

# Phd Program in Transportation

## Transport Demand Modeling

Anabela Ribeiro - 2021

[anabela@dec.uc.pt](mailto:anabela@dec.uc.pt)

PANEL MODELS



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# General Framework



## Class Structure

	Dia	Description
1	17 November 2021 (3 hours)	<p>A - Panel data models: Main issues One way component error models Two way component error models <i>Exercises and notes on the Home Assignment</i> <i>Panel Data Models</i></p> <p>B – Spatial regression models: Main issues Exploratory Spatial Data Analysis</p>
2	19 November 2021 (3 hours)	<p>B – Spatial regression models: Spatial Regression Models <i>Exercises and notes on the Home Assignment</i> <i>Spatial Data Models</i></p>

# General Framework



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Regression types

**Time Series** - Regression between a set of variables across time in a certain spatial unit - cross-sectional data in time

**Cross section** – Regression between a set of variables in a certain moment in time, for a set of territorial units or individuals

**Panel Data** – Regression analysis that combines spatial data with time data.

1 - **Panel Data** – Cross section with time series

2 - **Spatial Models** – Cross section with spatial relation variables

3 - **Panel Data** with spatial relation variables ....(not on these classes)

# Panel Data Models



## Main issues

One way component error models

Two way component error models

# Panel Data Models: Main Issues



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Advantages

### Longitudinal or cross-sectional time-series data

Panel data are better able to study the **dynamics of adjustment**. Cross-sectional distributions that look relatively stable hide a multitude of changes

Spells of unemployment, job turnover, residential and income mobility are better studied with panels. Panel data are also well suited to study the duration of economic states like **unemployment and poverty**, and if these panels are long enough, they can shed light on the **speed of adjustments to economic policy changes**.

Micro panel data gathered on variations for individuals, firms and households may be measured more accurately than those variations at the macro level, capturing other variations in presence and avoiding biases.

*Biases resulting from aggregation over firms or individuals may be reduced or eliminated, especially in life cycle models.*

# Panel Data Models: Main Issues

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...

If we do a simple regression, that regression is not able to capture variations related with differences across countries, related with other factors not present in the regression.

How to identify if those variations are significant?



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: Main issues



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Advantages - Example

For example, suppose that a cross-section of public transit agencies reveals that, on average, public transit subsidies are associated with 20% increased ridership.

If a homogeneous population of public transit firms is considered, this might lead to think that a firm's ridership will increase by 20% given transit subsidies

However, an alternative explanation in a sample of heterogeneous public transit firms is that the subsidies have no effect (0% increase) on four fifths of the firms, and raise ridership by 100% on one fifth of the firms. Although these competing hypotheses cannot be tested using a cross-sectional sample (**in the absence of a cross-sectional variable/parameter that “explains” this difference**), it is possible to test between them by identifying the effect of subsidies on a **cross section of time series for the different transit systems**. Thus, **testing for cross-sectional homogeneity is equivalent to testing the null hypothesis that these additional parameters are equal to zero**

# Panel Data Models: Main Issues



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

**Disadvantages – several new specification issues to be accounted for**

**Heterogeneity**

**Correlation in the disturbance terms**

**Heteroscedasticity**

Probably related with the existence of groups with similar behavior among their elements, and with a significantly different behavior from other groups.



# Panel Data Models: Main Issues



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Disadvantages – new specification issues

**Heterogeneity** - Compared with cross-sectional or time-series data, panel data raise new specification issues that need to be considered during analysis

The most important of these is **heterogeneity** bias

**Heterogeneity** refers to the differences **across cross-sectional units** that may not be appropriately reflected in the available data (explanatory variable/s).

**Heterogeneity** - if not accounted for explicitly, may lead to model parameters that are inconsistent and/or meaningless. With panel data, it is possible to account for cross-sectional **heterogeneity** by **introducing additional parameters into the model**.

**Heterogeneity** in panel data can be tested using several test of hypothesis – Z Test (if the distributions follow a normal distribution).

# Panel Data Models: Main Issues



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Disadvantages – new specification issues

**Correlation in the disturbance terms** - A second issue is serial correlation of the disturbance terms, which occurs in **time-series** studies when the disturbances associated with observations in one time period are dependent on disturbances from prior time periods. It can be with:

Positive correlation – the estimates of the standard errors are smaller than the true standard errors, or

Negative correlation – the estimates of the standard errors are bigger than the true standard errors.

**The regression estimates are unbiased but **its errors are biased**. It does not affect consistency of regression but **it affects its efficiency**.**

# Panel Data Models: Main Issues



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Disadvantages – new specification issues

**Heteroscedasticity**, which refers to the variance of the disturbances not being constant across observations.

It also affects the efficiency of the estimated parameters.

