PRINCETON UNIVERSITY

Getting Started in Fixed/Random Effects Models using R

(ver. 0.1-*Draft*)

Oscar Torres-Reyna
Data Consultant
otorres@princeton.edu

http://dss.princeton.edu/training/

Panel data (also known as longitudinal or cross-sectional time-series data) is a dataset in which the behavior of entities are observed across time.

These entities could be states, companies, individuals, countries, etc.

Panel data looks like this

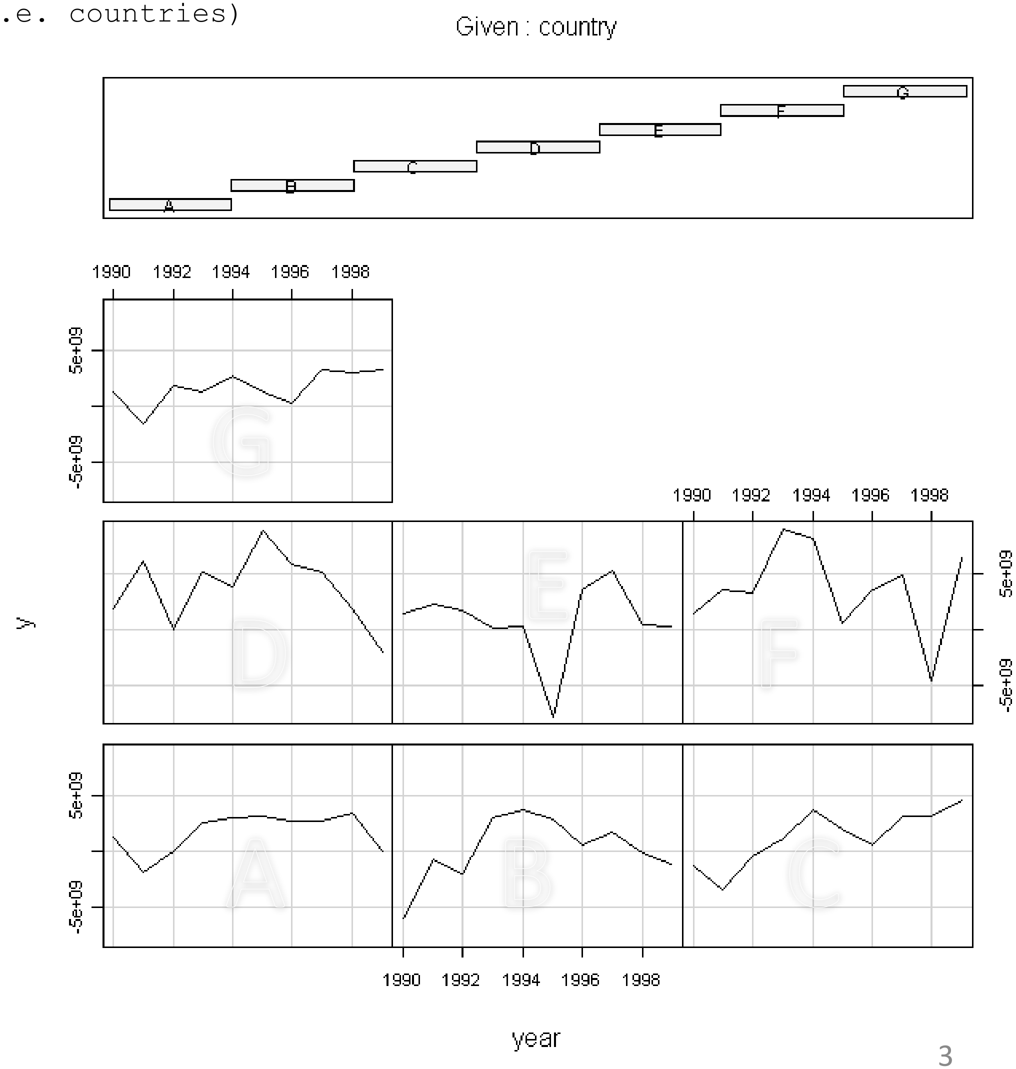
country	year	Y	X1	X2	X3
1	2000	6.0	7.8	5.8	1.3
1	2001	4.6	0.6	7.9	7.8
1	2002	9.4	2.1	5.4	1.1
2	2000	9.1	1.3	6.7	4.1
2	2001	8.3	0.9	6.6	5.0
2	2002	0.6	9.8	0.4	7.2
3	2000	9.1	0.2	2.6	6.4
3	2001	4.8	5.9	3.2	6.4
3	2002	9.1	5.2	6.9	2.1

For a brief introduction on the theory behind panel data analysis please see the following document:
http://dss.princeton.edu/training/Panel101.pdf

The contents of this document rely heavily on the document:
"Panel Data Econometrics in R: the plm package" http://cran.r-project.org/web/packages/plm/vignettes/plm.pdf and notes from the *ICPSR's Summer Program in Quantitative Methods of Social Research* (summer 2010)

