EUSERS SUMMER SCHOOL

Performance and Governance of Services of General Interest.

Critical perspectives on Energy, Telecommunications, Transport and Water Reforms in the EU

***June, 27th – July 1st* 2016**

Static Panel Data Analysis
A step by step practical guideline

Stefano Clò

Università degli studi di Milano,

Department of Economics, Management and Quantitative Methods

JEAN MONNET NETWORK

Co-funded by the European Union

Roadmap

- Advantages of Panel data
- How to prepare panel data
  - Find and download data (example with OECD ETCR indicator)
  - Import, clean and organize data
  - Build final dataset
- Choose a proper panel data model
  - Run a regression and interpret the results correctly
  - Pooled OLS, Fixed Effects or Random Effects?
  - Hausman Test
- Estimate my regression
  - How to report the results in a professional manner

Panel data Analysis

$$y_{it} = \alpha_i + X'_{it}\beta + \varepsilon_{it}$$

- α*:* constant term (intercept)
- X_{it}: set of independent variables x_{it} likely to affect dependent variable
- ε_{it}: error term, everything that is not captured by independent variable
- β : measures partial effect of x_{it} in period *t* for unit *i*
- $i = 1,\ldots,I$ number of groups
- $t = 1,\ldots,T$ number of time periods
- *Cross-section analysis* if $t = 1$ & $i > 1$
  - Observation of multiple groups in a single time period
- *Time-series analysis* if $t > 1$ & $i = 1$
  - Observation of one single group in several different time periods
- *Panel data analysis* if $t > 1$ & $i > 1$
  - Observation of multiple groups in several different time periods

Panel data Analysis

- *Panel data – longitudinal data or cross-sectional time series data*
  - Multiple units, each one with measurements at different time periods
  - More informative data, more variability, more efficiency
  - allow to explore more issues than cross-sectional or time-series data
  - Exploit both time variability and cross-section heterogeneity to analyze changes on an individual level.
  - Ex. EU Average energy consumption decreases by 3% from one year to another
    - Panel data allow to assess if consumption decreases by 3% in all countries, if it decreased by 6% in half population and 0% in the other half
  - Suitable to model and explain why individual units behave differently
  - Suitable to model and explain why behave differently at different time periods.

Panel data Analysis

- Types of Panel data

- Micro vs Macro Panel Data
  - Typically collected at micro-level (census survey, labor economics)
  - Increased pooling of individual time series of countries or industries (same industry across countries, different industries within same coutry)

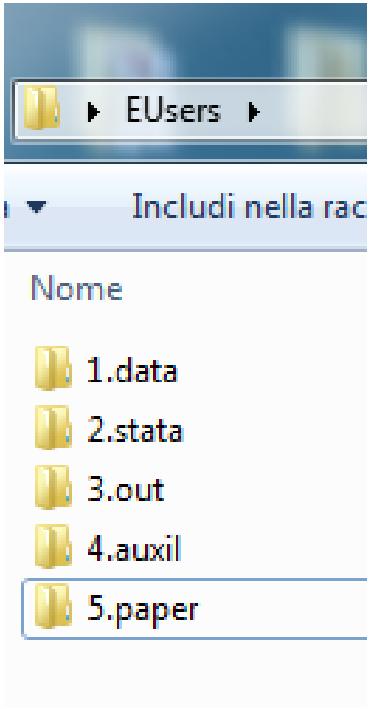
- Long vs Short Panel Data
  - Long: many time periods (large T) but few entities (small N) (T>N)
  - Short: many entities (large N) but few time periods (small T) (N>T)

- Balanced vs Unbalanced Panel Data
  - Balanced: Size sample equals *nT*. All entities have observations in all time periods
  - Unbalanced: Size sample smaller than *nT*. Some cells have missing observations.
  - Well-organized balanced sample is better. Unbalanced panel data entail some computation and estimation issues (e.g. interpolation) and most software packages can deal with unbalanced data as well

How to prepare panel data (using Stata)

- Variables to be downloaded from different sources, imported in Stata, cleaned, organized, and merged into a unique dataset according to the observation unit (country-sector-year)

- Need to organize material according to clear criteria



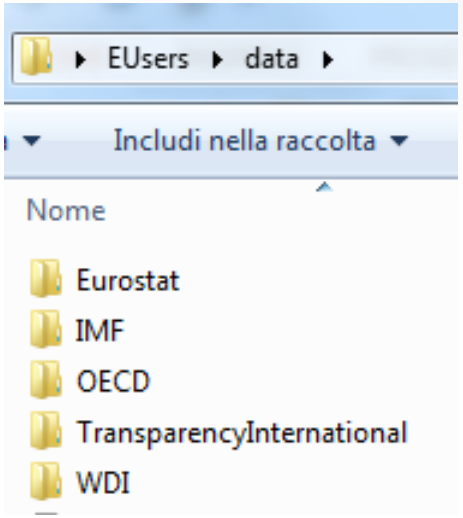
Main directory: name of the project (Eusers)

