PSID-Tools: An interface to the the Panel Study of Income Dynamics

Ulrich Kohler

ulrich.kohler@uni-potsdam.de

Methods for Empirical Social Research
Faculty for Economics and Social Sciences
University of Potsdam

13th Stata Users Group Meeting
Nuremberg June 26th 2015

Structure

Inhalt

What is the PSID?

- World's longest running panel study
- Starting 1968
- Annual data collection until 1997
- Biennial data collection since 1997
- Income related data, predominately on Family/Household level
- Individual data about Head of Household, proxies about partners
- Some individual data about all household members

Hurdles for using the PSID

- Data delivery in (many) separate ASCII-Text files
- Lots of variables
- No structured variable name model
- No value labels

PSID-Tools help dealing with these problems.

PSID data structure

- Main data

  `http://simba.isr.umich.edu/Zips/ZipMain.aspx`

- Supplemental data

  `http://simba.isr.umich.edu/Zips/zipSupp.aspx`

- User generated files

  `http://simba.isr.umich.edu/help/UgenVars.aspx`

Inhalt

Inhalt

2 PSID-Tools

- Overview
- psid install
- psid use/psid add
- psid long
- psid vardoc

Overview

As it stands, there are 5 PSID-Tools:

psid install Creates Stata datasets from the packed PSID files.

psid use Loads items from several waves of the PSID or CNEF into memory

psid add Merges items from several waves of the PSID/CNEF to a PSID/CNEF file in memory

psid long Makes data long (and keep the labels)

psid vardoc Shows "official" variable description and create value labels

Inhalt

2 PSID-Tools

Short summary

- `psid install` installs Stata datasets of the PSID or the CNEF
- Installation is fully automatic for the CNEF
- PSID data download requires registration: Thus, zip-files have to be downloaded using a browser.
- Syntax:

  psid install [*wavelist*] [using *dirname*]
  [, to(*dirname*) cnef replace upgrade <u>l</u>ower clean
  replacelong replacesingle longonly]

Examples

. `psid install using c:/downloads`

- unpacks the downloaded PSID zip-files in the directory c:/downloads,
- makes some necessary edits in the unpacked Stata do-files,
- runs the do-files that create the Stata datasets,
- applies function `proper()` to variable labels
- compresses the created Stata data sets,
- saves the Stata file into the current working directory, and
- erases all files generated on the fly.

. `psid install, cnef`

installs the CNEF in the current working directory.

Aside

Aside but very important:

- Cross year individual file (`indCCYY.dta`) must be installed to take advantage of the other programs of PSID-Tools.
- This remains true even if you do not use variables from `indCCYY.dta`
- This remains true even if you think that this is bad!

Inhalt

2 PSID-Tools

- Overview

- psid install

- psid use/psid add

- psid long

- psid vardoc

Short summary

- `psid use` and `psid add` create data sets from user input copy-pasted from the PSID Data Center
- `psid use` creates a new data set
- `psid add` merges variables to a PSID file in memory
- Syntax:

  ```
  . psid use || newstub varspecs using dirname
  [, dofile(filename, ...) ftype(cnef|wealth) keepnotes
  lower waves(wavelist) design(designtype) clear ]
  . psid add || newstub varspecs
  [, dofile(filename, ...) ftype(cnef|wealth) keepnotes
  lower waves(wavelist) ]
  ```

PSID data center

- https://simba.isr.umich.edu/VS/s.aspx
- Search facility for variables
- Provides (among others) item-correspondance-lists:

Print

V745		**HEALTH LIM HSEWRK**
H52c. Does your health limit the work you (HEAD) can do around the house? (1969)		

Count	%	Value/Range Text
605	13.57	1 Yes
94	2.11	3 Yes, qualified
3,741	83.88	5 No
20	.45	9 NA

Years Available:	[69]V745 [70]V1411 [71]V2123
Index Summary:	Family Public Data Index 01>HEALTH STATUS 02>Physical Health 03>disability 04>disability now, whether limits work around home (head):

Small example

```
. psid use
   || health
   [69]V745 [70]V1411 [71]V2123
   using ~/data/PSID, clear keepnotes

. describe

Contains data from ~/data/PSID/ind2013er.dta
  obs:        15,090
 vars:             10                          4 Jun 2015 08:32
 size:        241,440

              storage   display    value
variable name   type    format     label    variable label

x1110lll      long      %12.0g              Person identification number
x11102_1969   int       %9.0g               1969 Interview Number
x11102_1970   int       %9.0g               1970 Interview Number
x11102_1971   int       %9.0g               1971 Interview Number
xsqnr_1969    byte      %9.0g               Sequence Number 69
xsqnr_1970    byte      %9.0g               Sequence Number 70
xsqnr_1971    byte      %9.0g               Sequence Number 71
health1969    byte      %8.0g               Health Lim Hsewrk
health1970    byte      %8.0g               Health Lim Hsewrk
health1971    byte      %8.0g               Health Limit Hswrk

Sorted by:
     Note: Dataset has changed since last saved.
```

