NATIONAL HEALTH AND AGING TRENDS STUDY (NHATS)
Round 1 Income Imputation

November 2, 2012

Overview

In preparing survey data files for analysis, imputation is often used to address item nonresponse, particularly when complex multi-variate recodes are required that are built up from a collection of more detailed questions (Marker, Judkins, and Winglee, 2001). Round 1 of the National Health and Aging Trends Survey (NHATS) includes imputed values for total income. Both a continuous measure and a bracket value are provided, with separate bracket values for single respondents and those who are married or are living with a partner. We used a cyclical n-partition hot deck (see Judkins 1997) to generate five imputations of each measure. This technical paper provides details on the imputation strategy.

Income Sources Collected in NHATS

Round 1 of NHATS collects information on sources of income (yes/no) and amounts for each source. Income sources that flow from assets are asked after identifying the existence of the asset and the current worth. Respondents with a spouse/partner were given the option of reporting amounts for themselves either together or separately from their spouse/partner. Table 1 shows the income sources included in NHATS.

Table 1. Summary of Income Sources Collected in NHATS

Sources of Income and Income Amounts[1]	Time frame
Social Security	Last Month
Supplemental Security Income	Last Month
Veteran's Administration	Last Month
Pension plan	Last Month
Earned Income	Last Month/Last Paycheck
Interest/dividend income from any: mutual funds/stocks, bonds, bank accounts, or CDs[1]	Last Year
Retirement account withdrawals[1]	Last month/Last year
Total income from all sources	Last Year

[1]Asked after questions about the existence and current worth of relevant asset

Respondents who refused or reported they did not know an exact source amount were given the option of selecting one of five bracketed ranges. These ranges were developed from reports for the 65 and older population in the Survey of Consumer Finance for 2007 adjusted to 2011 dollars. Separate ranges were provided for individually reported (respondent or spouse/partner amounts separately) and for jointly reported (respondent and spouse/partner together) amounts.

After reporting about the amount from each individual source, respondents are asked to report total income from all sources. Again, respondents who report don't know or who refuse are offered a set of five bracketed ranges, tailored to joint/separate amounts for those with a spouse/partner, from which to choose. The reports of total income are not required to be consistent with source amounts reported earlier in the interview.

Extent of Missing Data for Total Income

Fifty-six percent of the sample provided a total income amount and an additional 13% reported a bracketed value instead. Thus, a bracketed value could be created from reports for 69% of the sample and required imputation for 31%. An exact value was then imputed for 44% of the sample (13% within a reported bracketed value and 31% within an imputed bracketed value).

Imputation Methodology

Westat's AutoImpute software was used to impute five values of the total income items. AutoImpute uses a cyclical n-partition hot deck (an approach analogous to the Gibbs sampler but using the hot deck to generate the imputations). (See Judkins 1997; Judkins et al. 2007; Judkins, Piesse, and Krenzke 2008; Krenzke and Judkins 2008.) This software is designed to facilitate preservation of multivariate distributions while also ensuring that imputations maintain skip patterns and adhere to constraints. In this application an example of a constraint is ensuring imputations for specific amounts fall within reported (or imputed) bracket ranges.

The cyclical n-partition hot deck procedure initially imputes all target variables (i.e., items requiring imputation) using a simple hot deck that uses specified auxiliary variables and skip controllers. Using the initial imputed variables, a model is fit for each target variable using simple forward stepwise regression selection. The predicted values of the target variable from the final model are used to generate imputed values by using predictive mean matching for ordinal (ordered categorical) target variables and clustering for unordered categorical target variables. Predictive mean matching uses a hot deck with the skip controllers as hard boundaries and the predicted values from the final model as soft boundaries. For unordered categorical target variables, a k-means clustering algorithm is used on the vector of predicted values for each level, and then a hot deck is used to impute the target variable with the skip controllers as hard boundaries and the cluster membership indicators as soft boundaries. For more details on the procedure see Judkins et al. (2007).

Variables Used in Imputation

Three classes of variables were used in the imputation (see Appendix tables):
1. Source variables that indicate (yes/no) whether the respondent (and his/her spouse/partner, if applicable) has the particular source of income (referred to below as "fencepost" variables);
2. Source amount variables that indicate the amount of income from the particular source; and
3. Auxiliary variables that included respondent characteristics (e.g., age, race/ethnicity, gender, educational attainment, home ownership, veteran's status, labor force status, spouse/partner's labor force status, interviewer observations about the home condition) and skip pattern controllers.

