Applied Econometrics: Computer Assignment-1

Submitted by: Apoorva Jain

PANEL DATA REGRESSION

Introduction

Panel data analysis is an increasingly popular form of longitudinal data analysis among social and behavioral science researchers. A panel is a cross-section or group of individuals/countries/cities who are surveyed periodically over a given time span.

The Panel Analysis Equation

$$Y_{it} = a_i + \beta_{1*} X_{1it} + \beta_{2*} X_{2it} + e_{it}$$

The reference paper for this study is mentioned as below: -

The Economic Research Guardian – Vol. 2(1)2012

WOMEN INDICATORS OF ECONOMIC GROWTH: A PANEL DATA APPROACH

Our sample consists of 39 cross- sectional units in the form of countries which we have studied for 3 years-2013,2015 and 2017. We have a balanced panel.

Country Name	Year	GDP per capita growth (annual %)	Fertility rate, total (births per woman)	Population growth (annual %)
Angola	1989	-3.3359	7.328	0.059987665
Angola	1993	-26.4118	7.065	0.333166396
Angola	1997	3.87785	6.789	0.415564969
Angola	2001	0.822114	6.601	0.358315679
Angola	2005	11.03084	6.461	0.275481733
Angola	2009	-2.80863	6.26	0.535079062
Angola	2013	1.292086	5.953	0.416901361
Angola	2017	-3.4099	5.6	0.643350904
Australia	1989	2.126443	1.838	2.393243584
Australia	1993	3.016627	1.859	2.263604202
Australia	1997	2.809859	1.778	2.099269021

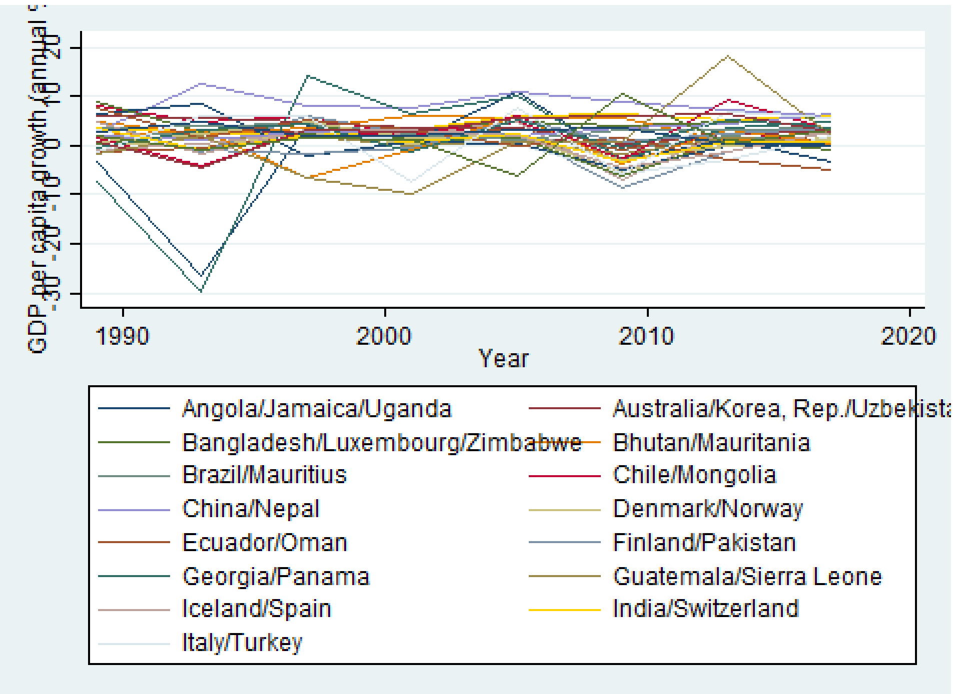
Australia	2001	0.565648	1.739	1.824913249
Australia	2005	1.851069	1.807	1.673564212
Australia	2009	-0.14256	1.971	1.628720225
Australia	2013	0.864663	1.855	1.496886223
Australia	2017	0.633037	1.765	1.768538648

Explanatory Variables:

- ***Fertility Rate***
- ***Population Growth***

Dependent Variable: GDP per capita growth

GRAPH 1



The above line graph highlights the variations in different cross-sectional units which in our case are the countries over the period of 8 years . The graph clearly shows that there are remarkable differences in the countries per capita GDP growth over these 8 year period. The variations may come due to many factors- Life expectancy rate,literacy rate,labor force participation etc.

GRAPH 2

GRAPH 2 represents the separate graphs for GDP per capita growth of each country over the 8 year period. It can be observed clearly that there is a huge ups and downs in the country's GDP per capita growth. The countries like Georgia, Turkey and Angola have experienced fluctuating variations in their gdp growth however Oman has experienced declining gdp growth over time. Some countries had a nearly constant growth over time such as INDIA, Korea,Uganda etc.