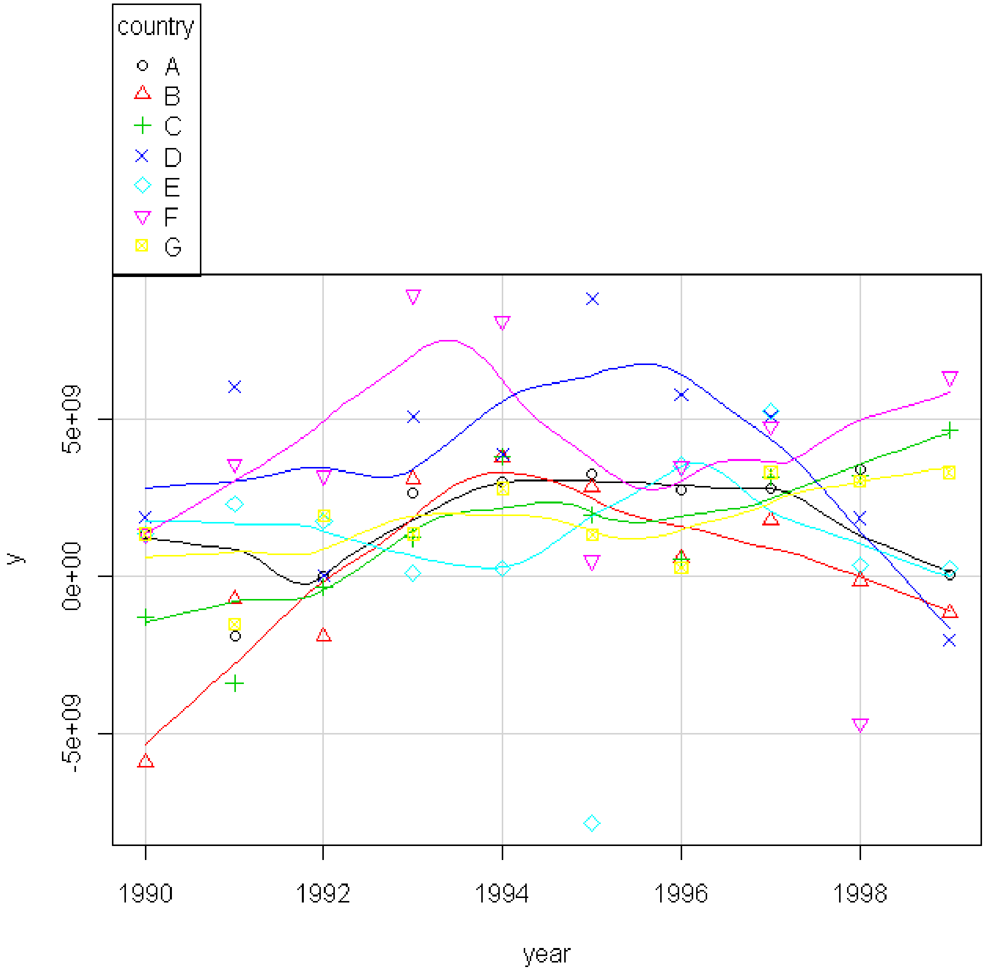
Exploring panel data

```
library(foreign)
Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")
coplot(y ~ year|country, type="l", data=Panel)        # Lines
coplot(y ~ year|country, type="b", data=Panel)        # Points and lines

# Bars at top indicates corresponding graph (i.e. countries)
from left to right starting on the bottom row
(Muenchen/Hilbe:355)
```



Exploring panel data

```
library(foreign)
Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")
library(car)
scatterplot(y~year|country, boxplots=FALSE, smooth=TRUE, reg.line=FALSE, data=Panel)
```



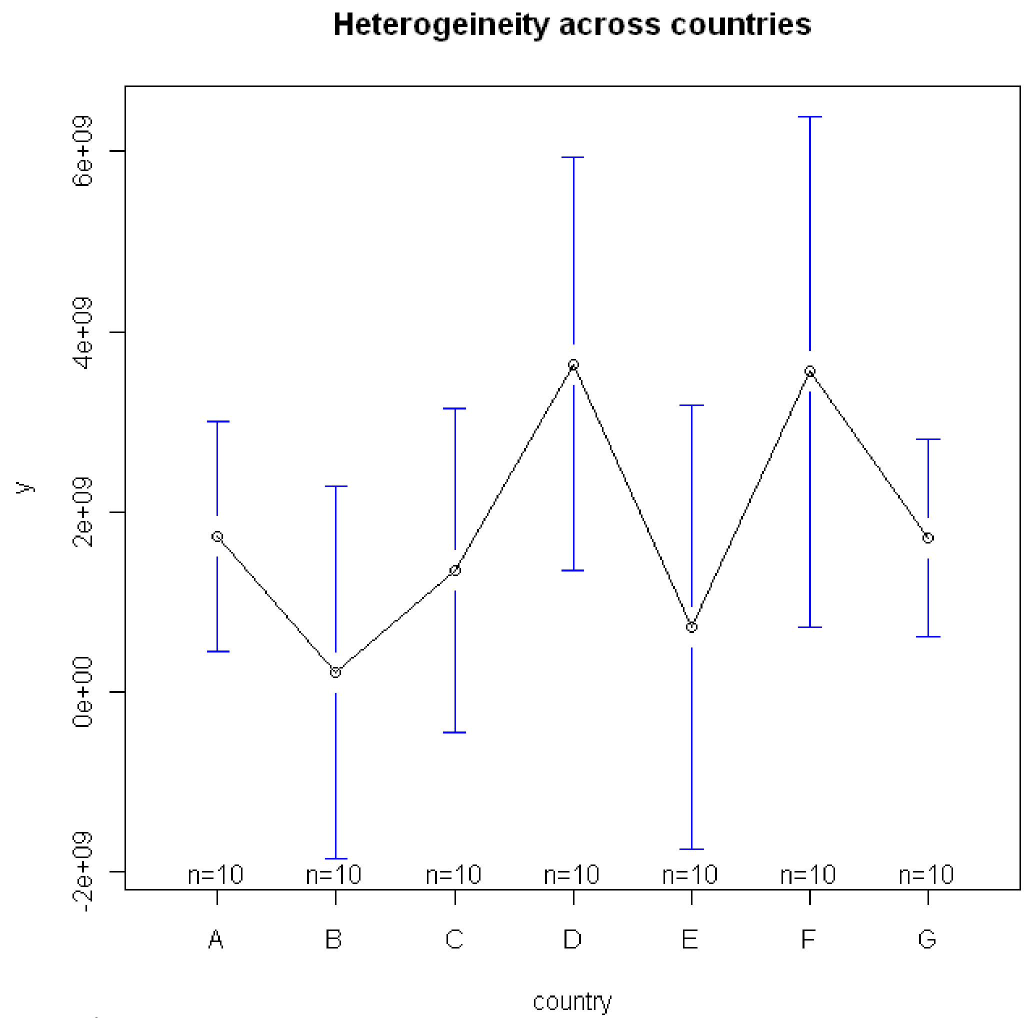
FIXED-EFFECTS MODEL

(Covariance Model, Within Estimator, Individual Dummy Variable Model, Least Squares Dummy Variable Model)