In order to preserve the joint distribution of the full set of income variables, all missing fencepost, source amount, auxiliary, and total income variables were imputed. For respondents with a spouse/partner, separate respondent and spouse/partner variables as well as the combined variable were imputed for each source.

When imputing total income variables, the source variables and source amounts were used, along with the auxiliary variables. Total income value was also constrained to fall within the reported/imputed bracket amount.

Income Imputation Variables in the SP File
The following imputed variables are included on the SP data file:

Variable name	Label	Description
2010 Total Income Value		
ia1totinc	R1 IA50 TOTAL INCOME	Actual
ia1toincimf	R1 F IMPUTED TOTAL INC FLG	Flag indicating imputation
ia1toincim1-5	R1 IA50 IMPUTED TOTAL INC1-INC5	Imputed values 1-5
2010 Total Income Range – Respondents who have spouse/partner		
ia1toincesjt	R1 IA51A JOINT EST TOT INCOME	Actual
ia1eincimjf	R1 F IMPTD JOINT EST TOT INC FLG	Flag indicating imputation
ia1eincimj1-5	R1 IA51A IMP EST JOINT TOT INC1-INC5	Imputed values 1-5
2010 Total Income Range – Respondents who are single		
ia1toincessg	R1 IA51B SNGLE EST TOT INC	Actual
ia1eincimsf	R1 F IMPUTED SGL EST TOT INC FLG	Flag indicating imputation
ia1eincims1-5	R1 IA51B IMP EST SP SGL TOT INC1-INC5	Imputed values 1-5

Using the Five Versions of the Imputed Variable in Analysis
For each of the three total income variables that was imputed, five sets of imputed variables were generated. For item respondents, these five sets contain copies of the reported data. For item nonrespondents, the five sets contain five independently generated imputed values. These five sets of imputed variables are provided to enable data users to use multiple imputation variance estimators and analysis techniques (see, for example, Rubin 1996) to account for the effects of item nonresponse and imputation error in variance estimates for analyses that use these income variables.

References

Judkins, D. (1997). Imputing for Swiss cheese patterns of missing data. *Proceedings of Statistics Canada Symposium '97*, 143-148.

Judkins, D., Krenzke, T., Piesse, A., Fan, Z., and Haung, W.C. (2007). Preservation of skip patterns and covariate structure through semi-parametric whole questionnaire imputation. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 3211-3218.

Judkins, D., Piesse, A., and Krenzke, T. (2008). Multiple semi-parametric imputation. *Proceedings of the Section on Survey Research Methods of the American Statistical Association.*

Krenzke, T. and Judkins, D. (2008). Filling in blanks: Some guesses are better than others – Illustrating the impact of covariate selection when imputing complex survey items. *Chance*, 21(3), 7-13.

Marker, D. A., Judkins, D. R., and Winglee, M. (2001). Large-scale imputation for complex surveys. In *Survey Nonresponse*, Eds. R. M. Groves, D. A. Dillman, E. L. Eltinge, and R. J. A. Little. New York: John Wiley & Sons.

Rubin, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association,* 91(434)*,* 473-489.