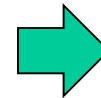


?

**Heterogeneity**

**Correlation in the disturbance terms**

**Heteroscedasticity**



**One Way and Two Way  
component models**

# Panel Data Models: Main Issues



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Overcoming specification problems in panel data:

**One-Way** error component models: variable-intercept models across individuals or time

**Two-Way** error component models: variable-intercept models across individuals and time

## Modeling specifications:

**With fixed effects:** effects that are in the sample

One way

Two ways

**With random effect:** effect randomly drawn from a population

One way

Two ways

# Panel Data Models: Main Issues



One way-models: variable-intercept models across individuals **or** time.

Two-way models: variable-intercept models across **both** individuals and time (two-way models)



FEUP



**Introduction of dummy variables in the model**

# Panel Data Models: Main Issues



Variable-intercept models across individuals or time (one-way models) and across both individuals and time (two-way models), are the simplest and most straightforward models for **accounting for cross-sectional *heterogeneity*** in panel data, which arise when the null hypothesis of overall ***homogeneity*** is rejected

- The variable-intercept model assumes that the effects of omitted variables may be individually unimportant but are collectively significant, and thus is considered to be a random variable that is independent of included independent variables.
- Because **heterogeneity effects are assumed to be constant for given cross-sectional units or for different cross-sectional units during one time period**, they are absorbed by the intercept term as a means to account for individual or time heterogeneity

# Panel Data Models: One-Way Error Component Models

## Fixed Effects

FE explore the relationship between predictor and outcome variables **within** an entity (country, person, company, etc.).

Each entity has its own individual characteristics that may or may not influence the predictor variables (for example, being a male or female)

When using FE we assume that something within the individual may impact or bias the predictor or outcome variables and **we need to control for this**.

This is the rationale behind the assumption of the correlation between entity's error term and predictor variables.

FE remove the effect of those time-invariant characteristics so we can **assess the net effect of the predictors on the outcome variable**.



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: One-Way Error Component Models

## Fixed Effects

Another important assumption of the FE model is that those time invariant characteristics are unique to the individual and should not be correlated with other individual characteristics.

Each entity is different therefore the entity's error term and the constant (which captures individual characteristics) should not be correlated with the others.

If the error terms are correlated, then FE is no suitable since inferences may not be correct and you need to model that relationship (**probably using random-effects**), this is the main rationale for the **Hausman test** .

Fixed-effects models are designed to study the causes of changes within a person or entity.



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP



# Panel Data Models: One-Way Error Component Models

## Fixed Effects

More formally, a panel data regression is written as:

$$Y_{it} = \alpha + X'_{it}\beta + u_{it}, i = 1, \dots, n; t = 1, \dots, T \quad (1)$$

where  $i$  refers to the cross-sectional units,  $t$  refers to the time periods,  $\alpha$  is a scalar,  $\beta$  is a  $P \times 1$  vector, and  $X_{it}$  is the  $i$ th observation on the  $P$ th explanatory variable

A **one-way error component model** for the disturbances, which is the most commonly utilized in panel data formulation, is specified as

$$\mu_{it} = \mu_i + v_{it} \quad (2)$$

where  $\mu_i$  is the unobserved cross-sectional specific effect and  $v_{it}$  are random disturbances



# Panel Data Models: One-Way Error Component Models

## Fixed Effects

When the  $\mu_i$  are assumed to be **fixed parameters** to be estimated, and the  $v_{it}$  are random disturbances that follow the usual regression assumptions, then combining both yields the following model, **where inference is conditional on the particular  $n$  cross-sectional units that are observed**, and is thus called a **fixed effects model**

$$Y_{it} = \alpha + X'_{it}\beta + \mu_i + v_{it}, i = 1, \dots, n; t = 1, \dots, T \quad (3)$$



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: One-Way Error Component Models

## Fixed Effects

**Estimation** on which ordinary least squares (OLS), which provide best linear unbiased estimators (BLUE), are used to obtain  $\alpha$ ,  $\beta$  and  $\mu_i$

- **Large Samples:** when  $n$  is large, many indicator's variables are included in the model, and the matrices to be inverted by OLS are of large dimension ( $n + P$ ). As such, a **least squares dummy variable (LSDV)** estimator for (3) is obtained for  $\beta$  (this estimator is also called the within-group estimator because only the variation within each group is utilized in forming the estimator)

$$Y_{it} = \alpha_i A_i + X'_{it} \beta + \mu_i + v_{it}, i = 1, \dots, n; t = 1, \dots, T \quad (4)$$

Unobserved time-invariant individual effect



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: One-Way Error Component Models

