Random Effects and Hausman

Panel Data Econometrics

Prof. Alexandre Gori Maia

State University of Campinas

Topics

Random Effects

Hausman Test

Reference

Baltagi, B. Econometric analysis of panel data. Third Edition. John Wiley & Sons. 2005, Chapters 1-4.

Wooldridge, J. M. 2001. Econometric analysis of cross section and panel data. Cap. 10.

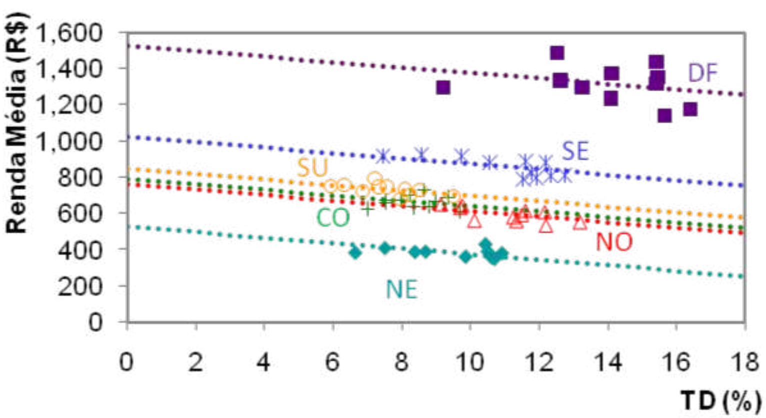
Fixed Effects Estimator

- The fixed effects within estimator is given by:

$$Y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + e_{it} \Rightarrow Y_{it} = \alpha + \sum_{j=1}^{k} \beta_j X_{j_{it}} + c_2 I_{2_i} + \ldots + c_n I_{n_i} + e_{it}$$

or

$$(Y_{it} - \overline{Y}_i) = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\boldsymbol{\beta} + (c_i - c_i) + (e_{it} - \overline{e}_i)$$

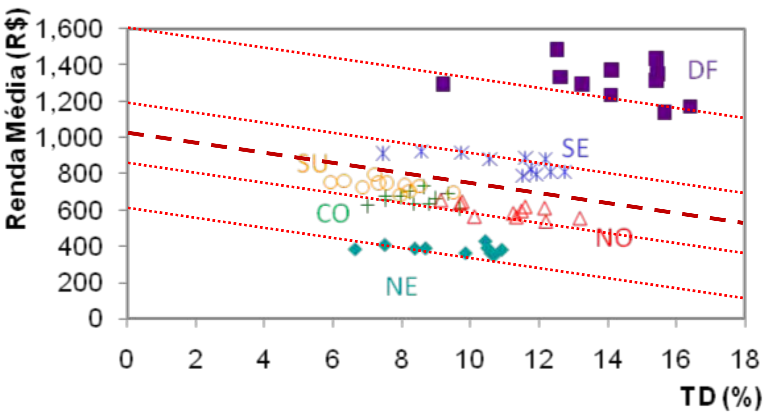


Both the fixed effects estimator with binary variables and the within estimator control unobservable characteristics that are constant over time.

- **Limitation**: the fixed effects estimator is operationally easy to implement, but may consume a large number of degrees of freedom (equivalent to the number of cross-sectional units minus one). This means, this estimator tend to be inefficient;

Random Effects Estimator

- The random effects estimator assumes that the intercept varies randomly around a constant value (α);
- This random variation ($c_{i.}$) can also be assumed to be part of the error:



Suppose we want to adjust the model:

$$Renda_{it} = \alpha + \beta TD_{it} + e_{it}$$

Assuming that the intercept is defined by the sum of a constant α and a random variable c_i:

$$Renda_{it} = \alpha + c_i + \beta TD_{it} + e_{it}$$

Or just:

$$Renda_{it} = \alpha + \beta TD_{it} + w_{it} \quad \text{where:} \quad w_{it} = c_i + e_{it}$$

- The random effects estimator assumes that the intercept is defined by a random variable, rather than by parameters that can be estimated using fixed effects.
- Both Generalized Least Squares and Maximum Likelihood can be used to obtain the random effects estimator.

Random Effects Estimator

The random effects estimator assumes that the individual heterogeneity (c_i) is randomly distributed around a constant value (α). In other words:

$$Y_{it} = \alpha + \sum_{j=1}^{k} \beta_j X_{j_{it}} + c_i + e_{it}$$

Or:

$$\mathrm{Y}_{it} = \alpha + \sum_{j=1}^{k} \beta_j X_{j_{it}} + w_{it} \quad \text{where} \quad \mathrm{w}_{it} = c_i + e_{it}$$

The error w_{it} is assumed to be compounded of two components: i) random variation between the cross-sectional units (c_i); ii) idiosyncratic error, which is independent of the cross-sectional units and time periods.