1. Folder with original data (data)
2. Folder with do files (stata): files containing command/script to organize and elaborate data, to create tables and graphs, to run regressions
3. Folder with transformed data set (out)
4. Folder with auxiliary materials: tables and graphs with excel etc. (auxil)
5. Folder with versions of the working paper: yyyymmdd_wp (paper)

How to prepare panel data: OECD ETCR

DATA

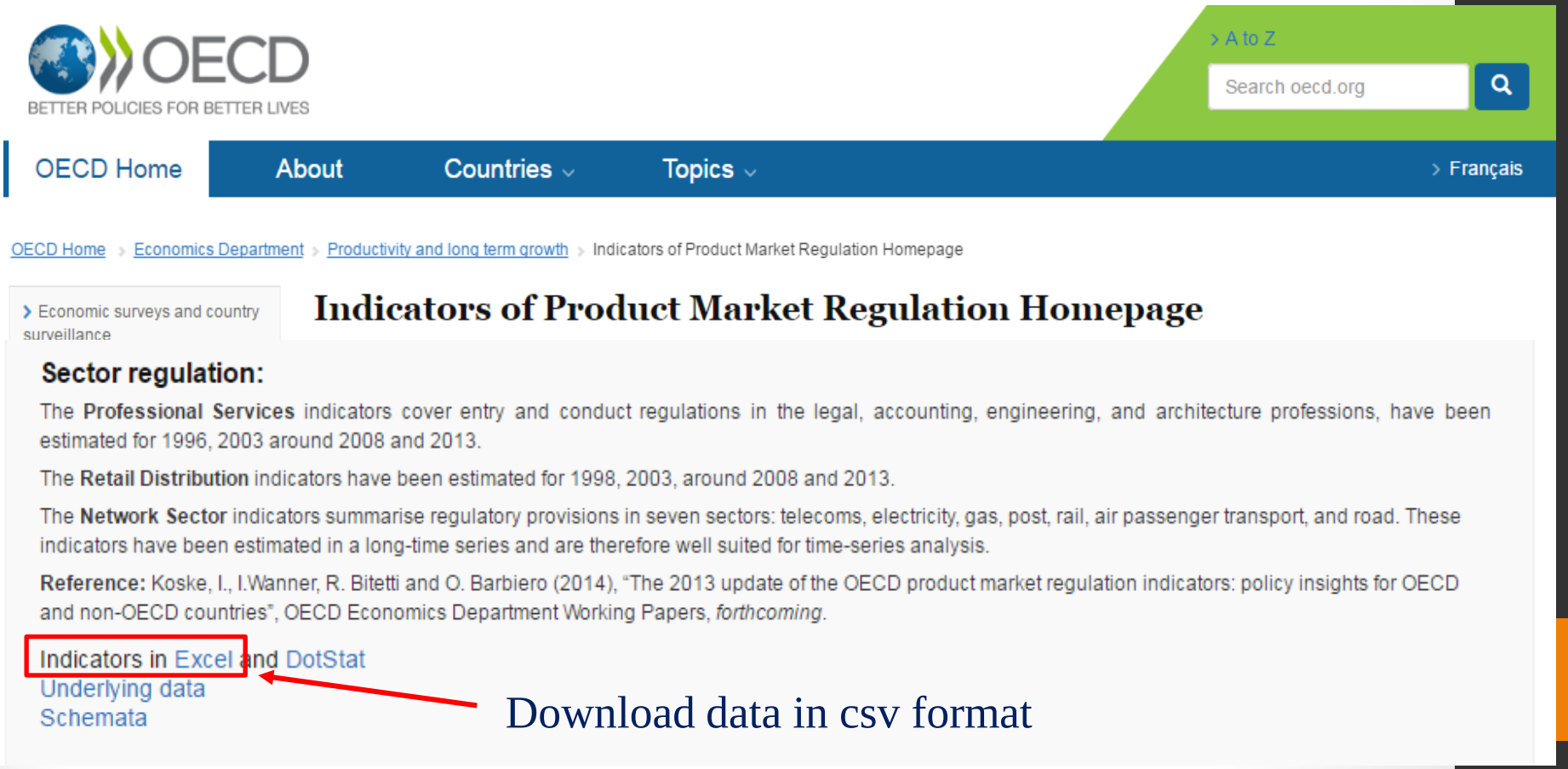
- Organized according to the main source
- Saved using the same format
- csv format: easy to import in Stata; clean format without special characters,
- Rename them according to clear criteria