## Fixed Effects

Testing for the joint significance of the included fixed effects parameters (the dummy variables) is straightforwardly conducted using the **Chow F test**

$$F_0 = \frac{(RRSS - URSS) / (n - 1)^{H_0}}{URSS / (nT - n - P)} \sim F_{n-1, n(T-1)-P} \quad (5)$$

where RRSS are the restricted residual sums of squares from OLS on the pooled model and URSS are the unrestricted residual sums of squares from the LSDV regression

**If the null is true (no fixed effects) then the correct procedure is to estimate a single regression from all the data.**

**If the null is not true (a significant value for F) then we have to account for fixed effects.**



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: One-Way Error Component Models

## Random Effects

There are too many parameters in the fixed effects model and the loss of degrees of freedom can be avoided if the  $\mu_i$  can be assumed random. Unlike the fixed effects model where inferences are conditional on the particular cross-sectional units sampled, an alternative formulation, called the **random effects model** can help solving the problem.

$$\mu_i \sim IDD(0, \sigma_\mu^2), v_{it} \sim IDD(0, \sigma_v^2) \quad (6)$$

- The  $\mu_i$  and  $v_{it}$  are independent, and  $X_{it}$  are independent of the  $\mu_i$  and  $v_{it}$  for all  $i, t$



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: One-Way Error Component Models



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Random Effects

The random effects model is an appropriate specification if we are drawing  $n$  individuals randomly from a large population

This is usually the case for household panel studies. Care is taken in the design of the panel to make it “representative” of the population we are trying to make inferences about

The individual effect is characterised as random and inference pertains to the population from which this sample was randomly drawn.

Furthermore, it can be shown that a random effects specification implies a **homoscedastic** disturbances variance,  $VAR(u_{it}) = \sigma_{\mu}^2 + \sigma_v^2$  for all  $i, t$ , and **serial correlation** only for disturbances of the same cross-sectional unit.

# Panel Data Models: One-Way Error Component Models



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Random Effects

In general, the following holds for a random effects model

homoscedastic disturbances variance

$$\begin{aligned} COV(u_{it}, u_{js}) &= \sigma_{\mu}^2 + \sigma_{\nu}^2 && \text{for } i = j, t = s \\ &= \sigma_{\mu}^2 && \text{for } i = j, t \neq s \\ &= 0 && \text{otherwise} \end{aligned} \quad (7)$$

serial correlation only for disturbances of the same cross-sectional unit

$$\begin{aligned} COR(u_{it}, u_{js}) &= 1 && \text{for } i = j, t = s \\ &= \frac{\sigma_{\mu}^2}{(\sigma_{\mu}^2 + \sigma_{\nu}^2)} && \text{for } i = j, t \neq s \\ &= 0 && \text{otherwise} \end{aligned} \quad (8)$$

# Panel Data Models: One-Way Error Component Models

## Fixed vs. Random Effects

This is not as easy a choice as it might seem. In fact, [the fixed versus random effects issue has generated a hot debate in the biometrics and statistics literature which has spilled over into the panel data econometrics literature....](#)

Most commonly accepted:

The most important issue when considering these alternative specifications is the context of the analysis. In the fixed-effects model, inferences are conditional on the effects that are in the sample, while in the random-effects model inferences are made unconditionally with respect to the population of effects (Hsiao, 1986). In other words, [the essential difference between these two modeling specifications is whether the inferences from the estimated model are confined to the effects in the sample or whether the inferences are made about a population of effects \(from which the effects in the model are a random sample\).](#) In the former case the fixed effects model is appropriate, whereas the latter is suited for the random effects model.



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP



# Panel Data Models: One-Way Error Component Models



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Fixed vs. Random Effects (Lagrange and Hausman Tests)

The fixed effects model has a considerable virtue in that **it does not assume that the individual effects are uncorrelated with the regressors**,  $E(u_{it} | X_{it}) = 0$ , as is assumed by the random-effects model.

In fact, the random-effects model may be biased and inconsistent due to omitted variables (Hausman and Taylor, 1981; Chamberlain, 1978).

With the intent of **identifying potential correlation between the individual effects and the regressors**, Hausman (1978) devised a test to examine the **null hypothesis of no correlation between the individual effects and the regressors**  $X_{it}$

A rejection of the null hypothesis of no correlation suggests the possible inconsistency of the random effects model and the possible preference for a fixed-effects specification (test value significant).

# Panel Data Models: One-Way Error Component Models



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Fixed vs. Random Effects (Lagrange and Hausman Tests)

So, in the process, and starting with classical OLS, we estimate:

1) First, if the fixed effects holds better than the no effects model (OLS)

*And for that we use a Lagrange Multiplier Test*

2) Second, if the random effects model can be used i.e. individual effects are uncorrelated with the regressors.

