

WERS 1998-2004 Panel Survey

Compiling a panel dataset

March 2007

The data that comprise the 1998-2004 Panel Survey are deposited in two separate data files which need to be linked prior to analysis. The WERS 2004 Information and Advice Service (<http://www.wers2004.info>) have compiled syntax for this purpose. This note accompanies the syntax files and provides further details regarding the construction of the panel dataset.

The first wave of data comes from the 1998 Cross-Section Survey and consists of observations on 2,191 workplaces with 10 or more employees; this data file has study number 3955 in the UK Data Archive's on-line catalogue. The second wave, conducted as part of WERS 2004, consists of a small number of data items for each of these 2,191 workplaces (essentially indicating its survival status and employment level in 2004) and a larger set of data items arising from repeat interviews at 938 of the surviving workplaces. These data are contained within the Panel Survey data file that can be found in the single WERS 2004 deposit at the UK Data Archive (study number 5294).

The data from 1998 and 2004 may be combined to form a single data file with either of two possible structures:

1. **One record per case (wide form).** In this format, each workplace occupies a single record (row) in the resulting data file. One block of variables contains data from 1998 and another block contains data from 2004. The two sets of variables might be distinguished by a different prefix (so **XASINGLE** might contain the data arising from the question named **ASINGLE** in 1998, whilst **YASINGLE** might contain the data from **ASINGLE** in 2004).
2. **One record per year (long form).** In this format, each workplace has two records in the data file. The first record contains data from 1998 and the second contains data from the same workplace in 2004. An index variable indicates the year to which each row relates. The number of variables is reduced by at least half in comparison with the first structure because: (a) the 1998 and 2004 values for consistently-defined data items can be contained within a single variable; and (b) questions or variables that are not consistent across the two years of the panel do not feature in the resulting data file.

A 1998-2004 panel dataset in wide form can be constructed from the deposited data files using either the `MATCH FILES` command in SPSS or the `-merge-` command in Stata. A panel dataset in long form can be constructed using the `VARSTOCASES` command in SPSS, or the `-reshape-` command in Stata.

The WERS 2004 Information and Advice Service (WIAS) have compiled syntax to perform both operations in either SPSS or Stata. The syntax requires SPSS v12 or Stata v7, or later versions of either software, as it creates some variables with names that exceed 8 characters. Note that the panel compilation syntax calls upon the files

Mq98fix.sps / Mq98fix.do. These files can be downloaded from the web-site of the WERS98 Data Dissemination Service (<http://www.niesr.ac.uk/research/wers98/>).

The steps that are followed in either package are essentially the same:

To construct a panel dataset in ‘wide form’:

1. Open the 1998 data file and fix known errors using the syntax compiled by the WERS98 Data Dissemination Service (see above). Then append the variables contained in the 2004 data file. As part of the matching process, variables containing 1998 data are prefixed with the letter X and those containing 2004 data are prefixed with the letter Y. Note that many variables will now have names that exceed 8 characters (permitted only in SPSS version 12.0 or later, and Stata version 7 or later).
2. Enforce consistency between the 1998 and 2004 blocks of variables in respect of the codes used for missing values: -1 for “Not asked” and -9 for “Don’t know or Not answered”. Note that the distinction between “Don’t know” and “Not answered” is sacrificed to simplify the assimilation process. A small number of variables with idiosyncratic missing value codes (e.g. **XXCODES**, **YXCODES**, **YEMPS2004**) are left untouched. [This step is omitted in Stata as there is no simple recode available]
3. Enforce consistency between the 1998 and 2004 blocks of variables in the names of otherwise equivalent variables. The practice is to make the 2004 data items consistent with 1998.
4. Save the new data file.

Note that in the syntax compiled by WIAS:

- The 2004 variables Xsample, Xoutcome, Xout3 and Xemps2004 are renamed to Ysample, Youtcome, Yout3 and Yemps2004 .
- Other variables already beginning with the letter 'x' receive the X or Y prefix as appropriate, e.g. Xxcode1, Yxcode1, Yxbtitle .
- The unique workplace id (serno) and the weighting variables from the 2004 data file (estwtmr, empwtmr, pqwtmr) do not receive a prefix .
- The original 1998 weight variables (est_wt, grosswt and emp_wt) are dropped since they have been superseded by estwtmr and empwtmr .

To construct a panel dataset in ‘long form’:

1. Follow the steps required to construct a ‘wide’ panel dataset.
2. Drop any workplaces that were not re-interviewed in 2004 (since those workplaces that were not interviewed have very limited information in 2004 and thus almost no consistent data items across the two years).
3. Drop any variables that are not entirely consistently across the two years (users may add further variables, after manipulating them to ensure consistency).
4. Remodel the data file to create a second row for each workplace, and populate this row with data from the 2004 block of consistent variables.
5. Create an index variable to identify the first row for each workplace as containing data from 1998 and the second as containing data from 2004.

Note that in the syntax compiled by WIAS:

- A new variable - YEAR - identifies the year to which each row of data relates.
- The following variables from 2004 are not included in the long dataset. Users will be able to 'recover' many of these by manipulating them to ensure consistency with 1998 (e.g. recoding Xauktot and Yauktot so that they have a consistent code list).

Ysoc2000, Yauktot, Yacomp01-12, Yahowch01-13, Ybmanage, Ycfactor1-
Yxcfact3, Ycspecia6 (empty), Ydbriefu1-Yxdbrief3, Ydappoin1-Ydappoin5,
Ydconsul1-Yxdconsu3, YevIEWS-Yxeviews, Yenew, Yenewnum,
Yewider, Yewidnum, Yeudrec, Yerequest, Yfvarpay6 (empty), Yfprofit,
Yfshar, Yfperfp, Yfmeasur1-Yxfmeasu3, Yfindper1-Yxfindpe4, Yfnmperf,
Yfnotpay, Yigroun10 (empty), Yifmoff, Yipatern, Yifamily8 (empty),
Yjnwtemp, Yjobsec, Yjtimear8 (empty), Ykmarket, Ykcompet, Yktarge10
(empty), Ykerfis, Ykoptb, Ykimp, Yktable, Ykdeti, Ymrel, Ymlinkdat,
Ymnextime

Further information and advice

If you have any questions about this note or the accompanying syntax files, please contact the WERS Information and Advice Service:

E-mail: wers2004@niesr.ac.uk
Phone: +44 (0) 20 7654 1933
Web-site: <http://www.wers2004.info>