HRS 1998 Preliminary Imputations

Honggao Cao

Institute for Social Research
University of Michigan

July 2000

CONTENTS

HRS98 Preliminary Imputations

I. Introduction

The Health and Retirement Study (HRS) is a national longitudinal study that describes for older Americans the dynamics of economic, health, marital, and family statuses, as well as public and private support systems. A companion study to the Asset and Health Dynamics among the Oldest Old (AHEAD), and administered at the Institute for Social Research (ISR), the University of Michigan, HRS uses an innovative method to collect data that are usually subject to high rates of non-response (Heeringa, Hill, and Howell, 1998; Hill, 1999; Juster and Smith, 1994). A key feature of the method is to allow a respondent who is unwilling or unable to answer an "amount" question to provide "bracket" information about the "amount". To utilize the "bracket" information collected, one needs to unfold brackets, or impute the amount variables with brackets.

In this document we describe imputation files for HRS 1998 preliminary release data. For simplicity, we shall call these files "HRS98 preliminary imputations". An imputation program developed at the ISR, IMPUTE,[1] was used to produce the results, which include variables for imputed values, imputation controls, bracket patterns, and imputation types. Naturally, the focus is on the imputations of *the amount variables with brackets;*[2] examples of these variables are given in Section III. But some of the variables without brackets are also covered if they are believed to be integral components of three aggregate variables interested by many researchers:

---

[1] A SAS application, the program IMPUTE, along with a documentation file, is available upon request. Send request to: hrsquest@isr.umich.edu

[2] "Amount variable" is a term used in the imputation program, IMPUTE, to indicate a variable for which imputations are done. For the definition of "amount variable", as well as other important terminologies used in the program (and hence in this file), see Section IV or consult Cao (2000).

household income, housing equity, and net worth. The rules for selecting the amount variables for imputations are stated formally in Section II, while the strategies for producing and organizing imputation results are discussed in Section IV.

In Section V, we list all of the amount variables included in the imputation files, with comments on the specific treatment to the imputations for each of the variables. The core of the document, this section may be viewed as a supplement to--but not a substitute for--the codebook of "HRS98 preliminary imputations", which is released simultaneously with this document.

Finally in Section VI, we list several important questions that we believe users may have regarding the imputations. A brief answer to each of the questions is provided.

Readers who are not interested in our imputation strategies may skip the rest of the paper. It is recommended, however, that users of the data consult Section IV, "Producing and Organizing Imputation Results", and Section V, "Questions and Answers", whenever encountering an imputation-related problem.

II. Rules for Variable Selections

For the current version of the imputations, we limit our attention to the variables in the HRS98 sections F (housing) and J (income and assets). Even in the two sections, not all of the variables have been selected. For some, imputations may easily be done by researchers themselves, tailoring to their own needs. Most of those variables do not have brackets. For other variables, there are too few valid records to make any imputations meaningful—even though they do have brackets. To be specific, the following rules were used to determine if an HRS98 variable should be imputed and included in the imputation files ---

```
1. The Bracket Rule: In general, all the variables with brackets
should be imputed, and included in the imputation file;

2. The Income/Asset Component Rule: A variable without bracket
should be imputed if it is determined to be an integral component
of household income, housing equity, or net wealth;

3. The Imputability Rule: A variable eligible for imputation based
on Rule 1 or 2 will not be imputed if there are no enough valid
observations.
```

III. The Bracketed and Unbracketd Problems: Examples

An imputation problem encountered in HRS98 may be categorized as one of the two types: bracketed and unbracketed. An unbracketed problem, consisting only of "ownership" and "amount" questions, has a familiar structure (Box 2). The ownership question asks a respondent if he or she "owns" the subject of interest (receives welfare income, pays medical expenditure,

Box 2. The Unbracketed Problem: an Example

```
(Ownership Question)
O1. Did you (or your husband/or your wife/or your partner) receive
any income from welfare in (1997/1998)?
     1.      Yes
     5.      No
     7.      Other
     8.      DK (don't know); NA (not ascertain)
     9.      RF (refused)
     Blank   INAP (Inapplicable)


(If "Yes" to the Ownership Question, then)
(Amount Question)
A1. How much did (you/he/she) receive in (1997/1998)?
     1-999996.   Actual value
       999997.   Other
       999998.   DK (don't know); NA (not ascertained)
       999999.   RF (refused)
        Blank.   INAP (Inapplicable)
```

etc). If the answer is "yes", then the amount question asks for the amount that he or she "owns". When a respondent gives a positive answer to the ownership question but an uncertain answer (e.g., "don't know" or "refused") to the amount question, however, an unbracketed problem would not pursue any further the uncertain answer.