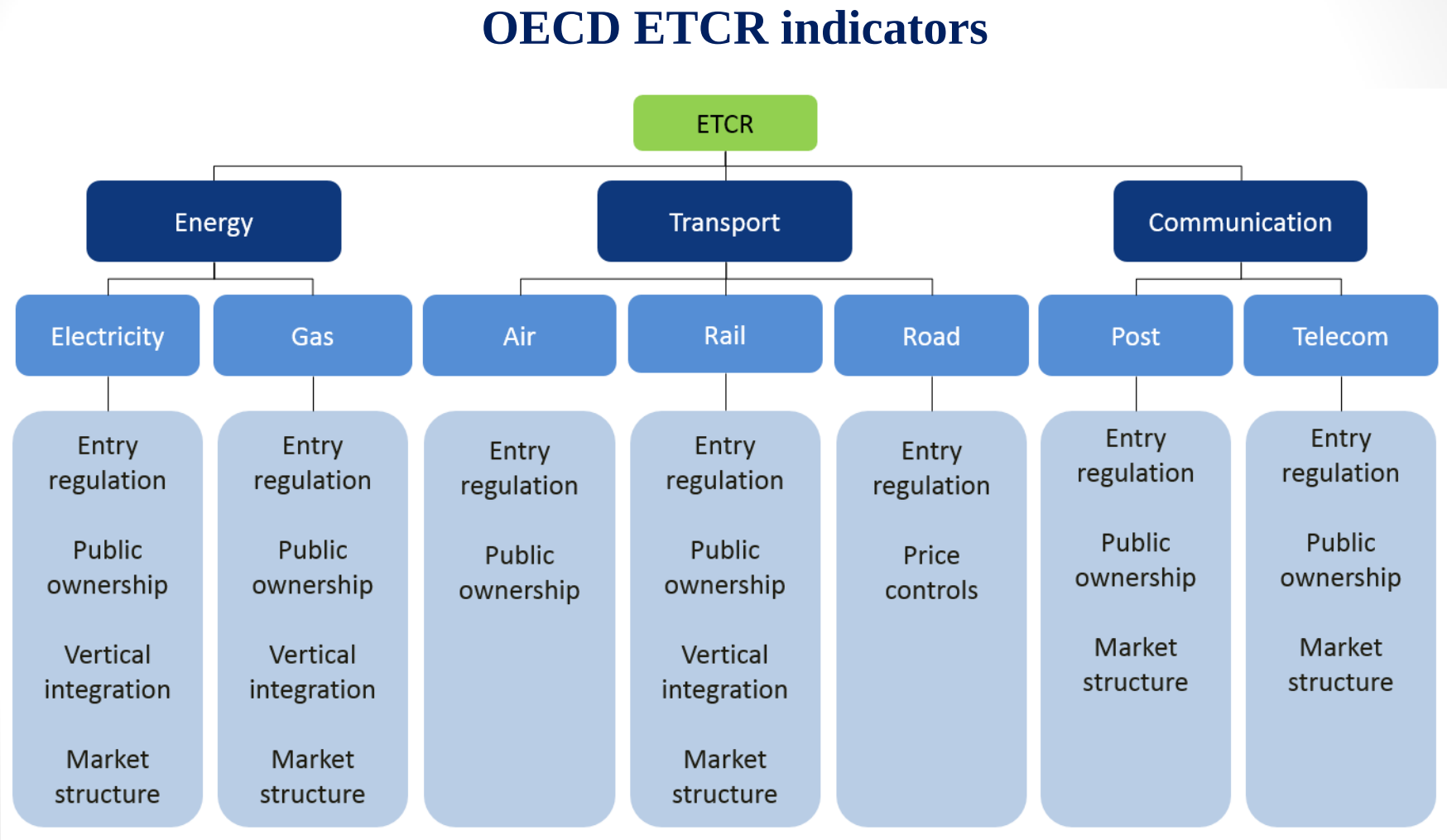
- Example: Focus on data on market reforms
- Monitored by the OECD ETCR (Energy, Transport, Communication, Regulation)
  - Several indicators to track intensity and progress of market and regulatory reforms over time at a sector and country level (Nicoletti et al., 1999).
  - Updated in 2003 (Conway et al. 2005), 2008 (Wölfl et al. 2009), 2013 (Koske et al. 2015).