*And for that we use the Hausman test.*



# Panel Data Models: One-Way Error Component Models



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Fixed vs. Random Effects (Lagrange and Hausman Tests)

### Lagrange Multiplier (LM):

Breusch and Pagan's Lagrange multiplier statistic, is used to test the null hypothesis that there are no group effects in the random effects model. *Arguably, a rejection of the null hypothesis is as likely to be due to the presence of fixed effects.* The statistic is computed from the ordinary least squares residuals from a pooled regression. *Large values of LM favor the effects model over the classical model with no common effects.*

### Hausman chi-square (H):

A second statistic is Hausman's chi squared statistic for testing whether the GLS estimator is an appropriate alternative to the LSDV estimator. Computation of the Hausman statistic requires estimates of both the random and fixed effects models. *Large values of H weigh in favor of the fixed effects model.*

# Panel Data Models: Two-Way Error Component Models

## Fixed Effects

The disturbances presented in (2) are further generalized to include **time-specific effects**. This generalization is called a two-way error components model, whose disturbances are written as

$$u_{it} = \mu_i + \lambda_t + v_{ij}, i = 1, \dots, n; t = 1, \dots, T \quad (10)$$

where  $\mu_i$  is the unobserved cross-sectional specific effect,  $\lambda_t$  denotes the **unobservable time effects**, and  $v_{it}$  are random disturbances. Here  $\lambda_t$  is individual invariant and accounts for any time-specific effect that is not included in the regression

When the  $\mu_i$  and  $\lambda_t$  are assumed to be fixed parameters to be estimated and are random disturbances that follow the usual regression assumptions, combining (1) and (10) yields **a model where inferences are conditional on the particular  $n$  cross-sectional units and are to be made over the specific time period of observation**

$$u_{it} = \mu_i + v_{ij}$$

where  $\mu_i$  is the unobserved cross-sectional specific effect and  $v_{it}$  are random disturbances



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: Two-Way Error Component Models

## Fixed Effects

This model is called a **two-way fixed effects error component model** and is given as

$$Y_{it} = \alpha + X'_{it}\beta + \mu_i + \lambda_t + v_{it}, i = 1, \dots, n; t = 1, \dots, T \quad (11)$$

where  $X_{it}$  are assumed independent of the  $v_{it}$  for all  $i, t$ . Inference for this **two-way fixed-effects model** is conditional on the particular  $n$  individuals and over the  $T$  time periods of observation. Similar to the one-way fixed-effects model, the computational difficulties involved with obtaining the OLS estimates for  $\beta$ .



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: Two-Way Error Component Models

## Fixed Effects

Normally the coefficients are estimated using a within transformation of Wallace and Hussain (1969) to use *generalised least squares* (GLS)

***GLS is equivalent to applying ordinary least squares to a linearly transformed version of the data.*** The GLS is applied when the variances of the observations are unequal (*heteroscedasticity*), or when there is a certain degree of *correlation between the observations (collinearity)* . In these cases ordinary least squares can be statistically inefficient, or even give misleading inferences.



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: Two-Way Error Component Models

## Fixed Effects

Testing for the joint significance of the included cross-sectional and time period fixed effects parameters (the dummy variables) is straightforwardly computed using an  $F$  test

$$F_0 = \frac{(RRSS - URSS) / (n + T - 2)^{H_0}}{URSS / (n - 1)(T - 1) - P} \sim F_{(n+T-2), (n-1)(T-1)-P} \quad (12)$$

where RRSS are the restricted residual sums of squares from OLS on the pooled model and URSS are the unrestricted residual sums of squares from the regression using the within transformation of Wallace and Hussain (1969)



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: Two-Way Error Component Models

## Random Effects

Similar to the one-way error component model case, if both the  $\mu_i$  and  $\lambda_t$  are random with

$$\mu_i \sim IDD(0, \sigma_\mu^2), \lambda_t \sim (0, \sigma_\lambda^2), v_{it} \sim IDD(0, \sigma_v^2) \quad (13)$$

The  $\mu_i$ ,  $\lambda_t$  and  $v_{it}$  are independent, and  $X_{it}$  are independent of the  $\mu_i$ ,  $\lambda_t$  and  $v_{it}$  for all  $i, t$ . This formulation is called the ***random-effects model***



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP



# Panel Data Models: Two-Way Error Component Models



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Random Effects

Furthermore, it can be shown that a random effects specification implies a **homoscedastic disturbances variance**,  $VAR(u_{it}) = \sigma_{\mu}^2 + \sigma_{\lambda}^2 + \sigma_{\nu}^2$  for all  $i, t$ , and **serial correlation only** for disturbances of the same cross-sectional unit.