By contrast, a bracketed problem not only has "ownership" and "amount" questions, it also has "bracket" questions regarding the "amount" when a respondent giving a positive answer to the "ownership" question fails to provide an exact amount (Boxes 3a and 3b). Varying in the

Box 3a. The Bracketed Problem with a Single Bracket Question: an Example

```
(Ownership Question)
O1. People sometimes receive property or lump sum payments of money
from such things as pension settlements, insurance settlements,
cash in annuities, or inheritances. ... In the past two years, did
you/or your husband/or your wife/or your partner receive a lump sum
of money or property that you have not already told me about?
      1.      Yes
      5.      No
      7.      Other
      8.      DK (don't know); NA (not ascertain)
      9.      RF (refused)
      Blank.  INAP (Inapplicable)

(If "Yes" to the Ownership Question, then)
(Amount Question)
A1. About how much did you/or your husband/or your wife/or your
partner receive from that?
      0-999996.   Actual value
        999997.   Other
        999998.   DK (don't know); NA (not ascertained)
        999999.   RF (refused)
         Blank.   INAP (Inapplicable)

(If "999998" or "999999" to the Amount Question, then)
(Bracket Question)
B1. Did it amount to less than $5,000, more than $5,000, or what?
      1.      Less than $5,000
      3.      About $5,000
      5.      More than $5,000
      8.      DK (don't know); NA (not ascertain)
      9.      RF (refused)
      Blank.  INAP (Inapplicable)
```

structure, a bracket question generally asks a respondent if the amount of the "thing" in interest that he or she "owns" is more (or less) than a certain break point. Since this question does not ask for an exact amount, it is likely that a respondent unable or unwilling to provide a definite answer to the amount question becomes able or willing to give a definite answer to the bracket question.

A bracketed problem may have one (Box 3a) or more (Box 3b) "bracket" questions; each of them corresponds to a distinct break point. (In the example Box 3a, the only break point is $50,000; while in the example Box 3b, the three break points are $3,000, $5,000, and $10,000.) Since an unbracketd problem is equivalent to a bracketed problem with no bracket question, it

Box 3b. The Bracketed Problem with Multiple Bracket Questions: an Example

```
(Ownership Question)
O1. Are you/or your husband/or your wife/or your partner currently
receive any income from veteran benefits?
      1.      Yes
      5.      No
      7.      Other
      8.      DK (don't know); NA (not ascertain)
      9.      RF (refused)
      Blank.  INAP (Inapplicable)

(If "Yes" to the Ownership Question, then)
(Ownership Question Continued)
O2. Who received that?
      1.      R only
      2.      Spouse or Partner
      3.      Both
      7.      Other
      8.      DK (don't know); NA (not ascertained)
      9.      RF (refused)
      Blank. INAP (Inapplicable)

(If "1" or "3" to the above question, then)
A1. (Amount Question for R's Veteran Benefits)
How much did you (yourself) receive last month from that?
      0-999996.   Actual value
        999997.   Other
        999998.   DK (don't know); NA (not ascertained)
        999999.   RF (refused)
         Blank.   INAP (Inapplicable)
```

```
(If "999998" or "999999" to the Amount Question, then)
(Bracket Question - 1)
B1. Did it amount to less than $3,000, more than $3,000, or what?
      1        Less than $3,000
      3.       About $3,000
      5.       More than $3,000
      8.       DK (don't know); NA (not ascertain)
      9.       RF (refused)
      Blank.   INAP (Inapplicable)


If "1" to the Bracket Question, then)
(Bracket Question - 2)
B2. Did it amount to less than $5,000, more than $5,000, or what?
      1.       Less than $5,000
      3.       About $5,000
      5.       More than $5,000
      8.       DK (don't know); NA (not ascertain)
      9.       RF (refused)
      Blank.   INAP (Inapplicable)


(If "1" to the Bracket Question - 2, then)
(Bracket Question - 3)
B3. Did it amount to less than $10,000, more than $10,000, or what?
      1.       Less than $5,000
      3.       About $5,000
      5.       More than $5,000
      8.       DK (don't know); NA (not ascertain)
      9.       RF (refused)
      Blank.   INAP (Inapplicable)
```

may be viewed as a special type of bracketed problem.