How to prepare panel data: OECD ETCR

- OECD ETCR data available at:
- http://www.oecd.org/eco/growth/indicatorsofproductmarketregulationhomepage.htm



How to prepare panel data: OECD ETCR



How to prepare panel data: OECD ETCR

- **OECD ETCR indicators**:
- Built by computing information on regulation and policies
- Information collected through a questionnaire sent to governments in OECD and non-OECD countries.
- Responses integrated by publicly available data to create cross-country time series data of annual frequency
- Quantitative or qualitative information are standardized by assigning a numerical value to each response over a 0-6 scale
- A lower value reflects more intensive reform

How to prepare panel data: OECD ETCR

	Topic weight ai	Question weight bj	Coding of data					
Entry regulation	1/4							
How are the terms and conditions of third party access (TPA) to the electricity transmission grid determined?		1/3	Regulated TPA (0)		Negotiated TPA (3)			No TPA (6)
Is there a liberalised wholesale market for electricity (a wholesale pool)?		1/3	yes (0)					No (6)
What is the minimum consumption threshold that consumers must exceed in order to be able to choose their electricity supplier ?		1/3	No threshold (0)	<250 GW (1)	Between 250 and 500 GW (2)	Between 500 and 1000 GW (3)	> 1000 GW (4)	No consumer choice (6)
Public ownership	1/4							
What is the percentage of shares owned, either directly or indirectly, by the government in the largest firm in the sector?1		1	% of shares owned by government / 100 * 6					
Vertical Integration	1/4							
What is the degree of vertical separation between a certain segment of the electricity sector and other segments of the industry?1		1	ownership separation (0)		legal separation (3)	accounting separation (4.5)		no separation (6)
Market structure	1/4							
What is the market share of the largest company in the electricity industry?2		1	< 50% (0)		between 50 and 90% (3)			> 90% (6)
Country scores (0-6)			$\Sigma_i a_i \Sigma_j b_j \text{answer}_{ij}$					

1. Simple average over 4 segments: generation/import, transmission, distribution and supply.
2. Simple average over 2 segments: generation/import and supply.

How to prepare panel data: OECD ETCR

- Data Caveats
- Time, Statistical units and data organization change among ETCR versions
- Time: annual frequency1975-2007 (OECD 2008); 1975-2013 (OECD 2015)
- Sectors: Electricity, Gas, Airlines, Rail, Road, Telecommunication, Post
- Countries: 30 countries (OECD 2008, Balanced Sample); 46 countries (OECD 2015, unbalanced sample)
  - Increased by non-OECD and new OECD countries (China, India, Indonesia, South Africa, Latvia, Romania, Bulgaria, Croatia, etc.)
- Variables:
  - New indicators added (Market Structure in the electricity market)
  - Data reorganization (sub-indicators removed)
  - Computation changes among ETCR versions (Ownership was discrete, now it is a continuous variable)

How to prepare panel data (using Stata)

- **Data arrangement**
- This is how a dataset should like (OECD ETCR 2013 version)
  - Both cross-section (country-sector) and time (year) variables (long form)
  - Regulatory variables are numeric and encoded

		All sectors	Electricity				
Country	***year***	***Overall***	***Overall***	***Entry***	***Public Ownership***	***Vertical Integration***	***Market structure***
Austria	1998	3.91	5.02	6.00	3.06	6.00	.
Austria	1999	3.45	4.35	4.00	3.06	6.00	.
Austria	2000	3.13	3.85	4.00	3.06	4.50	.
Austria	2001	2.79	3.19	2.00	3.06	4.50	.
Austria	2002	2.63	2.52	0.00	3.06	4.50	.
Austria	2003	2.47	2.26	0.00	3.06	4.50	1.50
Austria	2004	2.08	2.17	0.00	3.06	4.13	1.50
Austria	2005	2.04	2.17	0.00	3.06	4.13	1.50
Austria	2006	1.89	1.75	0.00	3.06	3.94	0.00
Austria	2007	1.89	1.75	0.00	3.06	3.94	0.00
Austria	2008	1.84	1.75	0.00	3.06	3.94	0.00
Austria	2009	1.67	1.75	0.00	3.06	3.94	0.00
Austria	2010	1.66	1.75	0.00	3.06	3.94	0.00
Austria	2011	1.65	1.75	0.00	3.06	3.94	0.00

How to prepare panel data (using Stata)

- **Data arrangement**
- Data set is not always clean & in a long form (OECD ETCR 2008)

4	ENTRY REGULATION				
5			Data		
6	Year	Item	Australia	Austria	Belgium
37	2005	How are the terms and conditions of third party access (TPA) to the electricity transmission grid determined?	Regulated TPA	Regulated TPA	Regulated TPA
38	2006	How are the terms and conditions of third party access (TPA) to the electricity transmission grid determined?	Regulated TPA	Regulated TPA	Regulated TPA
39	2007 [1]	How are the terms and conditions of third party access (TPA) to the electricity transmission grid determined?	Regulated TPA	Regulated TPA	Regulated TPA

- Data set in a wide form (includes either individual or time variable)
- We have time variable but not cross-section variable
- Indicator are string variable (Regulated TPA) instead of numeric variable

- Data need to be cleaned and encoded (converted into numeric values)
- Data need to be rearranged from wide to long form

