Generating calibrated weights
SHARE User Guide Release 1

November, 29th, 2013

Content

1 Introduction

This user guide describes how to generate calibrated cross-sectional and longitudinal weights both at the individual and at the household level with the available SHARE data.

The following section gives an overview over the provided Stata files. Please note that the programs to generate calibrated weights are so-called do-files which require the statistical software package Stata installed on your computer. The programs have been written using Stata 12 - we do not guarantee that they work properly with earlier Stata versions.
Basic information about sampling design, weights and calibrated weights released in the SHARE public release data can be found in the SHARE Release Guide 1.1.1 Wave 4 (http://www.share-project.org/fileadmin/pdf_documentation/SHARE_wave_4_ release_guide_1.1.1.pdf).

For any questions related to the construction of calibrated weights involving the SHARE data please write to Giuseppe De Luca (giuseppe.deluca@unipa.it), our SHARE weighting expert, and please always CC info@share-project.org.

This way we can continuously update this guide with a frequently asked questions (FAQs) section in the future.

2 Data files

The SHARE Calibrated Weights (Release 1) data are distributed in 14 files including three different file types. The files are listed below along with the corresponding file type:

File name	File type
cweightopt	ado
cweights	ado
nlsyseq	ado
data_w4_hhs_scr	dta
data_w4_ind_scr	dta
data_w24_hhs_scr	dta
data_w24_ind_scr	dta
1_CSW4_IND	do
2_CSW4_HHS	do
3_LW24_IND	do
4_LW24_HHS	do
Mortality	do
Population	do
Region_nuts1	do

Files with the extension .ado are "automatically loaded" do-files which typically contain code for Stata commands that are not already included in Stata. The core of the weighting procedure is carried out by the Stata ado files listed above: "cweightopt.ado", "cweights.ado" and "nlsyseq.ado". Users should <u>not</u> modify these programs. These files have to be placed in the respective ado file directory on your computer. Type `sysdir` in the Stata command line to set or query system directories (type `help sysdir` if you are not familiar with this command). Alternatively, these files be can stored in any folder on your computer and then be loaded in Stata by using the `adopath +` "folder directory" command (see the example do-files).

Stata datasets (with the extension dta) provide example datasets on how your data should look like when you want to generate weights. For example, the dataset "data_w4_hhs.dta" consists of all necessary variables from SHARE modules that a user will need in order to construct calibrated weights for wave 4 on household level (hhs). In general, users are required to construct their own dataset by merging the needed modules of SHARE. The dataset has to contain all relevant information needed to compute the required longitudinal weights (e.g. sample and wave indicators, original sampling design weights, calibration variables such as gender or year of birth, indicators for NUTS 1 region, etc.). See in addition to the example datasets table 1 in the appendix which provides an overview of the variables needed for the different weight types. Note that variables in the example datasets have been renamed, e.g. "yrbirth" from the "cv_r" module of SHARE wave 4 is renamed to "yrbirth_w4", etc.

Moreover, some variables have to be generated by your own. In order to compute longitudinal weights, a dummy variable is needed which indicates whether a household (or respondent at the individual level) has participated in all waves the longitudinal weight is intended to cover. The dataset "data_w24_hhs.dta" for example includes the indicator variable "w_2_4_hhs" which takes the value one for all households participating in SHARE

wave 2 and wave 4 and zero otherwise. The variable "region" in the example datasets will be generated by the Region_nuts1.do-file (more on ancillary do-files below). Note that the necessary NUTS 1 information has to be included in your data in order for the "Region_nuts1.do-file" to work. NUTS 1 information can be found in the "gv_housing" modules in the SHARE public release data. Be aware that NUTS 1 regional data are only available for the first wave a respondent participates in SHARE. For panel cases the regional information has to be merged from previous waves (see the respective SHARE Release Guides for additional and more detailed information on this).

Do-files starting with numbered labels illustrate the core of the weighting procedure. The example files for generating cross-sectional weights are focusing on Italy. Cross-sectional weights for other SHARE countries can be computed in a similar fashion. The example files for generating cross-sectional and longitudinal calibrated waves include a country loop structure which allows generating weights for more than one country at once.