IV. Producing and Organizing Imputation Results

a). Translating an Imputation Problem for IMPUTE

In order to use the imputation program, IMPUTE, one first needs to translate an imputation problem into a set of variables required by the program. Three types of variables are essential for IMPUTE, including the amount variable, the control variable, and, for a bracketed problem, the bracket variable(s). The amount variable is one for which imputations are done. In general, this variable corresponds to the amount question of an imputation problem, taking on the

value that answers the amount question. Similarly, the control variable corresponds to the ownership question, dictating if a respondent has a positive value on the amount variable; and a bracket variable corresponds to a bracket question, providing with bracket information about the amount when the exact value on the amount variable is not available. The relationship between an imputation problem and the variables translated from the problem is summarized in Box 4.

```
       Box 4. The Relationship Between an Imputation Problem and
                 the Variables Translated from the Problem

┌─────────────────────────────────────────────────────────────────────────┐
│  Question in an Imputation Problem                    Translated Variable │
│                                                                           │
│  Ownership Question       -------- ->                 Control Variable    │
│                                                                           │
│  Amount Question          -------- ->                 Amount Variable     │
│                                                                           │
│  Bracket Question(s)      -------- ->                 Bracket Variable(s) │
│                                                                           │
└─────────────────────────────────────────────────────────────────────────┘
```

The translation of an imputation problem into the variables required appears to be an easy task. But, in reality, that is not always true, especially for creating control variables. To impute "veteran benefits a *respondent* received last month" for the problem in Box 3b, for example, one needs to integrate the two ownership questions, O1 and O2, to create the control variable. That is, whether a respondent received veteran benefits last month may only be determined by combining the information collected from both questions O1 and O2. In fact, as indicated in Section V, about half of the control variables used for HRS98 imputations were created from two or more variables (or ownership questions).

b). The Imputation Strategies

The strategy appropriate for imputing an amount variable depends on the distributional characteristics of the variable as well as the structure of the underlying imputation problem. In

producing the HRS98 imputations, we adopted various strategies that we'd believed adequately reflect the characteristics of various amount variables and their underlying problems (see the last column in Box 9). These strategies, however, have three common features (Box 5).

First, all the imputations were based on a "mixed" method. As detailed in Appendix, Cao (2000), a "mixed" method replaces a missing value with a valid value based on a "hotdeck + regression" technique. Specifically, when it has a closed (e.g., 5,000 – 10,000) or bottom-open (e.g., -- 5,000) bracket, a missing value would be replaced by a value *randomly* selected from an appropriately-formed donor pool, a technique often referred to as "hotdeck"; when it has a top-open bracket (e.g., 5,000 -), it would be replaced by a value from the donor pool based on the predicted score from a *regression* model.

Second, whenever a regression technique was needed, the covariate vector always contained four basic demographic variables – including age, sex, educational attainment, and marital status. When imputing for the income from an asset (e.g., business or farm, stock, and bonds), however, the vector also included the asset variable.

And finally, when imputing a respondent variable and a spouse variable that is otherwise identical (e.g., respondent's veteran benefits and spouse's veteran benefits), a joint donor pool was always formed. The underlying rationale for pooling a respondent variable with a spouse

Box 5. The Common Features of the Imputation Strategies

```
1. A "mixed" method for all the imputations

2. A covariate vector containing four basic demographic variables
(age, sex, educational attainment, and marital status), and, when
imputing the income from an asset, the asset variable

3. A joint donor pool for imputing the same variables for a
respondent and his/her spouse
```

variable is of two-fold: (i) the question regarding the respondent's veteran benefits may be treated as the same as the question regarding the spouse's veteran benefits, so a joint donor pool is conceptually reasonable; and (ii) since pooling together the valid values of both respondent and spouse variables always increases the size of the donor pool, the reliability of the imputations for both the respondent variable and the spouse variable tends to be increased.

c). The Conventions Used for Organizing Imputation Results

Anticipating that the HRS98 preliminary imputations are most likely to be used together with other HRS98 variables in the public release data, we organized the results according to a set of useful conventions. First, an easy-to-understand connection between an imputation-related variable and the original amount variable is built. An imputation-related variable is always named as an original amount variable plus a single letter extension, a rule that becomes clearer and clearer as we proceed.

Second, for each amount variable imputed, six imputation-related variables are reported. Combined, these variables provide not only information on the structure of the imputation problem, but also information on the way the imputations have been done. A brief description of the variables is given in Box 6.

An imputation-related variable with extension "x"-- or the "x" variable for short -- stands for imputed values. When an original amount variable has a valid value, the "x" variable simply maintains the value; otherwise, it takes on a value out of imputation. By default, a valid value is one that an amount variable may possibly take on. Neither "missing", "don't know", nor "refused" is a valid value, although some missing values may be legitimate. (For example, a missing value on the spouse's social security income for a respondent who is single should be

legitimate. As a rule, for a legitimate missing on an amount variable, the "x" variable will be set to zero (0), while the control variable to missing (.).

Box 6. Imputation-Related Variables

Variable Name	Description
Q1234	Original amount variable in the public release data
Q1234x	Variable for imputed value (imputed amount variable)
Q1234t	Variable for imputation type or quality
Q1234s	Variable summarizing bracket information
Q1234c	Control variable for the amount variable
Q1234d	Variable for the lower bound value of the bracket
Q1234e	Variable for the upper bound value of the bracket

The "s" variable synthesizes the information on the bracket associated with the amount variable. More than often, a bracketed problem has two or more questions. To make it easier to use the information collected from all the bracket questions, the "s" variable combines the answers from the questions into a single index. The rule for reading such an index is shown in Box 7.

