AGRODEP
African Growth & Development Policy modeling consortium
FACILITATED BY IFPRI

AGRODEP Stata Training

April 2013

Module 5

Panel Data Regressions

Manuel Barron[1] and Pia Basurto[2]

[1] University of California, Berkeley, Department of Agricultural and Resource Economics
[2] University of California, Santa Cruz, Department of Economics

Module 5 – Panel Data Regressions

In this last module we introduce commands useful for panel data analysis. We show how to tell Stata that the data are in longitudinal form (i.e., that it is a panel) with the *xtset* command. We then present random effects, fixed effects, and differences in differences. We also show how to use *outreg* in this context. In the appendix, we show how to generate the fictitious treatment variable used in the differences in differences estimation.

For this module we will use data from the World Bank's World Development Indicators (WDI). The countries_panel.dta and countries_panel_2.dta are downloadable from the AGRODEP website. The original source of the data is http://data.worldbank.org/data-catalog/world-development-indicators.

1. Setting the data as a Panel

Before you run panel data models, you need to tell Stata that the data are in panel form. This means you have observations for multiple individuals and multiple points in time. You need to tell Stata which variable identifies the individuals and which variable identifies the points in time. For this, you will use the *xtset* command followed by the individual id and by the time id, in that order.

***Do-file or command window**

```
help xtset
```

***Help file**

```
xtset panelvar timevar [, tsoptions]
```

In the syntax for *xtset, panelvar* is the variable that identifies the individuals (in our dataset "country_id"), and *timevar* is the variable that identifies the time periods (in our dataset "year"). There are other advanced options that will not be covered in these introductory notes (See *help xtset*).

***Do-file or command window**

```
use countries_panel, clear

xtset country_id year
```

If you want to check whether the data has already been *xtset*, type *xtset* with no options

***Do file or command window**

```
xtset
```

2. xtreg

The main Stata command for panel data regressions is called *xtreg*. You can use it to run fixed effects and random effects least-squares panel regressions, as well as other models. Remember that before using the *xtreg* command, you need to *xtset* the data as explained in the previous section.

2.1 Fixed Effects

To estimate a fixed effects model, use the *xtreg* command with the "fe" option:

***Do-file or command window**

```
help xtreg
```

***Help file**

```
xtreg depvar [indepvars] [if] [in] [weight] , fe [FE_options]
```

2.2 Random Effects

To estimate a random effects model, use the *xtreg* command with the "re" option (or you can also omit the "re"option, as it is the default option for *xtreg* as explained in the help file). In your do-file or command window, type:

*** Do-file or command window**

```
xtreg secondary immunizations it_bed_net,re
```

***Output window**

```
Random-effects GLS regression                   Number of obs      =        76
Group variable: country_id                      Number of groups   =        38

R-sq:  within  = 0.4093                         Obs per group: min =         1
       between = 0.0026                                        avg =       2.0
       overall = 0.0547                                        max =         4

                                                Wald chi2(2)       =     23.13
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
   secondary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
immunizati~s |   .2886739   .1433344     2.01   0.044     .0077435    .5696042
  it_bed_net |   .1873014   .0820088     2.28   0.022      .026567    .3480358
       _cons |   43.72713   10.61591     4.12   0.000     22.92033    64.53393
-------------+----------------------------------------------------------------
     sigma_u |  18.789034
     sigma_e |  8.1310123
         rho |  .84226469   (fraction of variance due to u_i)
------------------------------------------------------------------------------
.
```

4. Difference- in-Differences

We will illustrate how to run a difference-in-differences regression to explain the effect of a treatment intervention on progression to secondary school. All the data are real, except for the treatment variable (See the appendix at the end of this module for an explanation of how to generate this variable. Note that some commands that were used are not explained in these notes. For further explanation search for the appropriate help files). For illustration purposes, suppose that the countries with treatment=1 implemented a particular education program, and that countries with treatment=0 did not implement it. In this example, we will use the *xi* command to include interactions between explanatory variables.

*** Do-file or command window:**

```
xi: reg secondary i.treat*i.post
```

As we saw in module 3, this will include the dependent variable plus a dummy for treatment, a dummy for post, and the interaction of both explanatory variables. In this case, "post" is simply an indicator variable that takes the value 1 after a certain point in time, and the value 0 otherwise. Including this variable allows us to control for the difference in secondary school enrollment between these two periods of time for each country. The model can then be interpreted as a difference-in-differences regression.

*** Stata output**

```
i.treat           _Itreat_0-1         (naturally coded; _Itreat_0 omitted)
i.post            _Ipost_0-1          (naturally coded; _Ipost_0 omitted)
i.treat*i.post    _ItreXpos_#_#       (coded as above)

      Source |       SS       df       MS              Number of obs =     317
-------------+------------------------------           F(  3,   313) =   23.93
       Model |  32853.6977     3  10951.2326           Prob > F      =  0.0000
    Residual |  143236.438   313  457.624401           R-squared     =  0.1866
-------------+------------------------------           Adj R-squared =  0.1788
       Total |  176090.135   316  557.247264           Root MSE      =  21.392

------------------------------------------------------------------------------
   secondary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   _Itreat_1 |    4.43538   3.947698     1.12   0.262    -3.332001    12.20276
    _Ipost_1 |   21.79388   3.631637     6.00   0.000     14.64837    28.93939
_ItreXpos_~1 |  -1.888718   4.981064    -0.38   0.705    -11.68932    7.911883
       _cons |   57.54843   2.884518    19.95   0.000     51.87294    63.22393
------------------------------------------------------------------------------
```

5. Outreg

As shown in Module 3, *outreg* can be used to present the results in a nice looking table.

***Do-file or command window**

```
xtset country_id year
xtreg  secondary immunizations it_bed_net,fe
outreg using reg_module5, replace  se title("Panel regressions")
```

This is what your file will look like:

Panel regressions	
	secondary
immunizations	0.235
	(0.182)
it_bed_net	0.242
	(0.095)*
Constant	44.083
	(12.757)**
Observations	76
Number of Country Code	38
R-squared	0.41
Standard errors in parentheses	
* significant at 5%; ** significant at 1%	

6. Wrapping Up

In this module we have introduced panel data models. In particular, we have covered fixed effects, random effects, and difference-in-differences. We also showed how to export the results using *outreg*.

7. Appendix:

The fictitious treatment variable was generated like this:

***Do-file**

```
use countries_panel.dta, clear

isid country_id year

*1. Diff in diff (Post when year >= 1986)
gen treat = .
replace treat = 1 if country_id < 124
replace treat = 0 if country_id >= 124

gen post = .
replace post = 0 if year < 1986
replace post = 1 if year >= 1986

save countries_panel_treatment.dta, replace

* To generate the dataset for the difference in differences estimation,
* collapse the data:

collapse (mean) secondary immunizations it_bed_net (first) treat, by(country_id post)
save country_panel_collapsed.dta, replace
```