Appendix. Lists of Variables Used in NHATS Round 1 Income Imputation

Table 1. Source ("Fencepost") Variables

#	Variable name	Label	% missing
1	ia1recsspa1	R1 IA1 SP REC SOCIAL SECURITY	4.1
2	ia1recsspa2	R1 IA1 SPOUSE PART REC SOC SEC	4.1
3	ia1recsspa3	R1 IA1 NO SOC SECURTY PAYMNT REC	4.1
4	ia1recssils1	R1 IA4 SP RECEIVD SSI LAST MONTH	4.7
5	ia1recssils2	R1 IA4 SPOUSE PRT REC SSI LST MO	4.7
6	ia1recssils3	R1 IA4 NO SSI RECEIVD LAST MONTH	4.7
7	ia1rvapayls1	R1 IA5 SP REC PAY FRM VA LAST MO	4.2
8	ia1rvapayls2	R1 IA5 SPOUS PA REC VA PAY LSTMO	4.2
9	ia1rvapayls3	R1 IA5 NO VA PAY REC LAST MONTH	4.2
10	ia1penjobou1	R1 IA6 SP HAS PENSION PLAN	5.0
11	ia1penjobou2	R1 IA6 SPOUSE HAS PENSION PLAN	5.0
12	ia1penjobou3	R1 IA6 NO PENSION PLAN	5.0
13	ia1iraothac1	R1 IA7 SP HAS IRA OTH RETIRE ACC	7.1
14	ia1iraothac2	R1 IA7 SPOUSE HAS IRA OTHR ACC	7.1
15	ia1iraothac3	R1 IA7 NO IRA OTHR RETIRE ACCT	7.1
16	ia1mutfdstk1	R1 IA8 SP OWNS MUTUAL FUND STOCK	7.0
17	ia1mutfdstk2	R1 IA8 SPOUSE OWNS FUNDS STOCK	7.0
18	ia1mutfdstk3	R1 IA8 SP SPOUSE OWN FUNDS STOCK	7.0
19	ia1mutfdstk4	R1 IA8 NO FUNDS OR STOCK OWNED	7.0
20	ia1ownbond1	R1 IA9 SP OWNS BONDS	7.3
21	ia1ownbond2	R1 IA9 SPOUSE OWNS BONDS	7.3
22	ia1ownbond3	R1 IA9 SP SPOUSE OWN BONDS	7.3
23	ia1ownbond4	R1 IA9 NO BONDS OWNED	7.3
24	ia1bnkacccd1	R1 IA10 SP OWNS CHECK ACCT	5.8
25	ia1bnkacccd2	R1 IA10 SPOUSE OWNS CHECK ACCT	5.8
26	ia1bnkacccd3	R1 IA10 SP SPOUSE OWN CHECK ACCT	5.8
27	ia1bnkacccd4	R1 IA10 NO CHECK ACCT OWNED	5.8
28	ia1bnkacccd5	R1 IA10 SP OWNS SAVINGS ACCT	7.0
29	ia1bnkacccd6	R1 IA10 SPOUSE OWNS SAVING ACCT	7.0
30	ia1bnkacccd7	R1 IA10 SP SPOUSE OWN SAVNG ACT	7.0
31	ia1bnkacccd8	R1 IA10 NO SAVINGS ACCT OWNED	7.0
32	ia1bnkacccd9	R1 IA10 SP OWNS CDS	7.9
33	ia1bnkaccc10	R1 IA10 SPOUSE OWNS CDS	7.9
34	ia1bnkaccc11	R1 IA10 SP SPOUSE OWN CDS	7.9
35	ia1bnkaccc12	R1 IA10 NO CDS OWNED	7.9
36	ia1realestt1	R1 IA13 SP OWNS REAL ESTATE	5.3
37	ia1realestt2	R1 IA13 SPOUSE OWNS REAL ESTATE	5.3
38	ia1realestt3	R1 IA13 SP SPOUSE OWN REAL ESTTE	5.3
39	ia1realestt4	R1 IA13 NO REAL ESTATE OWNED	5.3
40	lf1workfpay	R1 LF1 WORKED FOR PAY RECENTLY	1.4
41	lf1abstlstwk	R1 LF2 ABSENT FRM JOB LAST WEEK	3.3
42	lf1wrkplstmn	R1 LF3 WORK FOR PAY IN LST MONTH	0.1
43	lf1huswifwrk	R1 LF13 HUSB/WIFE/PARTN PAY WORK	1.6