The "c" variable is the control variable about which we have already discussed. For "HRS98 preliminary imputations", a value "1" on the "c" variable stands for a positive answer (or "yes") to the ownership question, a value "5" for a negative answer (or "no"), while "8" and "9" for "don't know" and "refused", respectively.

The "d" and "e" variables give the lower and upper bound values, respectively, of the bracket. For a closed bracket (e.g., 5000 -10000), both the "d" and "e" variables should not be missing. For an open-end bracket (e.g., 5000 - , - 10000), the variable corresponding to the open

Box 7. The Rules for Reading an Index of the "s" Variable

```
1.The General Rule: Let an imputation problem with an amount
variable Q1234 have i bracket variables, var1, var2, ..., vari.
Each of the variables corresponds to a distinct break point
value, with var1 to the smallest value, var2 to the second
smallest, etc. Let a missing value on any of the bracket
variables be set to zero.

   The variable Q1234s is then defined as ---

   Q1234s = var1 + var2*10 + var3*100 + ... + vari*10⁽ⁱ⁻¹⁾.

2. The Special Rule: The following special codes are reserved for
   an "s" variable ---

   '-2': the missing value on the amount variable is legitimate;
   '-1': the value on the amount variable is valid;
   '0':  imputation for the amount variable was based on no
         bracket information.
```

end (lower or upper) should be missing. When the amount variable has no bracket information (thus open at both ends), therefore, both the "d" and "e" variables should be missing. *In addition, a special code, "-1", has been reserved for the two variables if the amount variable is about the spouse of an information-providing respondent and the respondent has no spouse.*

Combining the information from "x", "s", "c", "d", and "e" variables, the "t" variable is an index for identifying the type of imputation done to each record of the amount variable. Regardless the structure of an imputation problem, an imputed value belongs to one of ten categories, which are listed in the last column in Box 8. While a full appreciation of the availability of this imputation quality variable requires a thorough knowledge of the imputation program IMPUTE, users without any knowledge of the program may still find the variable useful if they want to subset, modify our imputation results, or do the imputation work of their own. [3]

---

[3] For example, if one does not trust in our work for open brackets, one may keep only those records with the values of the "t" variable less than 5.

Box 8. Imputation Type as Determined by the "x", "c", "d", "e", and "s" Variables

Original Amount Variable (Q1234)	Imputation Variable (Q1234x)	Control Variable (Q1234c)	Summary Bracket Variable (Q1234s)	Break point Value Bottom (Q1234d)	Top (Q1234e)	Imputation Type (Q1234t)
legitimate missing (1)	0	.	-2	.	.	1
legitimate missing(2)	0			-1	-1	1
valid zero	0	5	-1	.	.	2
valid positive value	>0	1	-1	.	.	3
closed bracket	>0	1	>0	>=0	>0	4
bottom-open bracket	>0	1	>0	.	>0	5
top-open bracket	>0	1	>0	>=0	.	6
bracket with no information	>0	1	>0	.	.	7
imputation without bracket	>0	1	0	.	.	8
imputed to non-ownership (Zero Amount)	0	8,9	0	.	.	9
imputed ownership and Amount	>0	8,9	0	.	.	10

d). Post-Imputation Treatment: the "z" Variable

Depending on the structure of an imputation problem, the amount variable for which imputations are actually done may or may not be what we want to know conceptually. Consider the example in Box 3b. The amount variable for this problem is naturally "the veteran benefits a respondent received last month". But what about if we want to know another variable, "the veteran benefits a respondent received in the last calendar year"? Can we construct this new

variable based on the imputations for the original amount variable?

The work involving the conversion of the imputations for an original amount variable into another related variable is called as "post-imputation treatment", while the new, converted variable is named with the original amount variable plus an extension "z" (e.g., Q1234z). In most of the cases in the "HRS98 preliminary imputations", the "x" to "z" conversions was straightforward, involving only simple mathematical computations. For other cases, however, the tasks were quite complicated, as illustrated for the example "the veteran benefits a respondent received in the last calendar year" in Box 9.