1_CSW4_IND.do	Cross-sectional wave 4 individual level
2_CSW4_HHS.do	Cross-sectional wave 4 household level
3_LW24_IND.do	Longitudinal wave2-4 individual level
4_LW24_HHS.do	Longitudinal wave 2-4 household level

Every example do-file listed above contains a detailed description of the general weighting process and is very well commented. Please work through the example files first before you construct your own weights.

Additional external information about the target population needed in the weighting procedure is provided by ancillary do-files:

Region_nuts1.do	provides coding of the indicators for NUTS 1 geographical area
Population.do	information about population size by age-gender group and NUTS 1 code (from Eurostat)
mortality.do	data on death of the population 50+ by period, country, age-gender group and NUTS 1 regional indicator (from Eurostat)

Crucial for the weighting procedure to work is that these do-files are available in your working directory, as the example do-files mentioned above will need them in order to work properly.

The ancillary do-files will be called automatically from within the example do-files which create the calibrated weights, e.g. "1_CSW_IND.do". In the <u>cross-sectional</u> weighting files, the user only needs to adjust the data and ado paths and has to decide for which countries and wave the cross-sectional calibrated weights should be computed.

When constructing <u>longitudinal</u> weights, the syntax in the Stata example do-files referring to a specific wave or wave combinations has to be adjusted manually.
The name suffixes of the necessary information about the size of the target population and mortality by calibration group correspond to the year the respective wave was conducted. In the example do-file "3_LW24_IND.do" (i.e. Longitudinal wave 2-4 at the individual level) the matrix `cc'_Pop_06_10 refers to the year 2006 i.e. SHARE wave 2 and the year 2010 i.e.

SHARE wave 4 for country `cc'. In almost the same manner, the matrix `cc'_Dea_age_sex_07_10 provides mortality data between 2007 and 2010 i.e. between SHARE wave 2 and SHARE wave 4.
See table 2 in the appendix for an overview of when each SHARE wave was conducted.

In general, <u>all</u> locals, <u>all</u> globals, <u>all</u> matrices and <u>all</u> variables corresponding to certain years or waves need to be adjusted in order to generate calibrated longitudinal weights. This also includes, for example, the participation indicator variable "w_2_4_hhs" in the example do-file for longitudinal calibrated weights at the household level.
Take an example do-file for longitudinal weights, e.g. "3_LW24_IND.do", as template, check each syntax line carefully and adjust the code accordingly.

Furthermore, your panel dataset (for which you plan to compute longitudinal weights) should include a household identifier which is fixed across waves. Currently the SHARE release data include only wave-specific household identifier (e.g. hhid4 in wave 4).
A fixed household identifier can be easily generated by extracting the first ten digits from the wave-specific household identifier with the following command:

```
generate str10 hhid_fix = hhidcom4
```

In contrast to the wave-specific household identifier, the fixed household identifier does not consider household splits.

3 Additional remarks

Note that the current version of the generated calibrated weights release does not yet support all SHARE countries. See the description at the beginning of the example do-files for additional information.

Computing generated calibrated weights for all available countries at once takes a long time, depending on the Stata version in use (SE or MP) and the allocated random access memory (RAM). If you are using a Stata version below Stata 12, try to allocate more memory with the `set memory` command in Stata.
In addition, the computing time can be substantially reduced in two ways:

1. Using a lower number of combinations for the upper and lower thresholds for the distance function. See the locals "ulist" and "llist" in Section C in the example do-files.
2. Reducing the number of calibration margins, for example by excluding the NUTS 1 calibration margins.

4 How-to in a nutshell

- decide what sort of weights you need: longitudinal vs. cross-sectional, which country and which wave(s)
- open the respective example dataset for the preferred weighting structure e.g. "data_w24_hhs.dta" if you want to generate longitudinal weights at the household level
- become familiar with the structure of the data and the variables you will need, then construct your own dataset by merging the respective SHARE data in need. See the SHARE Release Guide 2.6.0 and the SHARE Wave 4 Release Guide 1 to learn more on how to merge modules of the SHARE data. In addition see table 1 in the appendix for an overview which variables are needed for which weight type
- Please note that for longitudinal weights some variables need to be constructed by the user, see the indicator variable "w_2_4_ind" in the "data_w24_ind.dta"-dataset as an example
- when your data has the right shape, open a copy of the respective example do-file for your preferred weighting procedure as a working template e.g. "3_LW24_IND.do" if you want to generate calibrated longitudinal weights at the individual level
    - o adjust the do-file in the necessary sections to your data i.e. specify the path of your working directory and of your ado file folder, choose wave(s) and countries
    - o Note that for longitudinal weights, you have to <u>manually adjust</u> all locals, globals, matrices and variables referring to certain years or waves

