ReadMe file for Data and Programs for
"Estimating the Impacts of Program Benefits: Using Instrumental Variables with Underreported and Imputed Data"

by Melvin Stephens Jr and Takashi Unayama

July 2, 2018

This file describes how to use the included data and programs to replicate the results in "Estimating the Impacts of Program Benefits: Using Instrumental Variables with Underreported and Imputed Data".

The Impact of Social Security Benefits on Shared Living Arrangements

The programs for this analysis were executed using Stata 14 on Mac OS.

We use the March Current Population Survey data from 1980-2013. There are three sources of the data that we use. For 1980-1987, we acquire the data from IPUMS-CPS (https://cps.ipums.org/cps/).

Beginning in 1988, the CPS restructured the methods it used for imputing missing values. As part of this restructuring, the CPS created a variable, FL-665, which indicates whether the household completed the March CPS supplement in addition to the main questionnaire. However, IPUMS-CPS currently does not provide FL-665.

We acquire the 1988-2013 March CPS data from the NBER CPS data web page (http://www.nber.org/cps/). However, the variable FL-665 was not provided as part of the original releases of the 1988-1990 March CPS. We obtain this variable for these three years from https://www.census.gov/data/datasets/time-series/demo/income-poverty/data-extracts.html.

We use the entire sample period to construct Figure 1 while, following the sample period used by Englehardt, Gruber, and Perry (2005) (EGP), the analysis in Figure 2 and Table 1 is limited to 1980-1999.

We first pool the individual data from 1980-2013, harmonizing the data across all years to be consistent with the coding used by the IPUMS-CPS for variables such as race, education, and marital status. The included program ext_ss.do contains the coding for this step. This program assumes that the FL-665 variable has already been merged onto the CPS data for 1988-1990. Neither the original CPS data nor the file created by executing ext_ss.do is found here. 1960 Census state codes are mapped into FIPS codes using the data found in state_fips_1960census.dta.

The included program figure1.do uses the dataset created by running ext_ss.do to create an additional dataset, ss_figure1_data.dta, which is included here and is used to construct Figure 1.

We then follow the methods described in EGP to create households and the variables used in the analysis including the indicator for shared living arrangements. We also merge the EGP instruments to these data in this step. The included program ss_example_data.do contains the coding for the step. The dataset resulting from this step, cps_ss_final.dta, is included here.

The included program figure2.do uses cps_ss_final.dta to create an additional dataset, ss_figure2_data.dta, which is included here and is used to construct Figure 2.

The included program ss_example_figures.do creates Figures 1 and 2. The included program ss_example_analysis.do yields the parameter estimates shown in Table 1.

Here is a list of the variables found in cps_ss_final.dta:

year – year of CPS survey
wtsupp – CPS sampling weight for the March supplement
ageh – age of the family head
yobh – year of birth of the family head
educ0 – education of the family head
(1 – <High school, 2 – High school grad, 3 – Some college, 4 – College grad)
marstnew – marital status of the household head
(1 – Married/married spouse absent, 4 – Separated/Divorced,
5 – Widowed, 6 – Never married/Single)
agespouse – age of the spouse (if present)
hsgradsp – spouse is a H.S. grad (if present)
somecollsp – spouse attended some college (if present)
collgradsp – spouse is a college grad (if present)
division – Census Divisions (nine)
sharlive – shared living arrangements indicator
(1 – lives with more adults than just self and spouse (if present), 0 – otherwise)
incsstotcpi – Reported Social Security income (in 1982-1984 thousands of dollars)
inst – Engelhardt, Gruber, and Perry instrument (in 1982-1984 thousands of dollars)
qincssnew – flag for imputation status of incsstotcpi
(0 – not imputed, 1 – amount imputed,
2 – recipiency imputed (1982 and earlier only),
3 – whole March supplement imputed)

Here is a list of the variables found in ss_figure1_data.dta:

year – CPS survey year
ss_impute – share of CPS observations with imputed Social Security benefits
wc_impute – share of CPS obs. with imputed worker's compensation benefits
uc_impute – share of CPS obs. with imputed unemployment compensation benefits
ssi_impute – share of CPS observations with imputed Supplement Security Income benefits
paw_impute - share of CPS obs. with imputed public assistance/welfare benefits
wage_impute - share of CPS obs. with imputed earnings
fl665_impute - share of CPS obs. with entire March supplement imputed

Here is a list of the variables found in ss_figure2_data.dta:

yobh – Year of birth of household head
inst - Engelhardt, Gruber, and Perry instrument (in dollars)
ssinc_impute – Average Social Security (SS) income for those with imputed SS benefits
ssinc_nonimpute – Average SS income for those with non-imputed SS benefits

Here is a list of the variables found in state_fips_1960census.dta:

state – name of state
census_code – 1960 state census code
fips_code – state fips code
area_name – name of state (in fips_code match file)

The Impact of Japanese Child Benefits on Consumption

1. How to Obtain the Japanese Family Income and Expenditure Survey (JFIES) Data

The programs listed below require microdata from Japanese Family Income and Expenditure Survey (JFIES). The survey datasets are the property of Statistical Bureau of Japan (SBJ) and are not publicly available. They were made available to the authors with the permission based on Article 33 of Japanese Statistics Act (Act No. 18 of 1947). It would be basically available for academic researchers, while application and other documentation should be written in Japanese.

2. Data Format and Dictionary files

Once a researcher is permitted to use the data, a fixed-length format file will be given. The length of file would be same for all researchers but variables not appeared in the application are masked and so contents of file are different depending on their research plan. Masked variablea are filled with blank. To use the fixed-length format data, we have converted the original data into STATA format using “infile” command with dictionary files given here.

3. Used Software

We used STATA version 15 on MAC OS.

4. STATA Code

Using our STATA code, the directory structure should be as below:

"Woking Directory" ----- “ORG_JFIES” (Original files given by SBJ)
|
|
----Dictionary (Dictionary files we made)

Three .do files are in "Working Directory" and contents of each file are below

(a) Stephens_Unayama.do: This is the main file for data constructions and regressions.
(b) Make_Sub_ID.do: This is a sub-routine to make household ID (see Unayama, 2011).
(c) Make_Calculating_Benefit.do: This is a sub-routine to calculate the imputed amount of child benefits (see Stephens and Unayama, 2015).

"Working Directory" also contains "CPI.dta" and "Calendar.dta", in which the Japanese Consumer Price Index and calendar information is available, respectively. They are merged with the main JFIES data.

5. Results file

The regression results are provided in “Results_RESTAT_Final.txt”.

References

Engelhardt, Gary V., Jonathan Gruber, and Cynthia D. Perry (2005) ``Social Security and Elderly Living Arrangements: Evidence from the Social Security Notch," *The Journal of Human Resources*, 40(2):354-72.

Unayama, Takashi (2011) “Family income and expenditure survey as a panel data: Construction of permanent household ID” (Kakei Chosa no Panel-ka ni-tsuite), Kokumin Keizai Zasshi Vol. 204, pp. 51-64. (http://www.lib.kobe-u.ac.jp/repository/81008361.pdf). IN JAPANESE

Melvin Stephens Jr & Takashi Unayama (2015) "Child benefit payments and household wealth accumulation," *The Japanese Economic Review* vol. 66(4) pp. 447-465.