<div align="center">

Box 9. The "x" to "z" Conversion: an Example

</div>

```
Original Amount Variable: Q4803 (Veteran Benefits a Respondent
                                Received Last month (year 1998))
Imputation Variable: Q4803x (Veteran Benefits a Respondent
                                Received Last Month (year 1998))
Objective: Construct Q4803z (Veteran Benefits a Respondent
                                Received in Year 1997)

The "Starting-time-dependent" Conversion Strategy ---

1. Assumption: monthly veteran benefits a respondent received in
   year 1997 was about the same as he/she received last month

2. Determine when a respondent started to receive the benefits:
   Q4811 (starting year), Q4812 (starting month)

3. Impute Q4811x and Q4812x as some of the values of Q4811 and Q4812
   are missing

4. Q4803x to Q4803z Conversion:
   If Q4811x <= 1996 then Q4803z = Q4803x*12
   Else if Q4811x > 1997 then Q4803z = 0
   Else if Q4811x = 1997 then Q4803z = Q4803x*(13-Q4812x)
```

In order to know from "the veteran benefits a respondent received last month" (Q4803x) "the veteran benefits a respondent received in year 1997 (or last calendar year)" (Q4803z), we

adopted a so-called "starting-time dependent" x-to-z conversion strategy.[4] We first built the relationship between the monthly benefits in 1998 and in 1997. Without great loss of accuracy, we assumed that "monthly veteran benefits a respondent received--if any--in 1994 was about the same as he/she received last month". We then determined if a respondent was receiving any veteran benefits in 1997. To do this, we imputed two "starting-time" variables, Q4811 (year starting to receive the benefits) and Q4812 (month starting to receive the benefits), as some of the values of the variables were missing. Finally, we made the conversion based on the assumed the relationship between the monthly benefits in 1998 and in 1997, and the imputations for the "starting-time" variables.

V. The Structure of "HRS98 Preliminary Imputations"

"HRS98 preliminary imputations" are divided into two different data sets: "h1998i_f", and "h1998i_j". The names of the data files contain both HRS wave (1998) and section (f, and j) identifiers. That is, "h1998i_f" is a file containing imputations for the HRS 1998 data, section F (housing), while "h1998i_j" for section J (income and assets). Because of the structure of the original data, both "h1998i_f" and "h1998i_j" are household level data sets, with one record for each HRS 1998 household.

In this section, we list all the amount variables covered in the two imputation files. For each variable, we list its name, a label defining the variable[5], the variables involved in the creation of the control variable, and, when needed, a brief comment on the imputation strategy used. Users curious about our imputation strategy for any amount variable are encouraged to look up the codebook for the HRS98 preliminary release data as well as the documentation file

---

[4] As indicated in Box 10, this strategy was used for x-to-z conversion for many other variables as well.

of the program IMPUTE for a better understanding.

We first list the amount variables for which imputations were actually done (Box 10), then the variables for which the "x" to "z" conversions were made (Box 11). For convenience, we also provide the definitions of four aggregate variables included in the imputation files (Box 12). These aggregates are: household income, net wealth (in the file "h1998i_j"), main home equity, and second home equity (in the file "h1998i_f"). While a conscious effort was made to have them constructed as well as possible, users of the imputation files are not required to agree with us on the definitions.

Box 10. The Amount Variables Imputed

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
	SECTION F: HOUSING ("hrs4fx")	
F5B.MOBILE HOME VALUE $ Q2753	Q2742 Q2751	
F6.HOME VALUE $ Q2760	Q2743 Q2746	
F7A.$ PAYMENT ON MORTGAGE Q2769	Q2768M1 Q2768M2	annualized before imputation
F7D.$ OWE ON MORTGAGE Q2773	Q2768M1 Q2768M2	
F9A.$ PAYMENT 2ND MORTGAGE Q2777	Q2768M1 Q2768M2	

---

[5] This is the same label used in the HRS98 preliminary release data, combining the question number with a brief definition of the variable.

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
F9C.$ OWN ON 2ND MORTGAGE Q2780	Q2768M1 Q2768M2	
F10.$ PAYMENT OTHER LOAN Q2784	Q2768M1 Q2768M2	
F10E.$ OWN ON LOAN Q2787	Q2768M1 Q2768M2	
F11A.$ COULD BORROW Q2792	Q2791	
F11C.$ OWED EQUITY LOAD Q2794	Q2793	
F11D.PURCHASE PRICE Q2795	Q2743	HRS cohort skipped this question
F11E.YEAR ACQUIRE HOME Q2796	Q2796C=1	HRS cohort skipped this question
F13.REAL ESTATE TAX Q2809	Q2809	only imputing for "DK" and "RF" cases
F14.$ AMOUNT RENT Q2818	Q2743	annualized before imputation
F21D.HOW MUCH LUMP_SUM Q2843	Q2842	
F21K.$ ASSOC FEE Q2850	Q2849	annualized before imputation
F46E.PURCHASE PRICE 2ND HOME Q2949	Q2914	
F46F.$ VALUE 2ND HOME Q2950	Q2914	
F47A.TOTAL $ OWED ALL MORTGAGES/LOANS-2ND Q2957	Q2955M1 Q2955M2	
F49A.$ PAYMENT ALL MORTGAGES-2ND Q2966	Q2955M1 Q2955M2	annualized before imputation
F53.REAL ESTATE TAX-2 Q2997	Q2914	