Fixed effects: Heterogeneity across countries (or entities)

```
library(foreign)
Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")
library(gplots)
plotmeans(y ~ country, main="Heterogeineity across countries", data=Panel)
```

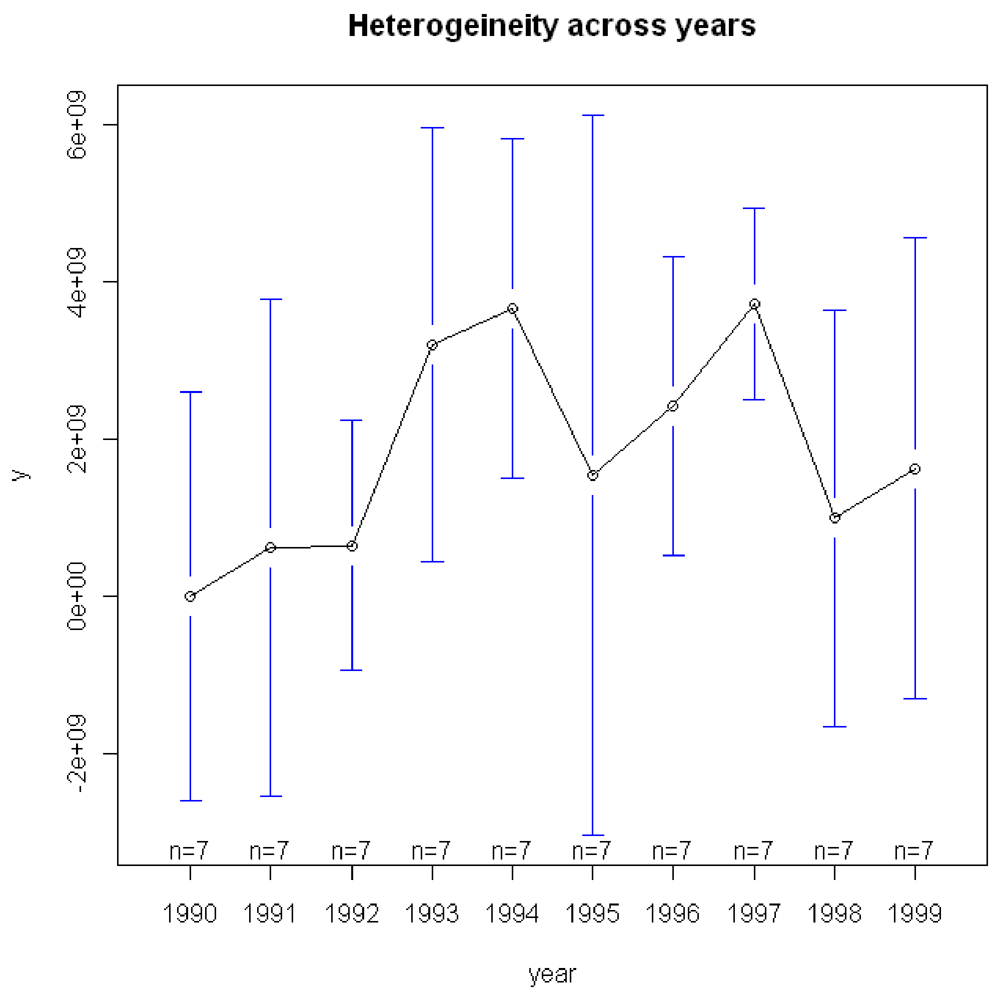
```
# plotmeans draw a 95%
confidence interval
around the means

detach("package:gplots")

# Remove package 'gplots'
from the workspace
```



Heterogeneity: unobserved variables that do not change over time

Fixed effects: Heterogeneity across years

```
library(foreign)
Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")
library(gplots)
plotmeans(y ~ year, main="Heterogeineity across years", data=Panel)

# plotmeans draw a 95%
confidence interval
around the means

detach("package:gplots")

# Remove package 'gplots'
from the workspace
```