In general, the following holds for a random effects model

$$\begin{aligned} COV(u_{it}, u_{js}) &= \sigma_{\mu}^2 && \text{for } i = j, t = s \\ &= \sigma_{\lambda}^2 && \text{for } i = j, t \neq s \\ &= 0 && \text{otherwise} \end{aligned} \quad (14)$$

and

$$\begin{aligned} COR(u_{it}, u_{js}) &= 1 && \text{for } i = j, t = s \\ &= \frac{\sigma_{\mu}^2}{(\sigma_{\mu}^2 + \sigma_{\lambda}^2 + \sigma_{\nu}^2)} && \text{for } i = j, t \neq s \\ &= 0 && \text{otherwise} \end{aligned} \quad (15)$$

# Panel Data Models: Two-Way Error Component Models



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Random Effects

Estimation of the two-way random-effects model is typically accomplished using the **GLS estimators** of Wallace and Hussain (1969) or by using maximum likelihood estimation (Baltagi and Li, 1992)

For this model specification, Breusch and Pagan (1980) derived a Lagrange-multiplier test for the null hypothesis;  $H_0 = \sigma_{\mu}^2 = \sigma_{\lambda}^2 = 0$  this test is based on the normality of the disturbances

If the two-way component model specification is significant, the quality of its estimates should be always better than a one-way Error Component model in Fixed Effects or Random Effects model.

As before, using the Hausman test, we can improve the estimation by using the Fixed or the random effects model.

# Panel Data Models: Example 1



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Nlogit with Grunfeld Investment Equation

$$I_{it} = \alpha + \beta_1 F_{it} + \beta_2 C_{it} + v_{it} \quad (16)$$

where  $I_{it}$  denotes real gross investment for firm  $i$  in year  $t$ ,  $F_{it}$  is the real value of the firm (shares outstanding) and  $C_{it}$  is the real value of the capital stock. These panel data consist of **10** large US manufacturing firms over **20** years, 1935–54

**Table 1**

Variable Abbreviation	Variable Description
invest	Gross investment, defined as additions to plant and equipment plus maintenance and repairs in millions of dollars deflated by the implicit price deflator of producers' durable equipment (base 1947)
value	Market value of the firm, defined as the price of common shares at December 31 (base 1947)
capital	Stock of plant and equipment, defined as the accumulated sum of net additions to plant and equipment deflated by the implicit price deflator for producers' durable equipment (base 1947)
firm	General Motors (GM), US Steel (US), General Electric (GE), Chrysler (CH), Atlantic Rening (AR), IBM, Union Oil (UO), Westinghouse (WH), Goodyear (GY), Diamond Match (DM), American Steel (AS)
year	Year of data
firmcod	Numeric code that identifies each firm

# Panel Data Models: Example 1



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Nlogit with Grunfeld Investment Equation

1° - *Ordinary least squares*

2° - *Fixed effects?*

*one way*

*two way*

3° - *Random?*

*one way*

*two way*

# Panel Data Models: Example 1



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Ordinary Least Squares Model Estimates Gross Investment

```

Ordinary least squares regression
Model was estimated Oct 15, 2010 at 09:10:35AM
LHS=INVEST Mean = 133.3119
Standard deviation = 210.5872
WTS=none Number of observs. = 220
Model size Parameters = 3
Degrees of freedom = 217
Residuals Sum of squares = 1768678.
Standard error of e = 90.28063
Fit R-squared = .8178870
Adjusted R-squared = .8162086
Model test F[ 2, 217] (prob) = 487.28 (.0000)
Diagnostic Log likelihood = -1301.299
Restricted(b=0) = -1488.643
Chi-sq [ 2] (prob) = 374.69 (.0000)
Info criter. LogAmemiya Prd. Crt. = 9.019390
Akaike Info. Criter. = 9.019388
Autocorrel Durbin-Watson Stat. = .3566636
Rho = cor[e,e(-1)] = .8216682
    
```

Variable	Coefficient	Standard Error	t-ratio	P[ T >t]	Mean of X
Constant	-38.4100540	8.41337092	-4.565	.0000	
VALUE	.11453436	.00551883	20.753	.0000	988.577805
CAPITAL	.22751413	.02422825	9.390	.0000	257.108541

# Panel Data Models: Example 1



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Fixed Effects Panel Data Model Estimates (One-way Error) Gross Investment