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
	SECTION J: INCOME AND ASSETS ("hrs4jx")	
J6.R HOW MUCH SELF-EMPLOYMENT Q4634	Q4633	joint donor pool with Q4670
J8.R HOW MUCH WAGES AND SALARY Q4643	Q4642	joint donor pool with Q4676
J10.R HOW MUCH PROF PRACTICE OR TRADE Q4652	Q4651	joint donor pool with Q4683
J11A.R HOW MUCH TIPS, BONUSES, COMMISSIO Q4658	Q4657	joint donor pool with Q4690
J12A.HOW MUCH OTHER INCOME LCY Q4664	Q4663	joint donor pool with Q4696
J14.SP HOW MUCH SELF-EMPLOYMENT Q4670	Q4669	joint donor pool with Q4634
J16.SP HOW MUCH WAGES AND SALARY Q4676	Q4675	joint donor pool with Q4643
J18.SP HOW MUCH PROF PRACTICE OR TRADE Q4683	Q4682	joint donor pool with Q4652
J19A.SP HOW MUCH TIPS, BONUSES, COMMISSI Q4690	Q4689	joint donor pool with Q4658
J20A.SP HOW MUCH OTHER INCOME Q4696	Q4695	joint donor pool with Q4664
J23.(J10)AMT UNEMPL COMP LCY Q4704	Q4702 Q4703	joint donor pool with Q4712
J28.(J12)AMT S/P UNEMPL COMP LCY Q4712	Q4702 Q4703	joint donor pool with Q4704

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
J35.(J15)AMT WORKERS COMP LCY Q4723	Q4721 Q4722	joint donor pool with Q4734
J40.(J17)AMT S/P WORKER COMP LCY Q4734	Q4721 Q4722	joint donor pool with Q4723
J47.(JX2)R AMT RECVD SS LAST MO Q4745	Q4743 Q4744	joint donor pool with Q4756
J50A1.(JX5)SP HOW MUCH SS BENEFITS Q4756	Q4743 Q4744	joint donor pool with Q4745
J61.(J11)R/S SSI $ Q4783	Q4781	
J66.(J12B)AMT WELFARE LCY Q4798	Q4796 Q4797	
J69.(J13A)R VET BENEFIT LAST MONTH Q4803	Q4801 Q4802	joint donor pool with Q4816
J72.(J13N)AMT SP VET BEN LAST MO Q4816	Q4801 Q4802	joint donor pool with Q4803
J77.(J15)REAL ESTATE $ Q4831	Q4830	
J82.(J16C)RENTAL RECEIVE $ LCYEAR Q4849	Q4845	integration with annualized "per-period" income (Q4847) to increase the size of donor pool
J84.(J18)BUSINESS OR FARM $ Q4857	Q4856	
J89.(J19D)BUS/FARM INCOME $ LCYEAR Q4877	Q4872	integration with annualized "per-period" income (Q4874) to increase the size of donor pool
J93.(J22-2)TOTAL $ IRA ACCOUNT Q4887	Q4884	

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
J95.(J23A-1)IRA NO REG INC WITHDRAW $-1 Q4898	Q4884 Q4897	annualized before imputation (value in two years)
J100.(J22-2)TOTAL $ IRA ACCOUNT-2 Q4909	Q4884 Q4885 Q4908	
J102.(J23A-2)IRA NO REG INC WITHDRAW $-2 Q4919	Q4884 Q4918	annualized before imputation (value in two years)
J107.(J22-3)TOTAL $ IRA ACCOUNT-3 Q4930	Q4884 Q4885 Q4929	
J109.(J23A-3)IRA NO REG INC WITHDRAW $-3 Q4938	Q4884 Q4937	annualized before imputation (value in two years)
J114.(J27D)ANNUITY AMOUNT Q4949	Q4947M1 Q4947M2	annualized before imputation
J124.(J29B)AMT WITHDRAW Q4963	Q4947M1 Q4947M2	
J132.(J30D)AMT PENSION LAST MONTH Q4974_1	Q4969 Q4970	
J132.(J30D)AMT PENSION LAST MONTH Q4974_2	Q4969 Q4970 Q4971	
J148.(J31D)AMT PENSION LAST MONTH Q4997_1	Q4969 Q4970	
J148.(J31D)AMT PENSION LAST MONTH Q4997_2	Q4969 Q4970 Q4994	
J168.(J33D)AMT ANNUITY INC Q5029_1	Q5023 Q5024	
J168.(J33D)AMT ANNUITY INC Q5029_2	Q5023 Q5024 Q5025	
J181.(J35C)AMT ANNUITY INC SP Q5058_1	Q5023 Q5024	