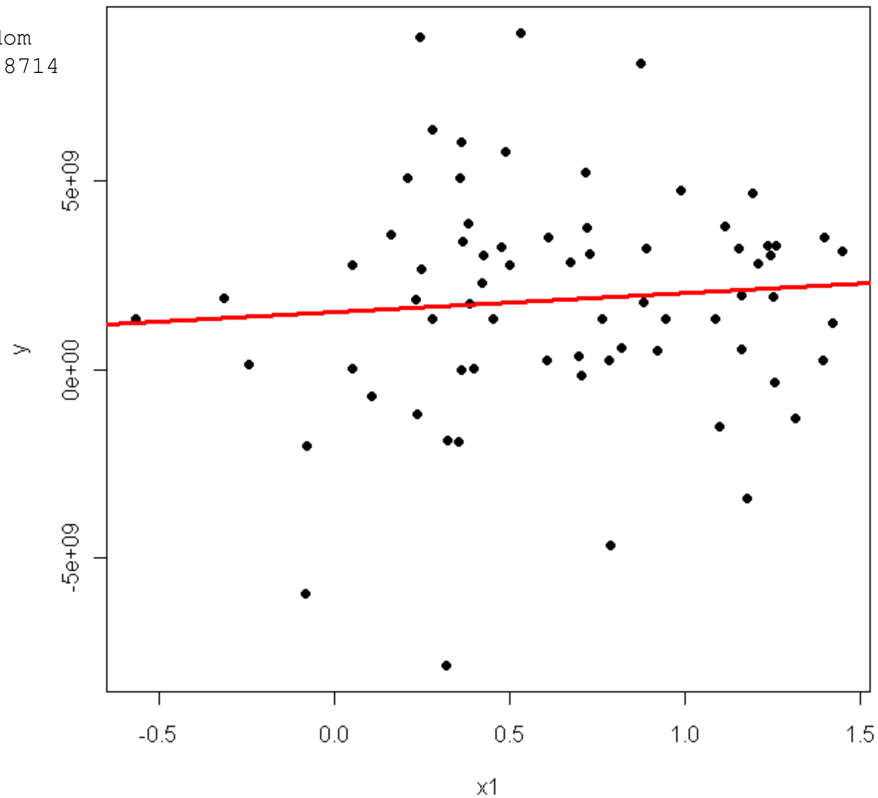


Heterogeneity: unobserved variables that do not change over time

OLS regression

```
> library(foreign)
> Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")
> ols <-lm(y ~ x1, data=Panel)
> summary(ols)

     Call:
     lm(formula = y ~ x1, data = Panel)

     Residuals:
            Min         1Q     Median         3Q        Max
     -9.546e+09 -1.578e+09  1.554e+08  1.422e+09  7.183e+09

     Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
     (Intercept) 1.524e+09  6.211e+08   2.454   0.0167 *
     x1          4.950e+08  7.789e+08   0.636   0.5272
     ---
     Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Residual standard error: 3.028e+09 on 68 degrees of freedom
     Multiple R-squared: 0.005905,   Adjusted R-squared: -0.008714
     F-statistic: 0.4039 on 1 and 68 DF,  p-value: 0.5272

> yhat <- ols$fitted

> plot(Panel$x1, Panel$y, pch=19, xlab="x1", ylab="y")
> abline(lm(Panel$y~Panel$x1),lwd=3, col="red")
```

Regular OLS regression does not consider heterogeneity across groups or time

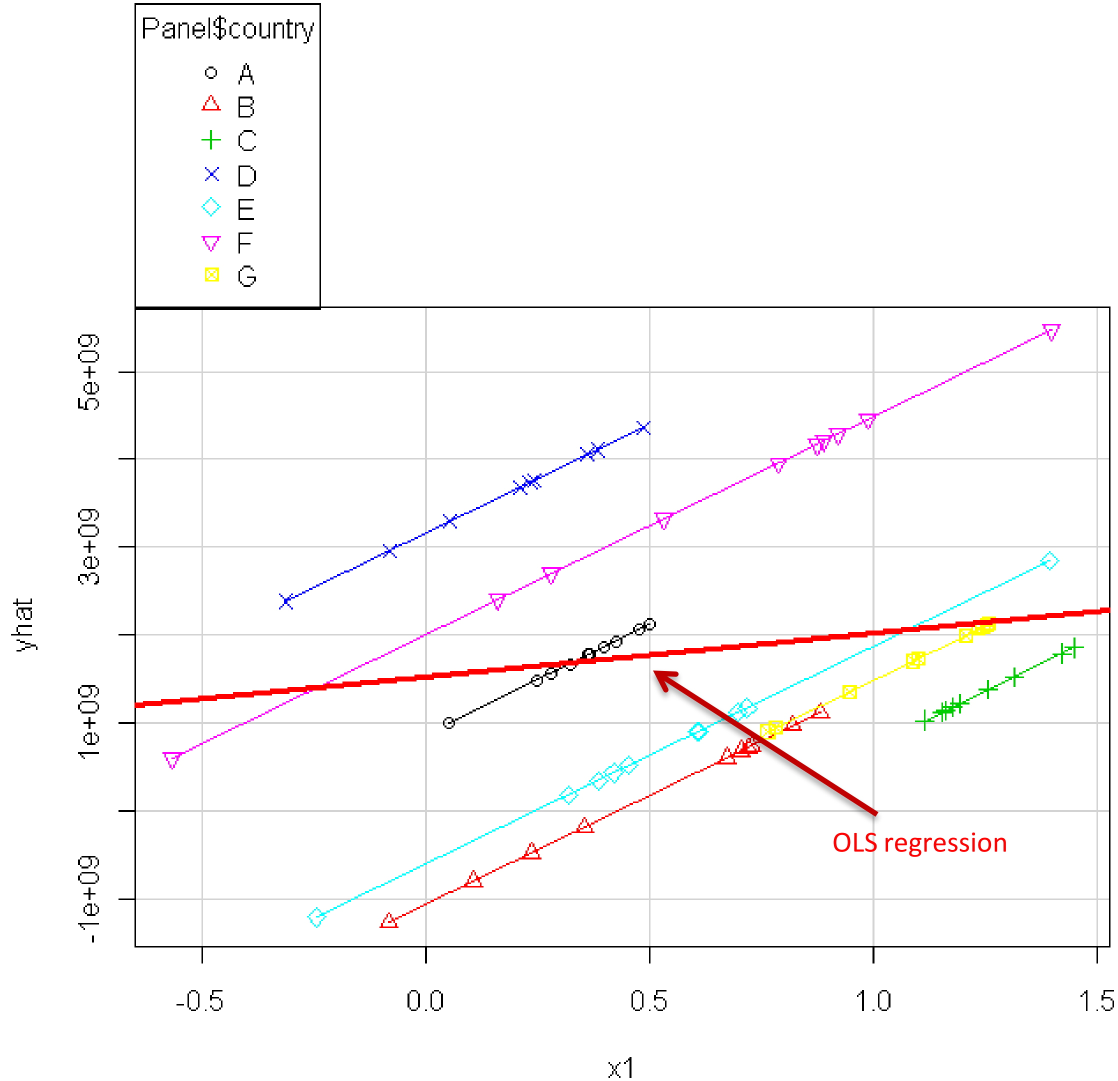


```
> library(foreign)
>Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")

> fixed.dum <-lm(y ~ x1 + factor(country) - 1, data=Panel)
> summary(fixed.dum)

      Call:
      lm(formula = y ~ x1 + factor(country) - 1, data = Panel)

      Residuals:
             Min          1Q      Median          3Q         Max
      -8.634e+09 -9.697e+08  5.405e+08  1.386e+09  5.612e+09

      Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
      x1                2.476e+09  1.107e+09   2.237  0.02889 *
      factor(country)A  8.805e+08  9.618e+08   0.916  0.36347
      factor(country)B -1.058e+09  1.051e+09  -1.006  0.31811
      factor(country)C -1.723e+09  1.632e+09  -1.056  0.29508
      factor(country)D  3.163e+09  9.095e+08   3.478  0.00093 ***
      factor(country)E -6.026e+08  1.064e+09  -0.566  0.57329
      factor(country)F  2.011e+09  1.123e+09   1.791  0.07821 .
      factor(country)G -9.847e+08  1.493e+09  -0.660  0.51190
      ---
      Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      Residual standard error: 2.796e+09 on 62 degrees of freedom
      Multiple R-squared: 0.4402,     Adjusted R-squared: 0.368
      F-statistic: 6.095 on 8 and 62 DF,  p-value: 8.892e-06
```

