical reference, if appropriate, to the address below.

> Health and Retirement Study
> Attn: Papers and Publications
> The Institute for Social Research, Room 3050
> P.O. Box 1248
> Ann Arbor, MI (USA) 48106-1248

Alternately, you may contact us by e-mail at hrsquest@isr.umich.edu with "Attn: Papers and Publications" in the subject line.

7. If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines for using the *Imputations for Pension Wealth* data release. If you have questions or concerns that are not adequately covered here or on our Web site, or if you have any comments, please contact us. We will do our best to provide answers.

7a. HRS Internet Site

Health and Retirement Study public release data and additional information about the study are available on the Internet. To access the data and other relevant information, point your Web browser to the HRS Web site: http://hrsonline.isr.umich.edu/

7b. Contact Information

If you need to contact us, you may do so by one of the methods listed below.

> Internet: Help Desk at our Web site
>
> E-mail: hrsquest@isr.umich.edu
>
> Postal service:
> Health and Retirement Study
> The Institute for Social Research, Room 3050
> The University of Michigan
> P.O. Box 1248
> Ann Arbor, MI 48106-1248
>
> FAX: (734) 647-1186