Table 2. Source Amount Variables

	Variable name	Label	% missing
1	ia1ssrrpymnt	R1 IA14 RECENT MTHLY SS RR PYMNT	20.8
2	ia1ssrrjtamt, ia1ssrrspamt, ia1ssrrptamt	IA14A-IA16A AMOUNT OF SOCIAL SECURITY/RAILROAD RETIREMNT	20.8
3	ia1ssrrjtest, ia1ssrrspest, ia1ssrrptest	IA14B-IA16B RANGE OF SOCIAL SECURITY/RAILROAD RETIREMNT	15.1
4	ia1ssipymnt	R1 IA17 RECENT MTHLY SSI PAYMENT	23.1
5	ia1ssijtamt, ia1ssispamt, ia1ssiptamt	IA17A-IA19A AMOUNT OF SUPPLEMENTAL SECURITY INCOME	20.6
6	ia1ssijtest, ia1ssispest, ia1ssiptest	IA17B-IA19B RANGE OF SUPPLEMENTAL SECURITY INCOME	16.3
7	ia1vapymnt	R1 IA20 RECENT MNTHLY VA PAYMENT	40.0
8	ia1vajtamt, ia1vaspamt, ia1vaptamt	IA20A-IA22A AMOUNT OF VETERANS ADMINISTRATION	27.8
9	ia1vajtest, ia1vaspest, ia1vaptest	IA20B-IA22B RANGE OF VETERANS ADMINISTRATION	22.0
10	ia1penpymt	R1 IA23 RCNT MTH JOBPENSION PYMT	21.9
11	ia1penjtamt, ia1penspamt, ia1penptamt	IA23A-IA25A AMOUNT OF PENSION PLAN	23.6
12	ia1penjtest, ia1penspest, ia1penptest	IA23B-IA25B RANGE OF PENSION PLAN	17.3
13	ia1retworth	R1 IA26 RETIREMENT ACCOUNT WORTH	40.2
14	ia1retjtwrt, ia1retspwrt, ia1retptwrt	IA26A-IA28A AMOUNT OF ANY RETIREMENT ACCOUNTS	43.6
15	ia1retjtest, ia1retspest, ia1retptest	IA26B-IA28B RANGE OF ANY RETIREMENT ACCOUNTS	29.2
16	ia1rtlmwdrw	R1 IA29 RETRMNT WDRW AMT LST MTH	25.0
17	ia1rtlmjtwdr, ia1rtlmspwdr, ia1rtlmptwdr	IA29A-IA31A AMOUNT OF RETIREMENT ACCTS WITHDRAW LST MTH	26.5
18	ia1rtlmjtest, ia1rtlmspest, ia1rtlmptest	IA29B-IA31B RANGE OF RETIREMENT ACCTS WITHDRAW LST MTH	21.7
19	ia1rtyrwdrw	R1 IA32 RETRMNT WDRWL AMT LST YR	28.1
20	ia1rtyrjtamt, ia1rtyrspamt, ia1rtyrptamt	IA32A-IA34A AMOUNT OF PARTNERS ACCTS WITHDRAW LST MTH	31.0
21	ia1rtyrjtest, ia1rtyrspest, ia1rtyrptest	IA32B-IA34B RANGE OF PARTNERS ACCTS WITHDRAW LST MTH	24.3
22	ia1skbdwrth	R1 IA35 NONRETR STKS BNDS WRTH	45.9
23	ia1skbdjtwrt, ia1skbdspwrt, ia1skbdptwrt	IA35A-IA37A AMOUNT OF WORTH OF FUNDS & STOCKS	46.7
24	ia1skbdjtest, ia1bndjtest, ia1skbdspest, ia1bndspest, ia1skbdptest, ia1bndptest	IA35B-IA37C RANGE OF WORTH OF FUNDS & STOCKS	32.4
25	ia1bkcdwrth	R1 IA38 BANK ACCT CD WORTH	35.3
26	ia1bkcdjtwrt, ia1bkcdspwrt, ia1bkcdptwrt	IA38A-IA40A AMOUNT OF WORTH OF BANK ACCDS & CDS	36.2
27	ia1bkcdjtest, ia1bnkjtest, ia1bkcdspest, ia1bnkspest, ia1bkcdptest, ia1bnkptest	IA38B-IA40C RANGE OF WORTH OF BANK ACCDS & CDS	26.6
28	ia1itdvinc	R1 IA41 AMT INT DIV INCOME LS YR	47.2
29	ia1itdvjtamt, ia1itdvspamt, ia1itdvptamt	IA41A-IA43A AMOUNT OF TOT WRTH COMBNED INTERST	48.3
30	ia1itdvjtest, ia1itdvspest, ia1itdvptest	IA41B-IA43B RANGE OF TOT WRTH COMBNED INTERST	32.3
31	ia1brewrt	R1 IA44 BUSINESS REALESTATE WRTH	25.5
32	ia1brejtwrt, ia1brespwrt, ia1breptwrt	IA44A-IA46A AMOUNT OF TOT WRTH ALL REAL ESTATE	31.1
33	ia1brejtest, ia1brespest, ia1breptest	IA44B-IA46B RANGE OF TOT WRTH ALL REAL ESTATE	22.6
34	ia1breiinc	R1 IA47 BUS REALESTATE INC LS YR	27.4
35	ia1breijtamt, ia1breispamt, ia1breiptamt	IA47A-IA49A AMOUNT OF COMB INCOME FROM REL ESTAT	30.1
36	ia1breijtest, ia1breispest, ia1breiptest	IA47B-IA49B RANGE OF COMB INCOME FROM REL ESTAT	24.4
37	lf1ernfrmwrk	R1 LF10 AMT EARN FRM WRK LST MTH	5.3
38	lf1huwpaearn	R1 LF14 HUS/WIFE/PAR PAY LST MTH	39.4