5 Appendix

Table 1: Variables needed for generating calibrated weights

Variable name	Description	Time dimension	Level
mergeid	Person ID	cross-sectional/ longitudinal	individual
hhid<*wave number*>	Wave-specific household ID (e.g. hhid4 for wave 4)	cross-sectional/ longitudinal	individual / household
split	Household split identifier	cross-sectional/ longitudinal	individual / household
hhid_fix	Household ID fixed across waves	cross-sectional/ longitudinal	individual / household
country	Country code	cross-sectional/ longitudinal	individual / household
waveid	Identifier of original wave	cross-sectional/ longitudinal	individual / household
wabeid_hh	Identifier of wave in which hh was sampled	cross-sectional/ longitudinal	individual / household
DW_W<*wave number*>	Design weight – wave <*wave number*>	cross-sectional/ longitudinal	individual / household
DW_w<*wave number*>_f	Design weight – wave <*wave number*> - Flag	cross-sectional/ longitudinal	individual / household
region	Indicator for NUTS 1 code	cross-sectional/ longitudinal	individual / household
gs_nuts1	NUTS 1 code (NUTS 1 variable from the gv_housing module)	cross-sectional/ longitudinal	individual / household
w_<wave number>_<wave number>_hhs	Indicator for households participating in waves <wave number>_<wave number>	longitudinal	household
w_<wave number>_<wave number>_ind	Indicator for respondents participating in waves <wave number>_<wave number>	longitudinal	individual
gender_w<wave number>	Gender	cross-sectional/ longitudinal	individual
yrbirth_w<wave number>	Year of birth	cross-sectional/ longitudinal	individual
yrbirthp_w<wave number>	Year of birth of spouse/partner	cross-sectional/ longitudinal	individual
gender_<loop number>	Gender of hh member <loop number>	cross-sectional/ longitudinal	household
yrbirth_<loop number>	Year of birth of hh member <loop number>	cross-sectional/ longitudinal	household

Table 2: Overview: Countries and Language Versions in SHARE Waves 1-4

ID	Country (Short)	Country	Wave 1	Wave 2	Wave 3 SHARELIFE	Wave 4
11	AT	Austria	2004	2006/07	2008/09	2011
12	DE	Germany	2004	2006/07	2008/09	2011/12
13	SE	Sweden	2004	2006/07	2008/09	2011
14	NL	Netherlands	2004	2007	2008/09	2011
15	ES	Spain	2004	2006/07	2008/09	2011
16	IT	Italy	2004	2006/07	2008/09	2011
17	FR	France	2004/05	2006/07	2009	2011
18	DK	Denmark	2004	2006/07	2008/09	2011
19	GR	Greece	2004/05	2007	2008/09	-
20	Cg	Switzerland (German)	2004	2006/07	2008/09	2011
21	Cf	Switzerland (French)	2004	2006/07	2008/09	2011
22	Ci	Switzerland (Italian)	2004	2006/07	2008/09	2011
23	Bf	Belgium (French)	2004/05	2006/07	2008/09	2011
24	Bn	Belgium (Flemish)	2004/05	2006/07	2008/09	2011
25	Ih	Israel (Hebrew)	2005/06	-	-	-
26	Ia	Israel (Arabic)	2005/06	-	-	-
27	Ir	Israel (Russian)	2005/06	-	-	-
28	CZ	Czech Republic	-	2006/07	2008/09	2011
29	PL	Poland	-	2006/07	2008/09	2011/12
30	IE	Ireland	-	2007	-	-
32	HU	Hungary	-	-	-	2011
33	PT	Portugal	-	-	-	2011
34	SI	Slovenia	-	-	-	2011
35	EE	Estonia	-	-	-	2010/11