How to prepare panel data (using Stata)

- Import, clean and organize data using Stata
- DO files need to be understood by other people and after time
- DO files need to be organized in a clear way

```
******************
******************
* Class tutorial
******************
******************

* ROADMAP
* 1. select my PC directory
*    tell stata where to take original data set,
*    where to find do files with commands
*    and where to save new data set)
* 2. import original dataset into stata
* 3. clean data and transform string variables into numeric variables
* 4. rearrange dataset from wide form to long form
* 5. merge data and create final dataset
```

How to prepare panel data (using Stata)

- Import, clean and organize data using Stata
- To be done with Stata (on PPT only screenshot)

```
*******************************************************************************
* 3. clean data and transform string variables into numeric variables
*******************************************************************************
* summarize dataset and look at dataset in stata
su
```

Command to summarize all the variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Year	0				
Item	0				
Australia	0				
Austria	0				
Belgium	0				
Canada	0				
zech_Repu~c	0				
Denmark	0				
Finland	0				
France	0				

All variables as strings:

No values

How to prepare panel data (using Stata)

- **Clean and Destring Year Variable (to be done for all string variables)**
- destring Year, replace (stata command)
- Year contains nonnumeric characters; no replace (stata result)

- Year is a string variable: convert it in a numeric variable
  - Drop empty rows
  - Drop non-numeric characters
  - now Year is a numeric variable and we can destring it
- destring Year, replace (stata command)
- Year has all characters numeric; replaced (stata result)

```
. su Year

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        Year |        198        1991    9.546041      1975       2007

.
```

How to prepare panel data (using Stata)

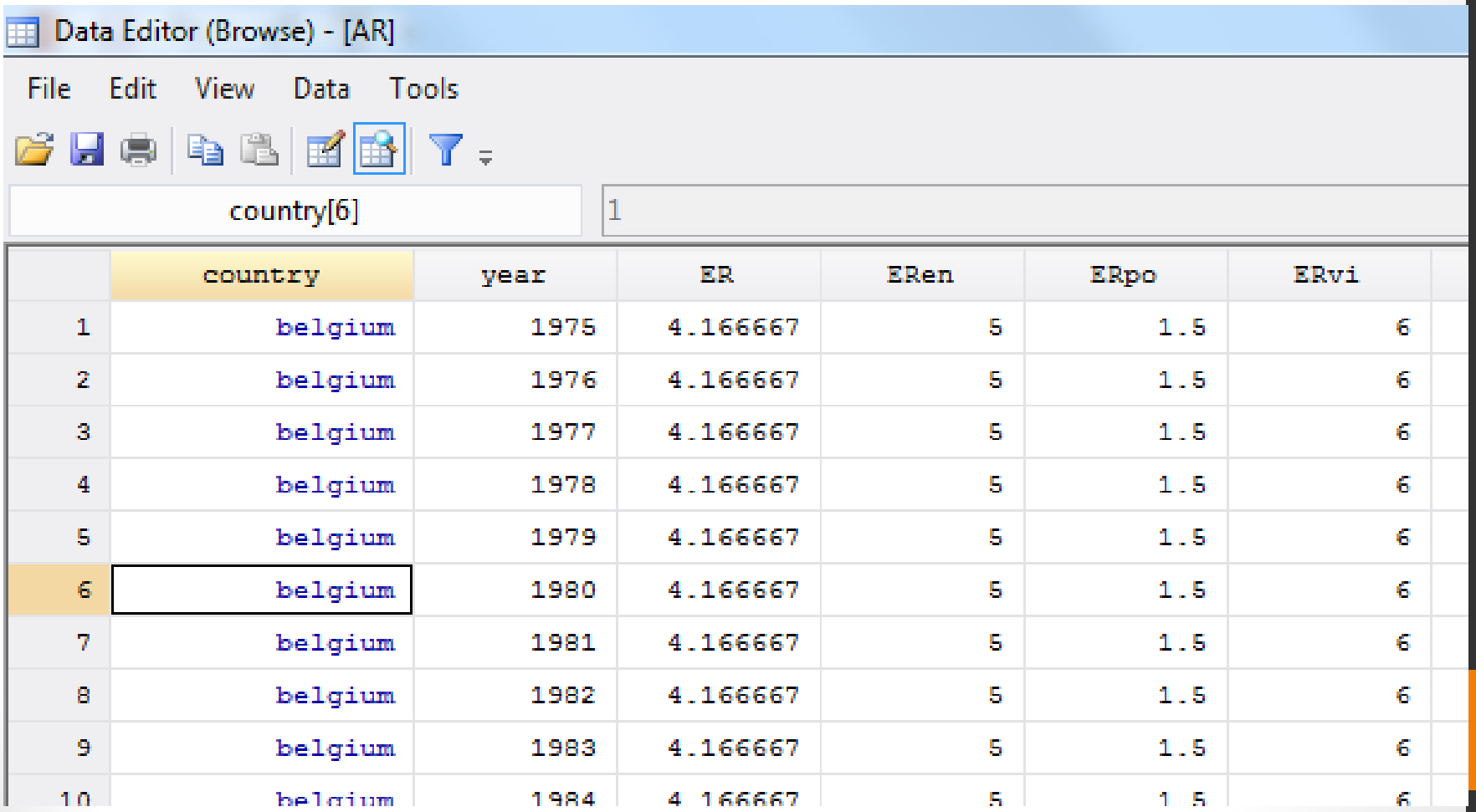
- **Reshape data set from wide to long**
- * now we have one variable for each country (wide)
- * we want one variable containing all the countries (long)
- * to reshape from wide to long, country variables must have the same prefix
  - rename Belgium ER`var'1 (stata command)
  - rename Denmark ER`var'2 (stata command)
  - reshape long ER`var', i(year) j(country)
  - Loop script to repeat these commands for each indicator `var'= po, en, vi
  - Merge separates files (each one with one indicator) into a unique data set

```
. reshape long ER`var', i(year) j(country)
(note: j = 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15)

Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                       33   ->      495
Number of variables                  17   ->        4
j variable (15 values)                    ->   country
xij variables:
                     ER1 ER2 ... ER15     ->   ER
-----------------------------------------------------------------------------
```