For the theory behind fixed effects, please see http://dss.princeton.edu/training/Panel101.pdf

Least squares dummy variable model

```
> yhat <- fixed.dum$fitted
> library(car)
> scatterplot(yhat~Panel$x1|Panel$country, boxplots=FALSE, xlab="x1", ylab="yhat",smooth=FALSE)
> abline(lm(Panel$y~Panel$x1),lwd=3, col="red")
```



Each component of the factor variable (country) is absorbing the effects particular to each country. Predictor *x1* was not significant in the OLS model, once controlling for differences across countries, *x1* became significant in the OLS_DUM (i.e. LSDV model).

```
> library(apsrtable)
> apsrtable(ols,fixed.dum, model.names = c("OLS", "OLS_DUM"))   # Displays a table in Latex form
```

```
\begin{table}[!ht]
\caption{}
\label{}
\begin{tabular}{ l D{.}{.}{2}D{.}{.}{2} }
\hline
  & \multicolumn{ 1 }{ c }{ OLS } & \multicolumn{ 1 }{ c }{ OLS_DUM } \\
\hline
%                 & OLS                & OLS_DUM          \\
(Intercept)       & 1524319070.05 ^*  &                  \\
                  & (621072623.86)     &                  \\
x1                & 494988913.90       & 2475617827.10 ^*\\
                  & (778861260.95)     & (1106675593.60)  \\
factor(country)A  &                    & 880542403.99     \\
                  &                    & (961807052.24)   \\
factor(country)B  &                    & -1057858363.16   \\
                  &                    & (1051067684.19)  \\
factor(country)C  &                    & -1722810754.55   \\
                  &                    & (1631513751.40)  \\
factor(country)D  &                    & 3162826897.32 ^*\\
                  &                    & (909459149.66)   \\
factor(country)E  &                    & -602622000.33    \\
                  &                    & (1064291684.41)  \\
factor(country)F  &                    & 2010731793.24    \\
                  &                    & (1122809097.35)  \\
factor(country)G  &                    & -984717493.45    \\
                  &                    & (1492723118.24)   \\
 $N$               & 70                 & 70                \\
$R^2$             & 0.01               & 0.44             \\
adj. $R^2$        & -0.01              & 0.37             \\
Resid. sd         & 3028276248.26      & 2795552570.60     \\ \hline
 \multicolumn{3}{l}{\footnotesize{Standard errors in parentheses}}\\
\multicolumn{3}{l}{\footnotesize{$^*$ indicates significance at $p< 0.05
$}}
\end{tabular}
 \end{table}
```

The coefficient of *x1* indicates how much *Y* changes when *X* increases by one unit. Notice *x1* is not significant in the OLS model

The coefficient of *x1* indicates how much *Y* changes overtime, controlling by differences in countries, when *X* increases by one unit. Notice *x1* is significant in the LSDV model

```
> cat(apsrtable(ols, fixed.dum, model.names = c("OLS", "OLS_DUM"), Sweave=F), file="ols_fixed1.txt")

# Exports the table to a text file (in Latex code).
```

Fixed effects: *n* entity-specific intercepts (using `plm`)

Outcome variable

```
> library(plm)
> fixed <- plm(y ~ x1, data=Panel, index=c("country", "year"), model="within")
> summary(fixed)
```

Predictor variable(s)

Panel setting

Fixed effects option

```
Oneway (individual) effect Within Model

Call:
plm(formula = y ~ x1, data = Panel, model = "within", index = c("country",
    "year"))

Balanced Panel: n=7, T=10, N=70
```

n = # of groups/panels, T = # years, N = total # of observations

```
Residuals :
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-8.63e+09 -9.70e+08  5.40e+08  1.81e-10  1.39e+09  5.61e+09

Coefficients :
     Estimate Std. Error t-value Pr(>|t|)
x1 2475617827 1106675594   2.237  0.02889 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    5.2364e+20
Residual Sum of Squares: 4.8454e+20
R-Squared      :  0.074684
      Adj. R-Squared :  0.066148
F-statistic: 5.00411 on 1 and 62 DF, p-value: 0.028892
```

The coeff of x1 indicates how much *Y* changes overtime, on average per country, when *X* increases by one unit.