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
J181.(J35C)AMT ANNUITY INC SP Q5058_2	Q5023 Q5024 Q5055 Q5056	
J196.(J35X1D)AMOUNT CASH SETTLEMENT Q5088_1X	Q5084_1	
J208.(J37)TOTAL $ STOCKS Q5100	Q5099	
J213.STOCK INCOME $ LCYEAR Q5122	Q5118	integration with annualized "per-period" income (Q5120) to increase the size of donor pool
J216.(J41)TOTAL $ BONDS Q5143	Q5142	
J221.(J19D)BONDS INCOME $ LCYEAR Q5165	Q5161	integration with annualized "per-period" income (Q5163) to increase the size of donor pool
J223.(J44)TOTAL $ CHECKING, SAVINGS Q5186	Q5185	
J228.(J19D)CHECKING INCOME $ LCYEAR Q5205	Q5201	integration with annualized "per-period" income (Q5203) to increase the size of donor pool
J230.(J48)TOTAL $ CD Q5225	Q5224	
J236.CD INCOME $ LCYEAR Q5247	Q5243	integration with annualized "per-period" income (Q5245) to increase the size of donor pool
J238.(J51A) TRANSPORTATION Q5261	Q5260	
J240.(J53)OTHER ASSETS $ Q5274	Q5273	
J242.(J55)OTHER ASSETS INCOME $ Q5283	Q5283	
J245.(J60A)TRUST FUNDS $ LCY Q5324	Q5323	

Box 10. The Amount Variables Imputed (Continued)

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
J248.(J61A)AMT ALIMON $ LCY Q5328	Q5327	
J250.(J62A)OTHER SOUR $ 1996 Q5335	Q5334	
J299.(J80B)$ TRUSTS Q5473	Q5469	
J299F.$ TRUSTS NOT PREVIOUSLY REPORTED Q5480	Q5479	
J303.(J82)DEBTS $ Q5487	Q5486	
J309.(J86-1)LUMP $-1 Q5512	Q5508	
J315.(J86-2)LUMP $-2 Q5519	Q5515	
J321.(J86-3)LUMP $-3 Q5526	Q5522	

Box 11. The "z" Variables

Label Variable Name	Variables Used for Creating the Control Variable	Comments on the Imputation Strategy
J47.R SS Benefits LCY Q4745Z	Q4750 Q4751	starting-time dependent
J50A1.SP SS Benefits LCY Q4756Z	Q4761 Q4762	starting-time dependent
J61.R/SP SSI LCY Q4783Z	Q4787 Q4788	starting-time dependent
J69.R Veteran Benefits LCY Q4803Z	Q4811 Q4812	starting-time dependent
J72.SP Veteran Benefits LCY Q4816Z	Q4824 Q4825	starting-time dependent
J95.R/SP IRA 1 Withdrawal LCY Q4898Z		$= Q4898X / 2$
J102.R/SP IRA 2 Withdrawal LCY Q4919Z		$= Q4919X / 2$
J109.R/SP IRA 3 Withdrawal LCY Q4938Z		$= Q4938X / 2$
J114.R/SP Other Annuity Income LCY Q4949Z	Q4950 Q4955 Q4956	starting-time dependent
J124.R/SP Other IRAs Withdrawal LCY Q4963Z		$= Q4963X / 2$
J132.R Pension 1 Income LCY Q4974_1Z	Q4983_1 Q4984_1	starting-time dependent
J132.R Pension 2 Income LCY Q4974_2Z	Q4983_2 Q4984_2	starting-time dependent
J148.SP Pension 1 Income LCY Q4997_1Z	Q5006_1 Q5007_1	starting-time dependent

Box 11. The "z" Variables (Continued)

Label Variable Name Strategy	Variables Used for Creating the Control Variable	Comments on the Imputation
J148.SP Pension 2 Income LCY Q4997_2Z	Q5006_2 Q5007_2	starting-time dependent
J168.R Annuity 1 Income LCY Q5029_1Z	Q5035_1 Q5036_1	starting-time dependent
J168.R Annuity 2 Income LCY Q5029_2Z	Q5035_2 Q5036_2	starting-time dependent
J181.SP Annuity 1 Income LCY Q5058_1Z	Q5064_1 Q5065_1	starting-time dependent
J181.SP Annuity 2 Income LCY Q5058_2Z	Q5064_2 Q5065_2	starting-time dependent
J196.R Type A Pension Cash 1 LCY Q5088_1Z	Q5085_1 Q5087_1	starting-time dependent
J309.R/SP Lump Sum 1 LCY Q5512Z	Q5511	starting-time dependent
J315.R/SP Lump Sum 2 LCY Q5519Z	Q5518	starting-time dependent
J321.R/SP Lump Sum 3 LCY Q5526Z	Q5525	starting-time dependent