How to prepare panel data (using Stata)

- **How new data set looks like**



	country	year	ER	ERen	ERpo	ERvi
1	belgium	1975	4.166667	5	1.5	6
2	belgium	1976	4.166667	5	1.5	6
3	belgium	1977	4.166667	5	1.5	6
4	belgium	1978	4.166667	5	1.5	6
5	belgium	1979	4.166667	5	1.5	6
6	belgium	1980	4.166667	5	1.5	6
7	belgium	1981	4.166667	5	1.5	6
8	belgium	1982	4.166667	5	1.5	6
9	belgium	1983	4.166667	5	1.5	6
10	belgium	1984	4.166667	5	1.5	6

Choose a proper panel data model

- Panel data model allow to examine group effects & time effects
- What is the most appropriate model?

- **POOLED OLS MODEL**
- OLS method for estimating parameters in a linear regression model, min. the square of the differences between observed data and the linear approximation of the data
- Errors assumed to be independent and identically distributed → efficient consistent estimates: ordinary least squares (OLS) produces best linear unbiased estimator (BLUE)

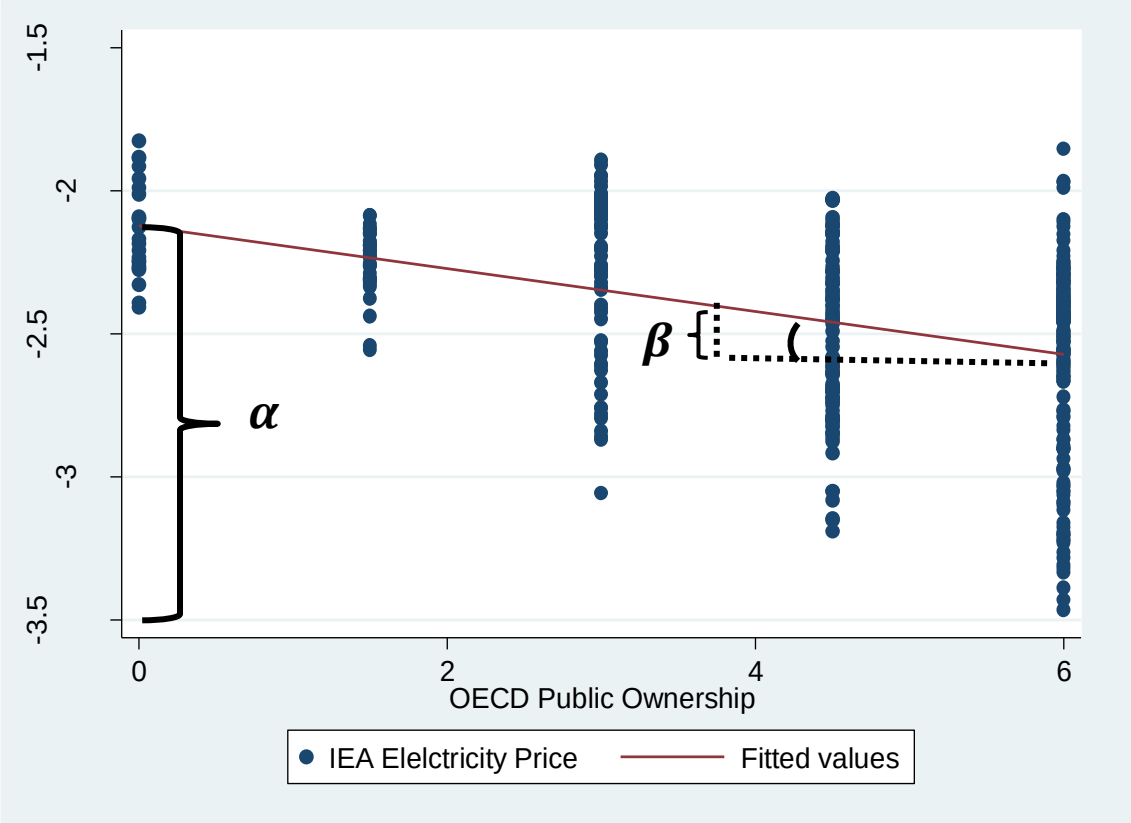
$$y_{it} = \alpha + X'_{it}\beta + \varepsilon_{it}$$