Pr(>|t|)= Two-tail p-values test the hypothesis that each coefficient is different from 0. To reject this, the p-value has to be lower than 0.05 (95%, you could choose also an alpha of 0.10), if this is the case then you can say that the variable has a significant influence on your dependent variable (y)

If this number is < 0.05 then your model is ok. This is a test (F) to see whether all the coefficients in the model are different than zero.

```
> fixef(fixed)     # Display the fixed effects (constants for each country)
          A           B           C          D          E          F
  880542404 -1057858363 -1722810755 3162826897 -602622000 2010731793
          G
 -984717493

> pFtest(fixed, ols)     # Testing for fixed effects, null: OLS better than fixed

        F test for individual effects

data:  y ~ x1
F = 2.9655, df1 = 6, df2 = 62, p-value = 0.01307
alternative hypothesis: significant effects
```

If the p-value is < 0.05 then the fixed effects model is a better choice

RANDOM-EFFECTS MODEL

(Random Intercept, Partial Pooling Model)

Random effects (using `plm`)

Outcome variable

Predictor variable(s)

Panel setting

Random effects option

```
> random <- plm(y ~ x1, data=Panel, index=c("country", "year"), model="random")
> summary(random)
          Oneway (individual) effect Random Effect Model
             (Swamy-Arora's transformation)

          Call:
          plm(formula = y ~ x1, data = Panel, model = "random", index = c("country",
              "year"))

          Balanced Panel: n=7, T=10, N=70

          Effects:
                              var   std.dev share
          idiosyncratic 7.815e+18 2.796e+09 0.873
          individual    1.133e+18 1.065e+09 0.127
          theta:  0.3611

          Residuals :
               Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
          -8.94e+09 -1.51e+09  2.82e+08  5.29e-08  1.56e+09  6.63e+09

          Coefficients :
                        Estimate Std. Error t-value Pr(>|t|)
          (Intercept) 1037014284  790626206  1.3116   0.1941
          x1          1247001782  902145601  1.3823   0.1714

          Total Sum of Squares:    5.6595e+20
          Residual Sum of Squares: 5.5048e+20
          R-Squared      :  0.02733
                Adj. R-Squared :  0.026549
          F-statistic: 1.91065 on 1 and 68 DF, p-value: 0.17141
```

n = # of groups/panels, T = # years, N = total # of observations

Interpretation of the coefficients is tricky since they include both the within-entity and between-entity effects. In the case of TSCS data represents the average effect of *X* over *Y* when *X* changes across time and between countries by one unit.

Pr(>|t|)= Two-tail p-values test the hypothesis that each coefficient is different from 0. To reject this, the p-value has to be lower than 0.05 (95%, you could choose also an alpha of 0.10), if this is the case then you can say that the variable has a significant influence on your dependent variable (y)

If this number is < 0.05 then your model is ok. This is a test (F) to see whether all the coefficients in the model are different than zero.

```
# Setting as panel data (an alternative way to run the above model

Panel.set <- plm.data(Panel, index = c("country", "year"))

# Random effects using panel setting (same output as above)

random.set <- plm(y ~ x1, data = Panel.set, model="random")
summary(random.set)
```


For the theory behind random effects please see: http://dss.princeton.edu/training/Panel101.pdf

FIXED OR RANDOM?

Fixed or Random: Hausman test

To decide between fixed or random effects you can run a Hausman test where the null hypothesis is that the preferred model is random effects vs. the alternative the fixed effects (see Green, 2008, chapter 9). It basically tests whether the unique errors (u_i) are correlated with the regressors, the null hypothesis is they are not.

Run a fixed effects model and save the estimates, then run a random model and save the estimates, then perform the test. If the p-value is significant (for example <0.05) then use fixed effects, if not use random effects.

```
> phtest(fixed, random)

        Hausman Test

data:  y ~ x1
chisq = 3.674, df = 1, p-value = 0.05527
alternative hypothesis: one model is inconsistent
```

If this number is < 0.05 then use fixed effects

For the theory behind fixed/random effects please see: http://dss.princeton.edu/training/Panel101.pdf

OTHER TESTS/ DIAGNOSTICS

Testing for time-fixed effects

```
> library(plm)
> fixed <- plm(y ~ x1, data=Panel, index=c("country", "year"),
model="within")
> fixed.time <- plm(y ~ x1 + factor(year), data=Panel, index=c("country",
"year"), model="within")
> summary(fixed.time)
Oneway (individual) effect Within Model

Call:
plm(formula = y ~ x1 + factor(year), data = Panel, model = "within",
    index = c("country", "year"))

Balanced Panel: n=7, T=10, N=70

Residuals :
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-7.92e+09 -1.05e+09 -1.40e+08  1.48e-07  1.63e+09  5.49e+09

Coefficients :
                   Estimate Std. Error t-value Pr(>|t|)
x1               1389050354 1319849567  1.0524  0.29738
factor(year)1991  296381559 1503368528  0.1971  0.84447
factor(year)1992  145369666 1547226548  0.0940  0.92550
factor(year)1993 2874386795 1503862554  1.9113  0.06138 .
factor(year)1994 2848156288 1661498927  1.7142  0.09233 .
factor(year)1995  973941306 1567245748  0.6214  0.53698
factor(year)1996 1672812557 1631539254  1.0253  0.30988
factor(year)1997 2991770063 1627062032  1.8388  0.07156 .
factor(year)1998  367463593 1587924445  0.2314  0.81789
factor(year)1999 1258751933 1512397632  0.8323  0.40898
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    5.2364e+20
Residual Sum of Squares: 4.0201e+20
R-Squared      :  0.23229
      Adj. R-Squared :  0.17588
F-statistic: 1.60365 on 10 and 53 DF, p-value: 0.13113
```

```
> # Testing time-fixed effects. The null is that no time-fixed
effects needed

> pFtest(fixed.time, fixed)

         F test for individual effects

data:  y ~ x1 + factor(year)
F = 1.209, df1 = 9, df2 = 53, p-value = 0.3094
alternative hypothesis: significant effects

> plmtest(fixed, c("time"), type=("bp"))

         Lagrange Multiplier Test - time effects (Breusch-Pagan)

data:  y ~ x1
chisq = 0.1653, df = 1, p-value = 0.6843
alternative hypothesis: significant effects
```