SUMMARY STATISTICS

TABLE 1.1

Variable		Mean	Std. Dev.	Min	Max	Observations
Countr~e	overall	.	.	.	.	N = 0
	Between		.	.	.	n = 0
	within		.	.	.	T = .
Year	overall	2003	9.182559	1989	2017	N = 264
	between		0	2003	2003	n = 33

	within		9.182559	1989	2017	T = 8
Y	overall	1.831194	4.566379	-29.84129	18.05314	N = 264
	between		1.774873	-2.367916	7.980037	n = 33
	within		4.217284	-27.92254	19.17115	T = 8
Fertil~e	overall	2.933792	1.588991	1.052	7.462	N = 264
	between		1.485706	1.295	6.507125	n = 33
	within		.6134492	1.350542	6.230042	T = 8
Popula~l	overall	1.453273	1.112182	-3.207518	7.34964	N = 264
	between		.9402871	-.6923246	3.546798	n = 33
	within		.6134663	-1.509181	5.256115	T = 8
id	overall	17	9.53999	1	33	N = 264
	between		9.66954	1	33	n = 33
	within		0	17	17	T = 8
yd1	overall	.125	.3313471	0	1	N = 264
	between		0	.125	.125	n = 33
	within		.3313471	0	1	T = 8
yd2	overall	.125	.3313471	0	1	N = 264
	between		0	.125	.125	n = 33
	within		.3313471	0	1	T = 8

yd3	overall	.125	.3313471	0	1	N = 264
	between		0	.125	.125	n = 33
	within		.3313471	0	1	T = 8
yd4	overall	.125	.3313471	0	1	N = 264
	between		0	.125	.125	n = 33
	within		.3313471	0	1	T = 8
yd5	overall	.125	.3313471	0	1	N = 264
	between		0	.125	.125	n = 33
	within		.3313471	0	1	T = 8
yd6	overall	.125	.3313471	0	1	N = 264
	between		0	.125	.125	n = 33
	within		.3313471	0	1	T = 8
yd7	overall	.125	.3313471	0	1	N = 264
	between		0	.125	.125	n = 33
	within		.3313471	0	1	T = 8
yd8	overall	.125	.3313471	0	1	N = 264
	between		0	.125	.125	n = 33
	within		.3313471	0	1	T = 8
_est_r~1	overall	1	0	1	1	N = 264
	between		0	1	1	n = 33

within | 0 1 1 | T = 8

_est_r~2 overall | 1 0 1 1 | N = 264

between | 0 1 1 | n = 33

within | 0 1 1 | T = 8

_est_r~3 overall | 1 0 1 1 | N = 264

between | 0 1 1 | n = 33

within | 0 1 1 | T = 8

_est_r~4 overall | 1 0 1 1 | N = 264

between | 0 1 1 | n = 33

within | 0 1 1 | T = 8

Table 1.1 provides the descriptive statistics for the per capita GDP growth rate. It shows the mean, standard deviation, minimum and maximum value of the variables. N is the number of total observations; n is the number of countries and T is the time period we have taken to study the variation of 39 cross-sections. It shows that our panel data is balanced with 38 observations for each time period.

In this document we focus on two techniques use to analyze panel data: – Fixed effects – Random effects.

Pooled & RE estimates for per capita gdp growth rates with and without dummies

VARIABLES	(Pooled)	(FE)	(RE)	(Pooled w/ Time Dummies)	(FE w/ Time Dummies)	(RE w/ Time Dummies)
Population growth annual	-0.597**	-1.473***	-0.675**	-0.593**	-1.509***	-0.697***

Std error	(0.251)	(0.443)	(0.270)	(0.245)	(0.437)	(0.269)
o.yd1				-		
yd2				-2.116*	-1.852*	-1.905*
				(1.081)	(1.042)	(1.051)
yd3				0.513	0.537	0.697
				(1.086)	(1.045)	(1.052)
yd4				-0.963	-0.860	-0.770
				(1.084)	(1.043)	(1.051)
yd5				1.269	1.427	1.468
				(1.083)	(1.042)	(1.051)
yd6				-2.931***	-2.770***	-2.732***
				(1.082)	(1.042)	(1.051)
yd7				0.496	0.742	0.705
				(1.081)	(1.042)	(1.051)
yd8				-0.204		
				(1.082)		
yd1					0.484	0.236
					(1.050)	(1.054)
o.yd8					-	-
Constant	2.699***	3.972***	2.813***	3.186***	4.311***	3.132***
	(0.459)	(0.698)	(0.503)	(0.876)	(0.972)	(0.857)
Observations	264	264	264	264	264	264
R-squared	0.021	0.046		0.107	0.147	
Number of id		33	33		33	33

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

FIXED EFFECTS MODEL.