Table 3. Auxilliary Variables

	Variable name	Label	% Missing
1	hh1martlstat	R1 HH1 MARITAL STATUS	0.1
2	pa1workfrpay	R1 PA17 EVER WORK FOR PAY	0.1
3	el1higstschl	R1 EL10 HGHST DGREE SCOOL COMPLD	1.3
4	rl1hisplatno	R1 RL3 CNSDR YRSF HSPAN OR LATNO	1.2
5	rl1yourrace1	R1 RL1 RACE OF SP WHITE	1.1
6	rl1yourrace2	R1 RL1 RACE OF SP AFRICN AMERICN	1.1
7	rl1yourrace3	R1 RL1 RACE OF SP AMERICN INDIAN	1.1
8	rl1yourrace4	R1 RL1 RACE OF SP ALASKA NATIVE	1.1
9	rl1yourrace5	R1 RL1 RACE OF SP ASIAN	1.1
10	rl1yourrace6	R1 RL1 RACE OF SP NATIVE HWAIIAN	1.1
11	rl1yourrace7	R1 RL1 RACE OF SP PACIFIC ISLNDR	1.1
12	rl1yourrace8	R1 RL1 RACE OF SP OTHER SPECIFY	1.1
13	va1serarmfor	R1 VA1 SERVED IN ARMED FORCES	1.1
14	va1memnatgrd	R1 VA3 MEMBER OF NATIONAL GUARD	0.0
15	lf1mrthnonjb	R1 LF4 MOR THN ONE JOB LAST WEEK	0.3
16	lf1hrswkwork	R1 LF5 HRS PR WEEK WORK MAIN JOB	1.0
17	lf1hrwrkltwk	R1 LF6 HOURS WORK LAST WEEK	0.6
18	lf1hrwrklstw	R1 LF7 HOW MNY HOURS DID YOU WRK	1.9
19	lf1oftpaid	R1 LF8 HOW OFTN PAID ON MAIN JOB	1.1
20	lf1lstpaychk	R1 LF9 HOW MUCH LAST PAYCHECK	19.8
21	hp1ownrentot	R1 HP1 OWN RENT OR OTHER	1.6
22	ia1howrecessp	R1 IA2 HOW REC SOC SEC PAY	0.6
23	ia1msrtrecss	R1 IA3A MTH STRD REC SOC SEC PAY	29.3
24	ia1yrstressp	R1 IA3B YR STRTD REC SOC SEC PAY	13.0
25	ir1areacond1	R1 IR15 LITTER GLASS ON SDWLK ST	1.2
26	ir1areacond2	R1 IR15 GRAFFITI ON BUILDG WALLS	1.2
27	ir1areacond3	R1 IR15 VACANT HOUSES OR STORES	1.2
28	ir1condhome1	R1 IR16 BROKEN WINDOWS IN HOME	1.2
29	ir1condhome2	R1 IR16 CRUMBLNG FOUNDTN IN HOME	1.2
30	ir1condhome3	R1 IR16 MISSNG BRCKS SIDNG IN HM	1.2
31	ir1condhome4	R1 IR16 ROOF PROBLEM IN HOME	1.2
32	ir1condhome5	R1 IR16 BROKEN STEPS TO HOME	1.2
33	fl1facility	R1 F SP IN FACILITY	0.0
34	sex	HISKEW GENDER	0.0
35	agecat	HISKEW AGE CATEGORY	0.0
36	rtirace	HISKEW RACEETH, 3-CATEGORY	0.0
37	per_cap_inc_5yr	PER CAPITA INCOME [WT1; ACS]	0.1
38	hh1livwthspo	R1 HH11 LIVE WITH SPOUSE PARTNER	0.1
39	hh1placekind	R1 HH12 KIND OF PLACE LIVE IN	2.2