Prof. Javier Gómez Biscarri
MSc in Finance/EPP/CMR
ECONOMETRICS

Assignment 4: panel data

Submission deadline: this assignment need not be submitted. However, if you do it, you can submit it to Daniel (daniel.dejuan@barcelonagse.eu) and he will give you some feedback.

1) **Panel Data Example.**
The Stata file **panel_fin.dta** contains financial data for a large number of US banks listed in a stock exchange over several quarters (1987Q2-2010Q3). The variables in the dataset are:
accyearquarter – Quarter
Quarter – Quarter identifier (it goes from 1, 1987Q2, to 94, 2010Q3)
Entity – Bank identifier
ret – return of the bank's stock over that quarter
l_ret – return of the bank's stock over the previous quarter
l_btm – book-to-market ratio at the beginning of the quarter
l_e_p – earnings/price ratio at the beginning of the quarter
l_roe – return on equity (earnings/equity) in the previous quarter
(the "*l*" in the above variables means "lagged").

We are going to examine the question of return predictability: Do variables such as the past return, book-to-market or some measure of past earnings have any predictive power over the bank's stock return? The data in the file are quite rich, since they contain all observations of listed banks over quite a long time period.

We are going to estimate the following simple predictive equation:

$$ret_{it} = \beta_0 + \beta_1 l\_ret_{it} + \beta_2 l\_btm_{it} + \beta_3 X_{it} + e_{it}$$

a. Estimate the equation by pooled OLS using *l_roe* and *l_e_p* as the third explanatory variable (X_{it}) (so estimate the equation twice). Is there evidence of return predictability? What variables seem to have information about the bank's subsequent return?
   b. Estimate now the model (from now on, choose *l_roe* as X_{it}) again as pooled OLS but where you:
      1. Include bank specific intercepts (bank dummies) and test for the joint significance of the effects. Remember that these estimates are inconsistent, but we do not want to worry much about that. Also, you may want to be careful here, since there are lots of banks, so if you use Stata the program might ask you to increase the memory allocated to the analysis.
      2. Include time effects but not bank effects (i.e. include a dummy for each period of time). Test for the joint significance of the time effects.
c. Estimate now the model as a random-effects panel and a fixed-effects panel, with no time effects. Check that the point estimates of β_1, β_2 and β_3 in fixed

effects are the same as those in b-1. Do the conclusions on predictability change now?

d. Given the analysis above, offer a "story" for your findings. That is, if you did find some predictability, are there any interesting theoretical arguments for such predictability?

Stata commands

A TRICK TO GENERATE DUMMIES

If you want to generate a set of dummies for each value of a certain variable xk (*i* in exercise 3b-1 or *year* in exercise 3b-2) you can do it by using the command **xi:** and the identifier **i.xk**. For example, the following

xi: regress y x1 x2 x3 i.x4

will do an OLS regression of y on x1, x2, x3 and a set of dummies for each value of x4 (dropping one, of course, in order to avoid perfect colinearity).

If you use this command, a list of variables with names *_Ix4_#* will appear. If you want to test that the coefficients attached to those variables are zero you can use

test _Ix4_1 _Ix4_2 ….

But if you have many of them, this is very cumbersome and test does not admit the typical name shortcut. However, **testparm** does admit the shortcut, so you can write

testparm _Ix4*

and Stata will test for joint significance of all the dummies generated for the different values of x4.

ESTIMATION OF PANEL DATA MODELS

In order to estimate panels with Stata, first you have to tell the program that indeed you are using a panel, and the variables that identify the individuals (*i*) and the second dimension (usually time periods, *t*). You do this using the command:

xtset i t

Then you can easily perform fixed-effects regressions:

xtreg *depvar indep_vars*, fe

or random effects regressions:

xtreg *depvar indep_vars*, re

(note that stata calls the individual heterogeneity u_i, whereas we called it α_i).

or group-means (between) regressions:

xtreg *depvar indep_vars*, be

Useful tests that can be performed are the Breusch-Pagan test for random effects (versus no random effects: the test is really a test of $V(\alpha_i)=0$). After estimating random effects, the command for the test is:

xttest0

Also, the Hausman test for random effects (the hypothesis is $C(\alpha_i,e_{it})=0$, so under this null RE is efficient, but under the alternative it is inconsistent, whereas FE is always consistent) can be done in three steps:

xtreg depvar *indepvars*, fe
estimates store fixed_eff

xtreg depvar *indepvars*, re
estimates store random_eff

hausman fixed_eff random_eff

R commands

A TRICK TO GENERATE DUMMIES

If you want to generate a set of dummies for each value of a certain variable xk (*i* in exercise 3b-1 or *year* in exercise 3b-2) you can do it by calling a factor into your model:

lm(y~x1+factor(x2)+..., data=df)

If you use this command, a list of variables with names starting by "factor" will appear. If you want to test that the coefficients attached to those variables are zero you can use

linearHypothesis (model, matchCoefs(model, "factor"))

and R will test for joint significance of all the dummies generated for the different values of the factor x2.

ESTIMATION OF PANEL DATA MODELS

In order to estimate panels with R, first you have to tell the program that indeed you are using a panel, and the variables that identify the individuals (*i*) and the second dimension (usually time periods, *t*). You do this using the following:
library(plm)
paneldf<-plm.data(dataframe, indexes=c("individualvar", "timevar"))
(alternatively,

paneldf<-pdata.frame(dataframe, index=c(“individualvar”, “timevar”))

although the `index = c("individual","time")` can also be included in the plm command (below).

Then you can easily perform fixed-effects regressions:
plm(y ~x1+x2+..., data=paneldf, model = c("within"))

or random effects regressions:
plm(y ~x1+x2+..., data=paneldf, model = c("random"))

or group-means (between) regressions:
plm(y ~x1+x2+..., data=paneldf, model = c("between"))

Useful tests that can be performed are the Breusch-Pagan test for random effects (versus no random effects: the test is really a test of $V(\alpha_i)=0$). After estimating random effects, the command for the test is:

plmtest(panelmodel, effect = c("individual"), type = c("bp"))

Also, the Hausman test for random effects (the hypothesis is $C(\alpha_i,e_{it})=0$, so under this null RE is efficient, but under the alternative it is inconsistent, whereas FE is always consistent) can be done by:

phtest(femodel, remodel)