Data and Programs for "Average and Quantile Effects in Nonseparable Panel Models"

Victor Chernozhukov, Iván Fernández-Val, Jinyong Hahn, and Whitney Newey

This supplement to the paper "Average and Quantile Effects in Nonseparable Panel Models" provides added details related to the data and the programs used in the numerical and empirical examples. The data sets come from the National Longitudinal Survey of Youth 1979 (Bureau of Labor Statistics, U.S. Department of Labor, 2002). The programs include the Stata and R code (R Development Team, 2007) used to construct Figures 1–4, A1, and A2, and Tables 1–3.

1 Data

The Stata data file *data_static_T8_long.dta* contains the data used in the union application of Section 10.1. These data consist of a sample of 2,065 full-time, young, working males, 20 to 29 years old in 1986, followed over the period 1986 to 1993. The data set contains the following variables

id :   individual identifier,

$time$:   time period, $1 = 1986, \ldots, 8 = 1993$,

y :   log-wage in \$1980,

x :   union status binary indicator, $= 1$ if unionized, $= 0$ otherwise,

tx :   type of individual with respect to history of x using binary coding, $1 + \sum_t 2^{x_t}$,

$tx1$:   type of individual with respect to union status in first period, $= 1$ if $x_1 = 1$,

$xf1$:   individual indicator for first period unionized,

$xf0$:   individual indicator for first period non-unionized,

$xl1$:   individual indicator for unionized in the last period,

$xl0$:   individual indicator for non-unionized in the last period.

The Stata data file *data_static.dta* contains the data used in the female labor force participation application of Section 10.2. These data consist of a sample of 1,587 married women, followed the years 1990, 1992, 1994, and 1996. The data set contains the following variables

id :   individual identifier,

$time$:   time period, $1 = 1990, \ldots, 4 = 1996$,

y :   labor force participation indicator, $1 = $ if in the labor force, $= 0$ otherwise,

x :   indicator for having any children younger than 3 years, $1 = $ if children, $= 0$ otherwise.

We only use the information for the years 1990, 1992, and 1994 in the analysis. The Stata data files *data_static_T2.dta* and *data_static_T3.dta,* created by the Stata program *table2.do,* label the sequences of x and y for each individual using binary coding.

2 Stata Command Files

The Stata command file *table2.do* includes the commands to generate Table 2, and produces the data files *data_static_T2.dta* and *data_static_T3.dta* used to compute Table 3.

3 R Command Files

We use five R command files in the numerical examples. The file *table1.R* includes the commands to generate Table 1, using the data set *data_static_T8_long.dta*. The files *table3logit.R* and *table3probit.R* include the commands to generate the columns "Logit" and "Probit" of Table 3, using the data sets *data_static_T2.dta* and *data_static_T3.dta*.[1] The file *figure1.R* includes the commands to generate Figure 1. The file *figures2&3.R* includes the commands to generate Figures 2 and 3. The files *figure4A&4D.R* and *figure4B&4C.R* include the commands to generate Figure 4, using the data set *data_static_T8_long.dta*. The file *figuresA1&A2* includes the commands to generate Figures A1 and A2 of the supplementary appendix.

For replication purposes, note that the programs *table3logit.R* and *table3probit.R* can take several days running. All the command files assume that the data sets are located in the current folder and generate output files in this folder. The files *table3logit.R*, *table3probit.R*, *figures2&3.R*, and *figuresA1&A2* use the R packages `quadprog` (S original by Berwin A. Turlach R port by A. Weingessel, 2011) and `lpSolve` (Berkelaar et al., 2011), which need to be installed.

References

[1] Bureau of Labor Statistics, U.S. Department of Labor (2002): *National Longitudinal Survey of Youth 1979 cohort, 1979-2002 (rounds 1-20) [computer file].* Produced and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH.

[2] M. Berkelaar et al. (2011): *lpSolve: Interface to Lp_solve v. 5.5 to solve linear/integer programs.* R package version 5.6.6, URL `http://CRAN.R-project.org/package=lpSolve`.

[3] R Development Core Team (2007): *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`.

[4] S original by Berwin A. Turlach R port by A. Weingessel (2011): *quadprog: Functions to solve Quadratic Programming Problems.* R package version 1.5-4, URL `http://CRAN.R-project.org/package=quadprog`.

---

[1]We do not provide the code to generate the rest of the columns, because they can be computed using standard software or code developed in the published papers cited in Section 10.2.