- α and β (intercept and slope) constant for all *i* and *t*
- Naive, ignores differences in time and space

Choose a proper panel data model

twoway scatter lEAprinet_kw ERpo || lfit lEAprinet_kw Erpo

Pooled OLS model: Same intercept and same slope for all countries (all groups equivalent)



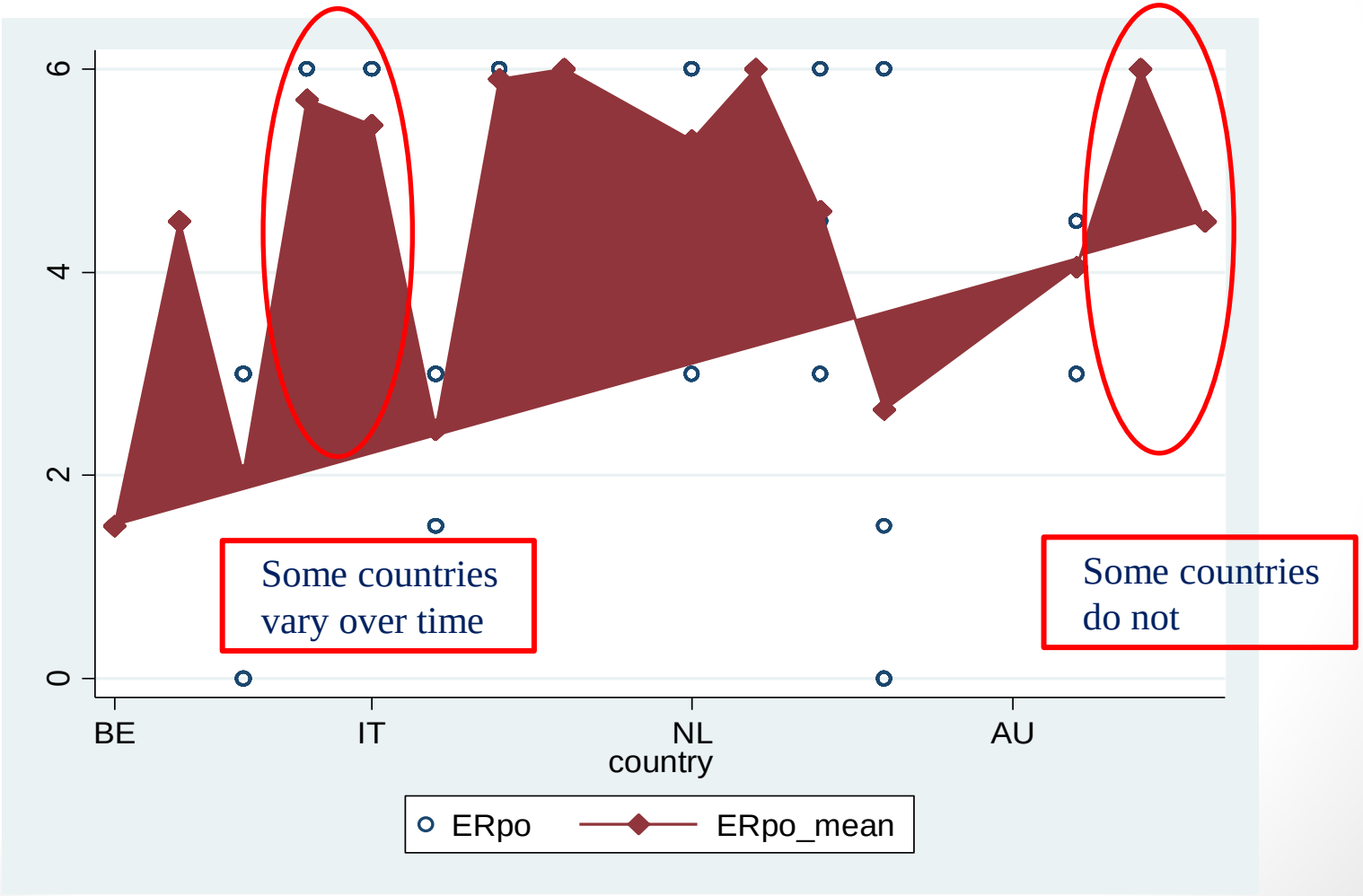
Choose a proper panel data model

- **POOLED OLS MODEL**
- ok If no group or time effect;
- If group or time effect ($u_i \neq 0$) → biased estimates
- Same units repeatedly observed → is it appropriate to assume that different observations are independent?
- If we do not control for that, Neglected individual (mean) effects may be captured by the error term
- error terms may not be uncorrelated and uniform
  - variance which vary across groups or are related with one another **(Heteroskedasticity)**
  - Error terms present correlation between values (**Autocorrelation)**

When we have heterogeneity among groups, we need panel data model able to examine group effects &/or time effects

Fixed Effect model

- We observe heterogeneity across countries: unobserved variables that do not change over time
- Blue dot: ER_po values for each country; Red dot: average ER_po for each country



Fixed Effect model

- Groups may have intrinsic and **time-invariant** characteristics which **specific to each group** (female vs male; cultural differences)
  - Individual who experience an event in the past are more likely to experience it in the future (due to change in preferences)
  - Individual differ in unobserved characteristics which influence the probabilty of experiencing the event
- Characteristics may be correlated with the regressor
- Characteristics may affect the dependent variable: need to control for them
  - Firm output depends on input (labor, assets, material) and on managerial quality
  - Managerial quality negatively correlated with inputs (higher quality brings to more efficient use of inputs)
- If omitted, group-specific characteristics captured by the error term → no more IID and biased estimator

Fixed Effect model

- **FIXED EFFECT** allow to deal with *individual effects (heterogeneity)* that may or may not be observed and influence the dependent variable
- FE model controls for all time-invariant differences between groups → estimator not biased because of omitted time-invariant characteristics
- FE designed to study the causes of changes within a group. A time-invariant characteristic cannot cause such a change, being constant for each group
- Fixed-effects work well when within-cluster variation is high or for high changing variables over time.

Fixed Effect model