		Estimated Fixed Effects			
		Group	Coefficient	Standard Error	t-ratio
Least Squares with Group Dummy Variables		1	-70.29907	47.37535	-1.48387
Ordinary least squares regression		2	101.90474	23.76871	4.28735
Model was estimated Nov 05, 2012 at 06:09:19PM		3	-235.56939	23.28607	-10.11632
LHS=INVEST Mean =	133.3119	4	-27.80911	13.41858	-2.07243
Standard deviation =	210.5872	5	-114.60252	13.50246	-8.48753
WTS=none Number of observs. =	220	6	-23.16020	12.07589	-1.91789
Model size Parameters =	13	7	-66.54422	12.24204	-5.43572
Degrees of freedom =	207	8	-57.54649	13.33791	-4.31451
Residuals Sum of squares =	523718.7	9	-87.21454	12.28873	-7.09712
Standard error of e =	50.29952	10	-6.56803	11.27363	-.58260
Fit R-squared =	.9460750	11	-20.57820	11.29779	-1.82144
Adjusted R-squared =	.9429489				
Model test F[ 12, 207] (prob) =	302.64 (.0000)				
Diagnostic Log likelihood =	-1167.426				
Restricted (b=0) =	-1488.643				
Chi-sq [ 12] (prob) =	642.44 (.0000)				
Info criter. LogAmemiya Prd. Crt. =	7.893402				
Akaike Info. Criter. =	7.893264				
Estd. Autocorrelation of e(i,t) =	.549274				

Variable	Coefficient	Standard Error	t-ratio	P[ T >t]	Mean of X
VALUE	.11012912	.01129984	9.746	.0000	988.577805
CAPITAL	.31003344	.01654048	18.744	.0000	257.108541

# Panel Data Models: Example 1

## Fixed Effects Panel Data Model Estimates (One-way Error) Gross Investment

Test Statistics for the Classical Model							
Model	Log-Likelihood		Sum of Squares		R-squared		
(1) Constant term only	-1488.64328		.9711984910D+07		.0000000		
(2) Group effects only	-1327.50901		.2244546885D+07		.7688890		
(3) X - variables only	-1301.29920		.1768678402D+07		.8178870		
(4) X and group effects	-1167.42554		.5237186622D+06		.9460750		

Hypothesis Tests							
Likelihood Ratio Test				F Tests			
	Chi-squared	d.f.	Prob.	F	num.	denom.	P value
(2) vs (1)	322.269	10	.00000	69.533	10	209	.00000
(3) vs (1)	374.688	2	.00000	487.284	2	217	.00000
(4) vs (1)	642.435	12	.00000	302.639	12	207	.00000
(4) vs (2)	320.167	2	.00000	340.079	2	207	.00000
(4) vs (3)	267.747	10	.00000	49.207	10	207	.00000

Consideration of group effects improve regression



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: Example 1



INSTITUTO SUPERIOR TÉCNICO



FEUP

## Fixed Effects Panel Data Model Estimates (Two-way Error) Gross Investment

```

+-----+
| Least Squares with Group and Period Effects
| Ordinary least squares regression
| Model was estimated Nov 07, 2012 at 01:14:13AM
| LHS=INVEST Mean = 133.3119
| Standard deviation = 210.5872
| WTS=none Number of observs. = 220
| Model size Parameters = 32
| Degrees of freedom = 188
| Residuals Sum of squares = 459399.9
| Standard error of e = 49.43295
| Fit R-squared = .9526976
| Adjusted R-squared = .9448978
| Model test F[ 31, 188] (prob) = 122.14 (.0000)
| Diagnostic Log likelihood = -1153.012
| Restricted(b=0) = -1488.643
| Chi-sq [ 31] (prob) = 671.26 (.0000)
| Info criter. LogAmemiya Prd. Crt. = 7.937036
| Akaike Info. Criter. = 7.934958
| Estd. Autocorrelation of e(i,t) .569917
+-----+
  
```

$\mu_i$

Estimated Fixed Effects - Full sets of effects, normalized t

Group	Coefficient	Standard Error	t-ratio
1	-53.14755	42.84415	-1.24049
2	149.17125	16.34968	9.12380
3	-192.46291	15.79790	-12.18282
4	35.02275	11.28051	3.10471
5	-63.87868	16.09003	-3.97008
6	42.16480	12.69795	3.32060
7	-8.17149	15.43398	-.52945
8	6.90560	11.49035	.60099
9	-29.34076	13.70930	-2.14021
10	65.11490	15.55637	4.18574
11	48.62208	15.58761	3.11928

