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

1 About the PSID

2 PSID-Tools
   - Overview
   - psid install
   - psid use/psid add
   - psid long
   - psid vardoc
Inhalt

1. About the PSID

2. PSID-Tools
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](http://simba.isr.umich.edu/Zips/ZipMain.aspx)

- **Supplemental data**
  

- **User generated files**
  
  [http://simba.isr.umich.edu/help/UgenVars.aspx](http://simba.isr.umich.edu/help/UgenVars.aspx)
Inhalt

1 About the PSID

2 PSID-Tools
Inhalt

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

- Overview
- psid install
- psid use/psid add
- psid long
- psid vardoc
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 lower 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!
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 vars specs using dirname
[ , dofile(filename, ...) ftype(cnef|wealth) keepnotes
  lower waves(wavelist) design(designtype) clear ]
.
. psid add || newstub vars specs
[ , 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:

<table>
<thead>
<tr>
<th>Count</th>
<th>%</th>
<th>Value/Range Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>605</td>
<td>13.57</td>
<td>1 Yes</td>
</tr>
<tr>
<td>94</td>
<td>2.11</td>
<td>3 Yes, qualified</td>
</tr>
<tr>
<td>3,741</td>
<td>83.88</td>
<td>5 No</td>
</tr>
<tr>
<td>20</td>
<td>.45</td>
<td>9 NA</td>
</tr>
</tbody>
</table>

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
   size: 241,440

<table>
<thead>
<tr>
<th>variable name</th>
<th>storage</th>
<th>type</th>
<th>display</th>
<th>value</th>
<th>variable label</th>
</tr>
</thead>
<tbody>
<tr>
<td>x11101ll</td>
<td>long</td>
<td>%12.0g</td>
<td></td>
<td></td>
<td>Person identification number</td>
</tr>
<tr>
<td>x11102_1969</td>
<td>int</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>1969 Interview Number</td>
</tr>
<tr>
<td>x11102_1970</td>
<td>int</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>1970 Interview Number</td>
</tr>
<tr>
<td>x11102_1971</td>
<td>int</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>1971 Interview Number</td>
</tr>
<tr>
<td>xsqnr_1969</td>
<td>byte</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>Sequence Number 69</td>
</tr>
<tr>
<td>xsqnr_1970</td>
<td>byte</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>Sequence Number 70</td>
</tr>
<tr>
<td>xsqnr_1971</td>
<td>byte</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>Sequence Number 71</td>
</tr>
<tr>
<td>health1969</td>
<td>byte</td>
<td>%8.0g</td>
<td></td>
<td></td>
<td>Health Lim Hsewrk</td>
</tr>
<tr>
<td>health1970</td>
<td>byte</td>
<td>%8.0g</td>
<td></td>
<td></td>
<td>Health Lim Hsewrk</td>
</tr>
<tr>
<td>health1971</td>
<td>byte</td>
<td>%8.0g</td>
<td></td>
<td></td>
<td>Health Limit Hswrk</td>
</tr>
</tbody>
</table>

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 ///
16:  [80]V7064 [81]V7655 ///
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 *filename.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

\texttt{psid long} makes data long.

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

\texttt{psid long} have no further options. Users are requested to use \texttt{reshape} for doing something special.
Example

```
. psid long
Data wide -> long

Number of obs. 15090 -> 45270
Number of variables 10 -> 5
j variable (3 values) -> wave
xij variables:
x111021969 x111021970 x111021971 -> x11102
```

Subcommand: describe

Contains data

| obs: 45,270 | vars: 5 | size: 452,700 |

<table>
<thead>
<tr>
<th>variable name</th>
<th>type</th>
<th>format</th>
<th>label</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1110111</td>
<td>long</td>
<td>%12.0g</td>
<td>Person identification number</td>
</tr>
<tr>
<td>wave</td>
<td>int</td>
<td>%9.0g</td>
<td>1969 Interview Number</td>
</tr>
<tr>
<td>x11102</td>
<td>int</td>
<td>%9.0g</td>
<td>1969 Interview Number</td>
</tr>
<tr>
<td>xsqnr</td>
<td>byte</td>
<td>%9.0g</td>
<td>Sequence Number 69</td>
</tr>
<tr>
<td>health</td>
<td>byte</td>
<td>%8.0g</td>
<td>Health Lim Hsewrk</td>
</tr>
</tbody>
</table>

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`
  [ , `addvaluelabel(lblname)` `show itemnum(#)` ]
  ```
Example

```
. psid vardoc health
```
### Example with `addvaluelabel`

```
.tab health

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5,687</td>
<td>12.56</td>
<td>12.56</td>
</tr>
<tr>
<td>3</td>
<td>620</td>
<td>1.37</td>
<td>13.93</td>
</tr>
<tr>
<td>5</td>
<td>38,745</td>
<td>85.59</td>
<td>99.52</td>
</tr>
<tr>
<td>9</td>
<td>218</td>
<td>0.48</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Total | 45,270 | 100.00 |
```

```
.psid var doc 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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>5,687</td>
<td>12.56</td>
<td>12.56</td>
</tr>
<tr>
<td>Yes, qualified, minor limitation</td>
<td>620</td>
<td>1.37</td>
<td>13.93</td>
</tr>
<tr>
<td>No</td>
<td>38,745</td>
<td>85.59</td>
<td>99.52</td>
</tr>
<tr>
<td>NA, DK, age</td>
<td>218</td>
<td>0.48</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Total | 45,270 | 100.00 |
Disclaimer

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

...thanks for listening