Controlling for variables that are constant across entities but vary over time can be done by including time fixed effects.

ESTIMATED REGRESSION EQUATION WITH TIME DUMMIES -

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B_{2t} + \delta_3 B_{Tt} + u_{it}$$

ESTIMATED REGRESSION EQUATION WITHOUT TIME DUMMIES -

$$Y_{it} = \beta_0 + \beta_1 X_{it} + u_{it}$$

From the above regression results, we can see that population growth is highly significant . At 1% level of significance , population growth is significant in fixed effects model both in time-dummies case as well as when time dummies are not taken into account. That means population growth in our model explains huge variation in per capita gdp growth over 8 year time period.

RANDOM EFFECTS MODEL

ESTIMATED REGRESSION EQUATION WITHOUT TIME DUMMIES:-

$$Yit = \beta X_{it} + \alpha + u_{it} + \varepsilon_{it}$$

ESTIMATED REGRESSION EQUATION WITH TIME DUMMIES

$$Yit = \beta X_{it} + \alpha_t + u_{it} + \varepsilon_{it}$$

From the above results, we can see that population growth is significant at 5% level of significance when time dummies are not included however it is significant at 1% level of significance when time dummies are included.

POOLED REGRESSION MODEL

ESTIMATED REGRESSION EQUATION WITHOUT TIME DUMMIES:-

$$Y_{it}= a_i+ \beta_1 * X_{1it}+ \beta_2 * X_{2it}+e_{it}$$

ESTIMATED REGRESSION EQUATION WITH TIME DUMMIES:-

$$Y_{it}= a_i+ \beta_1 * X_{1it}+ \beta_2 * X_{2it}+ Zt+e_{it}$$

We can see that under pooled regression analysis, population growth is significant at 5% level of significance and explains huge variation in gdp per capita growth.

HYPOTHESIS TESTING UNDER POOLED REGRESSION

```
( 1)  1993.Year = 0
( 2)  1997.Year = 0
( 3)  2001.Year = 0
( 4)  2005.Year = 0
( 5)  2009.Year = 0
( 6)  2013.Year = 0
( 7)  2017.Year = 0

      F(  7,   255) =    3.49
           Prob > F =    0.0014
```

The above hypothesis testing results shows that the p-value is very small and hence the null hypothesis that all the time dummies are 0 can be rejected safely at 1% level of significance. Pooled regression with time dummies is a better model.

HYPOTHESIS TESTING UNDER FIXED EFFECTS MODEL

```
( 1)  1993.Year = 0
( 2)  1997.Year = 0
( 3)  2001.Year = 0
( 4)  2005.Year = 0
( 5)  2009.Year = 0
( 6)  2013.Year = 0
( 7)  2017.Year = 0

      F(  7,   223) =    3.76
           Prob > F =    0.0007
```

The above hypothesis testing results shows that the p-value is very small and hence the null hypothesis that all the time dummies are 0 can be rejected safely at 1% level of significance. Fixed effects model with time dummies is a better model.

HYPOTHESIS TESTING UNDER RANDOM EFFECTS MODEL

```
( 1)  1993.Year = 0
( 2)  1997.Year = 0
( 3)  2001.Year = 0
( 4)  2005.Year = 0
( 5)  2009.Year = 0
( 6)  2013.Year = 0
( 7)  2017.Year = 0

           chi2(  7) =   25.75
         Prob > chi2 =    0.0006
```

The above hypothesis testing results shows that the p-value is very small and hence the null hypothesis that all the time dummies are 0 can be rejected safely at 1% level of significance. Random effects model with time dummies is a better model.

HAUSMAN TEST TO CHOOSE BETWEEN RE AND FE

WITHOUT TIME DUMMIES

	Coefficients (b) res1	(B) res2	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
Population~1	-1.472751	-.6753784	-.7973728	.35078

```
                 b = consistent under Ho and Ha; obtained from xtreg
      B = inconsistent under Ha, efficient under Ho; obtained from xtreg

  Test:  Ho:  difference in coefficients not systematic

                chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                        =        5.17
              Prob>chi2 =      0.0230
```

To decide between fixed or random effects we can run a Hausman test where the null hypothesis is that the preferred model is random effects vs. the alternative the fixed effects.

Since p value is less than 0.05, we will reject the null hypothesis and would use the fixed effects model for the estimation.