Estimated Fixed Effects - Full sets of effects, normalized t

Period	Coefficient	Standard Error	t-ratio
1	41.85916	15.38287	2.72115
2	24.89993	15.03606	1.65601
3	5.48352	15.46655	.35454
4	6.23543	14.87055	.41931
5	-21.24024	14.71160	-1.44377
6	2.03439	14.74346	.13799
7	25.37139	14.65503	1.73124
8	23.85983	14.71735	1.62120
9	4.08671	14.58562	.28019
10	3.53909	14.57617	.24280
11	-7.68033	14.59933	-.52607
12	14.10477	14.64576	.96306
13	6.98162	14.67759	.47567
14	3.52843	14.82120	.23807
15	-23.34160	14.88759	-1.56786
16	-25.52857	14.85283	-1.71877
17	-12.97548	14.79992	-.87673
18	-14.62988	15.04933	-.97213
19	-16.65342	15.74591	-1.05763
20	-39.93475	16.15229	-2.47239

Variable	Coefficient	Standard Error	t-ratio	P[ T >t]	Mean of X
VALUE	.11668113	.01293303	9.022	.0000	988.577805
CAPITAL	.35143569	.02104860	16.696	.0000	257.108541
Constant	-72.3935959	12.7315764	-5.686	.0000	

$\lambda_t$



# Panel Data Models: Example 1

## Fixed Effects Panel Data Model Estimates (Two-way Error) Gross Investment

Test Statistics for the Classical Model							
Model	Log-Likelihood	Sum of Squares	R-squared				
(1) Constant term only	-1488.64328	.9711984910D+07	.0000000				
(2) Group effects only	-1327.50901	.2244546885D+07	.7688890				
(3) X - variables only	-1301.29920	.1768678402D+07	.8178870				
(4) X and group effects	-1167.42554	.5237186622D+06	.9460750				
(5) X ind.&time effects	-1153.01185	.4593999310D+06	.9526976				

Hypothesis Tests							
Likelihood Ratio Test				F Tests			P value
	Chi-squared	d.f.	Prob.	F	num.	denom.	
(2) vs (1)	322.269	10	.00000	69.533	10	209	.00000
(3) vs (1)	374.688	2	.00000	487.284	2	217	.00000
(4) vs (1)	642.435	12	.00000	302.639	12	207	.00000
(4) vs (2)	320.167	2	.00000	340.079	2	207	.00000
(4) vs (3)	267.747	10	.00000	49.207	10	207	.00000
(5) vs (4)	28.827	19	.06875	1.385	19	188	.13801
(5) vs (3)	296.575	30	.00000	17.860	30	188	.00000

Consideration of group effects and time effects improve regression



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

# Panel Data Models: Example 1



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Random Effects Panel Data Model Estimates (One-way Error) Gross Investment

Least Squares with Group Dummy Variables		Random Effects Model: $v(i,t) = e(i,t) + u(i)$	
Ordinary	least squares regression	Estimates: Var[e]	= .253004D+04
Model was estimated Oct 15, 2010 at 09:35:46AM		Var[u]	= .562055D+04
LHS=INVEST	Mean = 133.3119	Corr[v(i,t),v(i,s)]	= .689588
	Standard deviation = 210.5872	<div style="border: 1px solid black; padding: 5px;">                     Lagrange Multiplier Test vs. Model (3) = 874.75                      ( 1 df, prob value = .000000)                      (High values of LM favor FEM/REM over CR model.)                      Baltagi-Li form of LM Statistic = 874.75                      Fixed vs. Random Effects (Hausman) = 2.89                      ( 2 df, prob value = .235736)                      (High (low) values of H favor FEM (REM).)                 </div>	
WTS=none	Number of observs. = 220		
Model size	Parameters = 13		
	Degrees of freedom = 207		
Residuals	Sum of squares = 523718.7		
	Standard error of e = 50.29952		
Fit	R-squared = .9460750		
	Adjusted R-squared = .9429489		
Model test	F[ 12, 207] (prob) = 302.64 (.0000)		
Diagnostic	Log likelihood = -1167.426		
	Restricted(b=0) = -1488.643		
	Chi-sq [ 12] (prob) = 642.44 (.0000)		
Info criter.	LogAmemiya Prd. Crt. = 7.893402		
	Akaike Info. Criter. = 7.893264		
Estd. Autocorrelation of e(i,t)	.549274	Sum of Squares	.188431D+07
		R-squared	.808040D+00

Variable	Coefficient	Standard Error	b/St.Er.	P[ Z >z]	Mean of X
VALUE	.10924931	.00978586	11.164	.0000	988.577805
CAPITAL	.30782652	.01634860	18.829	.0000	257.108541
Constant	-53.8343750	24.5716850	-2.191	.0285	