Box 12. Household Income, Housing Equity, and Net Wealth

1. Household income in the last calendar year:

hhinc = Q4634X+Q4670X+Q4643X+Q4676X+Q4652X+Q4683X +Q4658X+Q4690X
+Q4664X+Q4696X+Q5283X+Q4704X +Q4712X+Q4723X+Q4734X+Q4798X
+Q5324X+Q5328X+Q5335X +Q4745Z+Q4756Z+Q4783Z+Q4803Z+Q4816Z
+Q4849X +Q4877X+Q4898Z+Q4919Z+Q4938Z+Q4974_1Z+Q4974_2Z
+Q4997_1Z+Q4997_2Z+Q5029_1Z+Q5029_2Z+Q5058_1Z+Q5058_2Z
+Q5122X+Q5165X+Q5205X+Q5247X+Q4949Z

2. Net wealth at the time of interview:

assets = Q4831X+Q4857X+Q4887X+Q4909X+Q4930X+Q5100X+Q5143X+Q5186X
+Q5225X+Q5261X+Q5274X+Q5480X-Q5487X

3. Main home equity:

home1 = Q2753X+Q2760X-Q2773X-Q2780X-Q2787X-Q2794X

4. Second home equity:

home2 = Q2950X-Q2957X

VI. Questions and Answers

Q1. How can I merge the imputation data with the HRS 1998 preliminary release data?

A1. In general, you need to merge the two groups of data by "HHID", "FSUBHH", and, when necessary, "PN". It is important to note, however, that the HRS98 preliminary imputations are based on a set of identification variables that are slightly different from those in the preliminary release data. The differences in the identification variables have been described in a data alert at the HRS Web Site: http://www.umich.edu/~hrswww/center/alerts/fpre011.html. Before merging the imputation data with the preliminary release data, therefore, you first need to correct the identification variables in the preliminary release data according to the information provided in the aforementioned data alert.

Q2. What variables were imputed for the HRS98 preliminary release data?

A2. All the variables imputed for the HRS98 preliminary release data are listed in Box 10. These variables were selected based on the rules detailed in Box 1.

Q3. How did you impute a variable?

A3. For most of the variables selected for imputation, we used a "mixed" imputation method, which combines a "hotdeck" imputation procedure and a "linear regression" procedure. Please see Cao (2000), Appendix, for detail.

Q4. Can I impute a variable myself?

A4. Yes, you certainly can. If the variable you want to impute was not selected for imputation by us--most likely, the variable would have no bracket--you may impute it based on any imputation procedure that you believe reasonable ("hotdeck", for example). If you want to do imputations yourself systematically, using a different method, you may do so by taking advantage of the availability of the original bracket variables in the data sets. And you may use a summary bracket variable (the "s" variable) to simplify your work.

Q5. What is the main difference between an "x" variable and a "z" variable?

A5. An "x" variable is an original imputation variable, while a "z" variable is one converted from a corresponding "x" variable. The conversion is needed usually because the "x" variable is not what we want conceptually. A "z" variable must have a corresponding "x" variable, but not vice versa.

Q6. Can I define "hhinc" myself?

A6. Yes, you can, and sometimes you should. For example, if you don't want to have household income contain anything from assets, you should drop the asset income components from the equation in Box 12.

Q7. Why are there multiple imputation-related variables for each amount variable?

A7. For each amount variable, there are six imputation-related variables: the "x", "c", "s", "d", "e", and "t" variables. (For definitions of the variables, see Box 6.) We include these variables in the data sets to accommodate various needs of perspective users. Those who just want to have imputed values may keep only "x" and, if applicable, "z" variables when working with the data.

Q8. How do you create a "c" variable?

A8. The creation of a "c" variable may be complicated or very simple, depending on the structure of the imputation problem. Sometimes we need only a single variable to create the control variable for the problem. In other cases, we need to incorporate information from several variables to determine if a respondent or household receives welfare income, pays medical expenditures, etc.

Q9. What should I do if I find any problems with the imputations?

A9. Problems with the imputations may exist for various reasons. If you believe that the imputations for a variable do not look correct, please contact hrsquest@isr.umich.edu.

References

Cao, Honggao. IMPUTE: A SAS Application System from Missing Value Imputations --- With Special Reference to HRS Income/Assets Imputations. Institute for Social Research, University of Michigan, March 2000.

Herringa, S. G., Hill, D. H., and Howell, D. A. "Unfolding Brackets for Reducing Item Non-Response in Economic Surveys," Health and Retirement Study Working Paper, No. 94-029,

June 1998.

Hill, D. H. "Unfolding Bracket Method in the Measurement of Expenditures and Wealth," in J. P. Smith, and R. J. Willis (eds.), Wealth, Work, and Health, University of Michigan, 1999.

Juster, F. T., and Smith, J. P. "Improving the Quality of Economic Data: Lessons from the HRS," HRS Working Paper, No. 94-027, July 1994.