Example – Stata & R

- Suppose we have a pooled data with information for the regressand *y* and two exogenous variables (*x*1 and *x*2) across two periods (t=0 and 1). The random effect estimator in Stata is given by:

```
* random effects estimator (one-way)
xtreg y x1 x2, re

* random effects estimator (two-way)
xtreg y x1 x2 i.time, re
```

- The equivalent in R:

```
# random effects model - one-way
re1 <- plm(y ~ x1 + x2, data=mydata, index=c("csunit","time"), model="random")
summary(re1)

# random effects model - two-way
re2 <- plm(y ~ x1 + x2 + as.factor(time), data=agric,
           index=c("csunit","time"),
           model="random")
summary(re2)
```

- The equivalent in Python:

```
# random effects eestimator - two-way
re1 = RandomEffects(mydata.y, mydata[['x1','x2']]).fit()
print(re1)
```

Fixed or Random Effects?

1) Degrees of freedom: the fixed effects estimator is operationally easier, but may consume a large number of degrees of freedom depending on the number of cross-sectional units. In this respect, the random effects estimator tends to be more efficient;

2) Variability of the regressor: the random effects estimator is particularly attractive when the regressors present low variability within each cross-section unit. The inefficiency of the fixed effects estimators makes more difficult to obtain significant estimates for the slope coefficients. In a extreme case where one regressor do not vary for the same cross-sectional unit (sex of the person, for example), only the random effects estimator can be applied;

Fixed or Random Effects?

3) Correlation between errors and regressors: the random effects estimators assumes that the individual heterogeneity c_i is part of the composite error w_{it}. But the errors w_{it} must be uncorrelated with the independent variables. This means that:

$$Cov(X_{it}, w_{it}) = 0 \text{ and } Cov(X_{it}, c_i) = 0$$

This is a strong assumption and means that unobservable characteristics that related to the cross-sectional units are not related to the independent variables. The fixed effects estimator does not require this assumption.

Specification Test - Hausman

The fixed effects estimator is more accurate than the random effects estimator, but less efficient (larger variance). In turn, the random effects estimator is more efficient than the fixed effects estimator, but may be biased. The Hausman specification test basically compares the parameters for the models with fixed ($\boldsymbol{\beta}_{FE}$) and random ($\boldsymbol{\beta}_{RE}$) effects:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_{EF} + \mathbf{w} \quad \text{and} \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta}_{EA} + \mathbf{w}$$

The null hypothesis is that

$$\mathbf{H0}:\ \boldsymbol{\beta}_{RE} = \boldsymbol{\beta}_{FE}$$

If we do not reject H0 ($\boldsymbol{\beta}_{RE}$= $\boldsymbol{\beta}_{fE}$) , then the fixed and random effects estimators are consistent. In this case, we may choose the random effects estimator because it is more efficient. If we do not reject H0 ($\boldsymbol{\beta}_{RE} \neq \boldsymbol{\beta}_{fE}$), then the fixed effects estimator is the only consistent and must be chosen.

The statistic used in this test is:

$$m = (\boldsymbol{\beta}_{EF} - \boldsymbol{\beta}_{EA})'(S_{\boldsymbol{\beta}_{EF}} - S_{\boldsymbol{\beta}_{EA}})^{-1}(\boldsymbol{\beta}_{EF} - \boldsymbol{\beta}_{EA})$$

Which has a χ^2 distribution with k degrees of freedom, where k is the number of independent variables.

Example – Stata & R

- The Hausman test in Stata:

```
* random effects estimator (one-way)
xtreg y x1 x2, fe
estimates store fe

* random effects estimator (one-way)
xtreg y x1 x2, re
estimates store re

* hausman test
hausman fe re
```

- The equivalent in R:

```
# fixed effects model - within transformation
fe <- plm(y ~ x1 + x2, data=mydata, index=c("csunit","time"), model="within")
summary(fe)

# random effects model - one-way
re1 <- plm(y ~ x1 + x2, data=mydata, index=c("csunit","time"), model="random")
summary(re)

# hausman test
phtest(fe,re)
```

Exercise

1) The dataset *Data_AgricultureClimate.xls* contains information on agricultural production and climate variables (GORI MAIA, A., MIYAMOTO, B. C, GARCIA, J. R. Climate change and agriculture: Do environmental preservation and ecossystem services matter? Ecoloogical Economics, v. 152 (October 2018), 2018):
   a) Use random effects estimator to analyze the relation between (log) the total value of production, (log) area, temperature and precipitation;
   b) Compare the estimates obtained by fixed and random effects;