If this number is < 0.05 then use time-fixed effects. In this example, no need to use time-fixed effects.

Testing for random effects: Breusch-Pagan Lagrange multiplier (LM)

```
> # Regular OLS (pooling model) using plm
>
> pool <- plm(y ~ x1, data=Panel, index=c("country", "year"), model="pooling")
> summary(pool)
      Oneway (individual) effect Pooling Model

      Call:
      plm(formula = y ~ x1, data = Panel, model = "pooling", index = c("country",
          "year"))

      Balanced Panel: n=7, T=10, N=70

      Residuals :
           Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
      -9.55e+09 -1.58e+09  1.55e+08  1.77e-08  1.42e+09  7.18e+09

      Coefficients :
                    Estimate Std. Error t-value Pr(>|t|)
      (Intercept) 1524319070  621072624  2.4543  0.01668 *
      x1           494988914  778861261  0.6355  0.52722
      ---
      Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      Total Sum of Squares:    6.2729e+20
      Residual Sum of Squares: 6.2359e+20
      R-Squared      :  0.0059046
            Adj. R-Squared :  0.0057359
      F-statistic: 0.403897 on 1 and 68 DF, p-value: 0.52722

> # Breusch-Pagan Lagrange Multiplier for random effects. Null is no panel effect (i.e. OLS better).

> plmtest(pool, type=c("bp"))

                 Lagrange Multiplier Test - (Breusch-Pagan)

      data:  y ~ x1
      chisq = 2.6692, df = 1, p-value = 0.1023
      alternative hypothesis: significant effects
```

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. This is, no significant difference across units (i.e. no panel effect). (http://dss.princeton.edu/training/Panel101.pdf)

Here we failed to reject the null and conclude that random effects is not appropriate. This is, no evidence of significant differences across countries, therefore you can run a simple OLS regression.

Testing for cross-sectional dependence/contemporaneous correlation: using Breusch-Pagan LM test of independence and Pasaran CD test

According to Baltagi, cross-sectional dependence is a problem in macro panels with long time series. This is not much of a problem in micro panels (few years and large number of cases).

The null hypothesis in the B-P/LM and Pasaran CD tests of independence is that residuals across entities are not correlated. B-P/LM and Pasaran CD (cross-sectional dependence) tests are used to test whether the residuals are correlated across entities*. Cross-sectional dependence can lead to bias in tests results (also called contemporaneous correlation).

```
> fixed <- plm(y ~ x1, data=Panel, index=c("country", "year"), model="within")

> pcdtest(fixed, test = c("lm"))

        Breusch-Pagan LM test for cross-sectional dependence in panels

data:  formula
chisq = 28.9143, df = 21, p-value = 0.1161
alternative hypothesis: cross-sectional dependence

> pcdtest(fixed, test = c("cd"))

        Pesaran CD test for cross-sectional dependence in panels

data:  formula
z = 1.1554, p-value = 0.2479
alternative hypothesis: cross-sectional dependence
```

No cross-sectional dependence

*Source: Hoechle, Daniel, "Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence", http://fmwww.bc.edu/repec/bocode/x/xtscc_paper.pdf

Testing for serial correlation

Serial correlation tests apply to macro panels with long time series. Not a problem in micro panels (with very few years). The null is that there is not serial correlation.

```
> pbgtest(fixed)
Loading required package: lmtest

        Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  y ~ x1
chisq = 14.1367, df = 10, p-value = 0.1668
alternative hypothesis: serial correlation in idiosyncratic errors
```

No serial correlation

The Dickey-Fuller test to check for stochastic trends. The null hypothesis is that the series has a unit root (i.e. non-stationary). If unit root is present you can take the first difference of the variable.

```
> Panel.set <- plm.data(Panel, index = c("country", "year"))

> library(tseries)

> adf.test(Panel.set$y, k=2)

        Augmented Dickey-Fuller Test

data:  Panel.set$y
Dickey-Fuller = -3.9051, Lag order = 2, p-value = 0.01910
alternative hypothesis: stationary
```

If p-value < 0.05 then no unit roots present.

The null hypothesis for the Breusch-Pagan test is homoskedasticity.

```
> library(lmtest)
> bptest(y ~ x1 + factor(country), data = Panel, studentize=F)

        Breusch-Pagan test

data:  y ~ x1 + factor(country)
BP = 14.6064, df = 7, p-value = 0.04139
```

Presence of heteroskedasticity

If hetersokedaticity is detected you can use robust covariance matrix to account for it. See the following pages.

Controlling for heteroskedasticity: Robust covariance matrix estimation (Sandwich estimator)

The --vcovHC– function estimates three heteroskedasticity-consistent covariance estimators:

- "white1" - for general heteroskedasticity but no serial correlation. Recommended for random effects.
- "white2" - is "white1" restricted to a common variance within groups. Recommended for random effects.
- "arellano" - both heteroskedasticity and serial correlation. Recommended for fixed effects.