Big example

```
                        ../projects/WZB-PSID-File/wzbpsid_cr01V3_retrival.do
  ✄
  11:  psid use  ///
  12:    /// date of interview-head
  13:    || int01V1  ///
  14:    [68]V99 [69]V553 [70]V1236 [71]V1939 [72]V2539 [73]V3092 [74]V3505 ///
  15:    [75]V3918 [76]V4433 [77]V5347 [78]V5847 [79]V6459 ///
  16:    [80]V7064 [81]V7655 ///
  17:    [82]V8349 [83]V8958 [84]V10416 [85]V11600 [86]V13008 [87]V14111 ///
  18:    [88]V15127 [89]V16628 [90]V18046 [91]V19346 [92]V20648 [93]V22403 ///
  19:    [94]ER2005 [95]ER5004 [96]ER7004 ///
  20:    [97]ER10005 [99]ER13006 [01]ER17009 [03]ER21012 [05]ER25012 ///
  ✄
1121:    /// Year last released
1122:    || crm05 [95]ER33271 ///
1123:    /// Times sent to adult corrections institutions
1124:    || crm06 [95]ER33269 ///
1125:    /// Times sent to youth corrections institutions
1126:    || crm07 [95]ER33268 ///
1127:    /// Times FU was asked to talk about school behavior
1128:    || crm08 [95]ER33272 ///
1129:    || strat []ER31996 ///
1130:  using $psid , design(any) clear keepnotes
                        ../projects/WZB-PSID-File/wzbpsid_cr01V3_retrival.do
```

Selected options

`psiduse`/`psidadd` rename variables of each item-corresponding-list using a common variable name model.

This is convenient, but there is more:

`design(designtype)`   specifies the design of the panel data to be created:

- `design(balanced)` creates a balanced panel design (default)
- `design(any)` keeps all observations.
- `design(#)` keeps households interviewed # times or more.

`dofile(filename)`   Creates Do-file *filenname.do* of the retrieval.

`keepnote`   Keep notes of the original variable names (pre-requisite for `psid vardoc`)

Inhalt

2 PSID-Tools

- Overview
- psid install
- psid use/psid add
- psid long
- psid vardoc

short summary

psid long makes data long.

- Wrapper for reshape
- Keeps variable label (... of first variable of a item-correspondance list)
- Deals with information left behind by the keepnotes-option of psid use/psid add

psid long have no further options. Users are requested to use reshape for doing something special.

Example

```
. psid long
(note: j = 1969 1970 1971)
Data                                    wide  ->  long

Number of obs.                         15090  ->   45270
Number of variables                       10  ->       5
j variable (3 values)                        ->    wave
xij variables:
      x111021969 x111021970 x111021971  ->  x11102
          xsqnr1969 xsqnr1970 xsqnr1971  ->  xsqnr
      health1969 health1970 health1971  ->  health

. describe
Contains data
  obs:         45,270
  vars:             5
  size:       452,700

            storage   display    value
variable name  type    format    label    variable label

x1110111       long    %12.0g             Person identification number
wave           int     %9.0g
x11102         int     %9.0g              1969 Interview Number
xsqnr          byte    %9.0g              Sequence Number 69
health         byte    %8.0g              Health Lim Hsewrk

Sorted by: x1110111  wave
    Note: Dataset has changed since last saved.
```

Inhalt

2 PSID-Tools

- Overview
- psid install
- psid use/psid add
- psid long
- psid vardoc

Short summary

`psid vardoc` displays the official variable description of the PSID Data Center in a browser window.

- Only works for datasets created with `psid use`/`psid add` along with option `keepnotes`
- For long datasets: Requires dataset to be created with `psid long`.
- Requires connection to the internet
- Syntax:

  `psid vardoc` *varname*
  [, <u>addvaluelabel</u>(*lblname*) <u>s</u>how itemnum(*#*)]

Example

```
. psid vardoc health
```

Example with addvaluelabel

```
. tab health
Health Lim |
    Hsewrk |      Freq.     Percent        Cum.
-----------+-----------------------------------
         1 |      5,687       12.56       12.56
         3 |        620        1.37       13.93
         5 |     38,745       85.59       99.52
         9 |        218        0.48      100.00
-----------+-----------------------------------
     Total |     45,270      100.00

. psid vardoc health, addvaluelabel(emp)
Label emp attached to health and defined as:
emp:
           1 Yes
           3 Yes, qualified, minor limitation
           5 No
           9 NA, DK, age

. tab health
               Health Lim Hsewrk |      Freq.     Percent        Cum.
---------------------------------+-----------------------------------
                             Yes |      5,687       12.56       12.56
 Yes, qualified, minor limitation |        620        1.37       13.93
                              No |     38,745       85.59       99.52
                      NA, DK, age |        218        0.48      100.00
---------------------------------+-----------------------------------
                           Total |     45,270      100.00
```

Disclaimer

- Variables of an item-correspondence list may not have identical value labels accross time
- It is the users' responsibility to check whether psid vardoc, addvaluelabels is correct in long data.

. . . thanks for listening