WITH TIME DUMMIES

	Coefficients (b) res3	(B) res4	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
Population~l	-1.508898	-.6969653	-.8119325	.3439579
yd1	.4835328	.2358255	.2477073	.
yd2	-1.851767	-1.904532	.0527648	.
yd3	.5366654	.6967627	-.1600973	.
yd4	-.8600486	-.7704896	-.089559	.
yd5	1.426618	1.468091	-.0414735	.
yd6	-2.769569	-2.731898	-.0376704	.
yd7	.7420421	.7052868	.0367553	.

```
                      b = consistent under Ho and Ha; obtained from xtreg
       B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(8) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        5.57
                Prob>chi2 =      0.6950
                (V_b-V_B is not positive definite)
```

.

We can observe clearly that p value here is greater than 0.05. Hence, we fail to reject the null hypothesis and instead use random effects model for the estimation when there are time dummies.

TO COMPARE BETWEEN POOLED MODEL AND FE MODEL

NO TIME DUMMY

```
Fixed-effects (within) regression               Number of obs      =       264
Group variable: id                              Number of groups   =        33

R-sq:                                           Obs per group:
     within  = 0.0459                                         min =          8
     between = 0.0128                                         avg =        8.0
     overall = 0.0211                                         max =          8

                                                F(1,230)           =     11.06
corr(u_i, Xb)  = -0.4649                        Prob > F           =    0.0010
```

Y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Populationgrowthannual	-1.472751	.4427679	-3.33	0.001	-2.345151	-.6003516
_cons	3.971504	.6982436	5.69	0.000	2.595733	5.347275
sigma_u	2.1238681					
sigma_e	4.4049876					
rho	.18862079	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(32, 230) = 1.46                    Prob > F = 0.0613
```

Here p value is greater than 0.05, hence we do not reject the null hypothesis and do pooled OLS estimation for the model.

TIME DUMMY

```
Fixed-effects (within) regression               Number of obs      =       264
Group variable: id                              Number of groups   =        33

R-sq:                                           Obs per group:
     within  = 0.1467                                         min =          8
     between = 0.0128                                         avg =        8.0
     overall = 0.0875                                         max =          8

                                                F(8,223)           =      4.79
corr(u_i, Xb)  = -0.3721                        Prob > F           =    0.0000
```

Y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Populationgrowthannual	-1.508898	.4366949	-3.46	0.001	-2.369474	-.6483211
yd1	.4835328	1.050016	0.46	0.646	-1.585691	2.552756
yd2	-1.851767	1.041916	-1.78	0.077	-3.905028	.2014942
yd3	.5366654	1.045083	0.51	0.608	-1.522837	2.596167
yd4	-.8600486	1.042643	-0.82	0.410	-2.914742	1.194645
yd5	1.426618	1.041768	1.37	0.172	-.6263525	3.479588
yd6	-2.769569	1.041727	-2.66	0.008	-4.822457	-.7166811
yd7	.7420421	1.041717	0.71	0.477	-1.310827	2.794911
yd8	0	(omitted)				
_cons	4.310601	.9718216	4.44	0.000	2.395472	6.22573
sigma_u	2.1429962					
sigma_e	4.2307128					
rho	.2041865	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(32, 223) = 1.58                    Prob > F = 0.0299
```

Since p value is very small, we reject the poolability and work with fixed effects model.

TO COMPARE BETWEEN POOLED MODEL AND RE MODEL

```
Breusch and Pagan Lagrangian multiplier test for random effects

        Y[id,t] = Xb + u[id] + e[id,t]

        Estimated results:
```

	Var	sd = sqrt(Var)
Y	20.85182	4.566379
e	17.89893	4.230713
u	.9726687	.9862397

```
        Test:   Var(u) = 0
                             chibar2(01) =     2.35
                          Prob > chibar2 =   0.0625
```

The LM test helps you decide between a random effects regression and a simple OLS regression. The null hypothesis in the LM test is that variances across entities is zero. That is, no significant difference across units (i.e. no panel effect) and that random effects is not appropriate. Here p value is greater than 0.05 ,since there is no evidence of significant differences across countries, we can use a simple OLS regression. We will not reject the null hypothesis and come to the conclusion that there is no such panel effect.

SUMMARY AND CONCLUSIONS

We have tried running the regressions by taking different combinations of explanatory variables to check the significance of the variables in all the 3 models- FE,RE and POOLED. We came to the conclusion that when both fertility rate and population growth rate was regressed in the models with time dummies/without time dummies, only population growth rate came out to be the significant variable. Fertility rate when taken individually and regressed is an insignificant variable in all the 3 models.

Using hausman test, we conclude that in the case of time dummies, random effect model is used while in case of models without time dummies, fixed effects will be used.

Time dummies are significant in our model ,so the random effects model should be used over fixed effects model. However, if the choice is to be made between pooled and random effects model, pooled model will be used for the estimation as the p –value is insignificant highlighting the fact that there is no such significant variations between the countries. If the choice is to be made amongst pooled OLS and FE models, we will choose FE model in case of time dummy model and pooled OLS in case of without dummy model.