# Panel Data Models: Example 1

## Random Effects Panel Data Model Estimates (Two-way Error) Gross Investment



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

Least Squares with Group and Period Effects		Random Effects Model: $v(i,t) = e(i,t) + u(i) + w(t)$	
Ordinary	least squares regression	Estimates:	Var[e] = .244362D+04
Model was estimated	Oct 15, 2010 at 10:10:47AM		Var[u] = .447819D+04
LHS=INVEST	Mean = 133.3119		Corr[v(i,t),v(i,s)] = .646968
	Standard deviation = 210.5872		Var[w] = .122879D+04
WTS=none	Number of observs. = 220		Corr[v(i,t),w(i,t)] = .334600
Model size	Parameters = 32	Lagrange Multiplier Test vs. Model (3) = 881.07	( 2 df, prob value = .000000)
	Degrees of freedom = 188	(High values of LM favor FEM/REM over CR model.)	
Residuals	Sum of squares = 459399.9	Fixed vs. Random Effects (Hausman) = 5.72	( 2 df, prob value = .057275)
	Standard error of e = 49.43295	(High (low) values of H favor FEM (REM).)	
Fit	R-squared = .9526976	Sum of Squares	.186431D+07
	Adjusted R-squared = .9448978	R-squared	.808040D+00
Model test	F[ 31, 188] (prob) = 122.14 (.0000)		
Diagnostic	Log likelihood = -1153.012		
	Restricted(b=0) = -1488.643		
	Chi-sq [ 31] (prob) = 671.26 (.0000)		
Info criter.	LogAmemiya Prd. Crt. = 7.937036		
	Akaike Info. Criter. = 7.934958		
Estd. Autocorrelation of e(i,t)	.569917		

Variable	Coefficient	Standard Error	b/St. Er.	P[ Z >z]	Mean of X
VALUE	.11107050	.01021747	10.871	.0000	988.577805
CAPITAL	.33700305	.01975302	17.061	.0000	257.108541
Constant	-63.1362933	23.9608695	-2.635	.0084	

# Panel Data Models: Example 1



FEUP

```
Random Effects Model:  $v(i,t) = e(i,t) + u(i) + w(t)$ 
Estimates:  Var[e]           = .244362D+04
            Var[u]           = .447819D+04
            Corr[v(i,t),v(i,s)] = .646968
            Var[w]           = .122879D+04
            Corr[v(i,t),v(j,t)] = .334600
Lagrange Multiplier Test vs. Model (3) = 881.07
( 2 df, prob value = .000000)
(High values of LM favor FEM/REM over CR model.)
Fixed vs. Random Effects (Hausman)     = 5.72
( 2 df, prob value = .057275)
(High (low) values of H favor FEM (REM).)
            Sum of Squares     .186431D+07
            R-squared          .808040D+00
```

Large values of LM favor the effects model over the classical model with no common effects.

Large values of H weigh in favor of the fixed effects model.

A large value of the LM statistic in the presence of a small H statistic (as in our application) argues in favor of the random effects model.

**“LIMDEP, Version 9, Student , Reference Guide”**

# Panel Data Models: Example 1



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

## Nlogit with Grunfeld Investment Equation

1° - Ordinary least squares (OLS)

2° - Fixed effects?

*one way > yes, better than OLS*

*two way > yes, better than OLS and than one way*

3° - Random?

*one way > Big L, small H – yes, maybe random*

*two way > Big L, significant yet still small H – yes, maybe random*

*Final answer: Maybe we should stay with Random effects two-way model.*

*In alternative, Fixed effects with two effects might be not so bad.*

# Panel Data Models: Bibliography



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP

Washington, S., Karlaftis, M., Mannering, F. (2003), *Statistical and Econometric Methods for Transportation Data Analysis*, Chapman & Hall/CRC.

Wooldridge, J. (2001), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Massachusetts.

Baltagi, B. (2005), *Econometric Analysis of Panel Data*, Third Edition, John Wiley & Sons, Ltd., Chichester.

\_\_\_\_\_(1996), *A Handbook of the Theory with Applications*, *Advanced Studies in Theoretical and Applied Econometrics*, Vol. 33, Mátyás, László; Sevestre, Patrick (Eds.), Springer.

Hsiao, C. (2003), *Analysis of Panel Data*, Second Edition, Cambridge University Press, Cambridge.

Frees, E. (2004), *Longitudinal and Panel Data, Analysis and Applications in the Social Sciences*, Cambridge University Press, Cambridge.

# Phd Program in Transportation

## Transport Demand Modeling

Anabela Ribeiro - 2020

[anabela@dec.uc.pt](mailto:anabela@dec.uc.pt)

PANEL MODELS



INSTITUTO  
SUPERIOR  
TÉCNICO



FEUP