The following options apply*:

- HC0 - heteroskedasticity consistent. The default.
- HC1,HC2, HC3 – Recommended for small samples. HC3 gives less weight to influential observations.
- HC4 - small samples with influential observations
- HAC - heteroskedasticity and autocorrelation consistent (type ?vcovHAC for more details)

See the following pages for examples

For more details see:

- http://cran.r-project.org/web/packages/plm/vignettes/plm.pdf
- http://cran.r-project.org/web/packages/sandwich/vignettes/sandwich.pdf (see page 4)
- Stock and Watson 2006.
- *Kleiber and Zeileis, 2008.

Controlling for heteroskedasticity: Random effects

```
> random <- plm(y ~ x1, data=Panel, index=c("country", "year"), model="random")

> coeftest(random)       # Original coefficients

t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1037014284  790626206  1.3116   0.1941
x1          1247001782  902145601  1.3823   0.1714

> coeftest(random, vcovHC)       # Heteroskedasticity consistent coefficients

t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1037014284  907983029  1.1421   0.2574
x1          1247001782  828970247  1.5043   0.1371

> coeftest(random, vcovHC(random, type = "HC3")) # Heteroskedasticity consistent coefficients, type 3

t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1037014284  943438284  1.0992   0.2756
x1          1247001782  867137585  1.4381   0.1550

> # The following shows the HC standard errors of the coefficients

> t(sapply(c("HC0", "HC1", "HC2", "HC3", "HC4"), function(x) sqrt(diag(vcovHC(random, type = x)))))

    (Intercept)        x1
HC0   907983029 828970247
HC1   921238957 841072643
HC2   925403820 847733474
HC3   943438284 867137584
HC4   941376033 866024033
```

Standard errors given different types of HC.

Controlling for heteroskedasticity: Fixed effects

```
> fixed <- plm(y ~ x1, data=Panel, index=c("country", "year"), model="within")

> coeftest(fixed)      # Original coefficients

t test of coefficients:

     Estimate Std. Error t value Pr(>|t|)
x1 2475617827 1106675594   2.237  0.02889 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coeftest(fixed, vcovHC) # Heteroskedasticity consistent coefficients

t test of coefficients:

     Estimate Std. Error t value Pr(>|t|)
x1 2475617827 1358388942  1.8225  0.07321 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coeftest(fixed, vcovHC(fixed, method = "arellano")) # Heteroskedasticity consistent coefficients (Arellano)

t test of coefficients:

     Estimate Std. Error t value Pr(>|t|)
x1 2475617827 1358388942  1.8225  0.07321 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coeftest(fixed, vcovHC(fixed, type = "HC3")) # Heteroskedasticity consistent coefficients, type 3

t test of coefficients:

     Estimate Std. Error t value Pr(>|t|)
x1 2475617827 1439083523  1.7203  0.09037 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> # The following shows the HC standard errors of the coefficients
> t(sapply(c("HC0", "HC1", "HC2", "HC3", "HC4"), function(x) sqrt(diag(vcovHC(fixed, type = x)))))
         HC0.x1     HC1.x1     HC2.x1     HC3.x1     HC4.x1
[1,] 1358388942 1368196931 1397037369 1439083523 1522166034
```

Standard errors given different types of HC.

References/Useful links

- DSS Online Training Section http://dss.princeton.edu/training/
- Princeton DSS Libguides http://libguides.princeton.edu/dss
- John Fox's site http://socserv.mcmaster.ca/jfox/
- Quick-R http://www.statmethods.net/
- UCLA Resources to learn and use R http://www.ats.ucla.edu/stat/R/
- UCLA Resources to learn and use Stata http://www.ats.ucla.edu/stat/stata/
- DSS - Stata http://dss/online_help/stats_packages/stata/
- DSS - R http://dss.princeton.edu/online_help/stats_packages/r
- Panel Data Econometrics in R: the plm package http://cran.r-project.org/web/packages/plm/vignettes/plm.pdf
- Econometric Computing with HC and HAC Covariance Matrix Estimators http://cran.r-project.org/web/packages/sandwich/vignettes/sandwich.pdf

References/Recommended books

- *An R Companion to Applied Regression*, Second Edition / John Fox , Sanford Weisberg, Sage Publications, 2011
- *Data Manipulation with R* / Phil Spector, Springer, 2008
- *Applied Econometrics with R* / Christian Kleiber, Achim Zeileis, Springer, 2008
- *Introductory Statistics with R* / Peter Dalgaard, Springer, 2008
- *Complex Surveys. A guide to Analysis Using R* / Thomas Lumley, Wiley, 2010
- *Applied Regression Analysis and Generalized Linear Models* / John Fox, Sage, 2008
- *R for Stata Users* / Robert A. Muenchen, Joseph Hilbe, Springer, 2010
- *Introduction to econometrics* / James H. Stock, Mark W. Watson. 2nd ed., Boston: Pearson Addison Wesley, 2007.
- *Data analysis using regression and multilevel/hierarchical models* / Andrew Gelman, Jennifer Hill. Cambridge ; New York : Cambridge University Press, 2007.
- *Econometric analysis* / William H. Greene. 6th ed., Upper Saddle River, N.J. : Prentice Hall, 2008.
- *Designing Social Inquiry: Scientific Inference in Qualitative Research / Gary King, Robert O. Keohane, Sidney Verba, Princeton University Press*, 1994.
- *Unifying Political Methodology: The Likelihood Theory of Statistical Inference* / Gary King, Cambridge University Press, 1989
- *Statistical Analysis: an interdisciplinary introduction to univariate & multivariate methods /* Sam Kachigan, New York : Radius Press, c1986
- *Statistics with Stata (updated for version 9) /* Lawrence Hamilton, Thomson Books/Cole, 2006