- FE estimated by least squares dummy variable (LSDV) regression (OLS with a set of dummies) and within effect estimation methods.
- Inclusion of dummy variables in the model allow to demean our longitudinal data → demean variables and run regression on the demeaned data to calculate the coefficient → analyze time variation within a group
- With FE models, individual characteristics are incorporated in **the intercept** → they will vary across groups
- FE examines if intercepts vary across group or time period

Fixed Effect model

$$y_{it} = \alpha_i + X'_{it}\beta + \varepsilon_{it}$$

- α_i, i=1,…, N → N fixed unknown parameters to be estimated
- Each group has its own intercept, which captures the effects that are peculiar to each group and that do not vary over time
- β is the coefficient for the independent variable: for a given country, as X varies across time by one unit, y increases or decreases by β units
- β estimation for all: a change in x has same effect (one period to another or one individual from another)

$$y_{it} = \alpha + X'_{it}\beta + \gamma_2 C_2 + \gamma_3 C_3 + \ldots + \gamma_n C_n + \varepsilon_{it}$$

- $\alpha_i = \alpha + \gamma_i$
- C_i (i=1….n) is the dummy variable for each country C. Since they are binary (dummies) you have n-1 entities included in the model (otherwise collinearity).
- γ_iIs the coefficient for the binary variables.

Fixed Effect model

- Have many regressors as dummy variables may be anattractive (think of panel with thousands of groups)
- The same estimator for β can be obtained when the regression is performed in deviations from individual means
- We eliminate the individual effects α_i by transforming the data and demeaning them

$$y_{it} - \overline{y}_i = \left(X'_{it} - \overline{X}_i\right)\beta + (\varepsilon_{it} - \overline{\varepsilon_i})$$

- Regression model in deviation from individual means and does not include the individual effects α_i → within estimator
- FE concentrates on differences "WITHIN" individuals (exploit time variability: to what extent y_{it} differs from $\overline{y}_i$
- It does not explain why $\overline{y}_i$ is different from $\overline{y}_j$
- β identified only through within dimension of the data

Fixed Effect model

- Two ways to estimate the same regression: demeaning vs intercepts

xtreg lprice ERpo i.country, fe

lEAprinet_kw	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ERpo	-.0986983	.0101557	-9.72	0.000	-.1186617	-.0787349

regress lprice ERpo i.country → it shows intercepts for different countries

lEAprinet_kw	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ERpo	-.0986983	.0101557	-9.72	0.000	-.1186617	-.0787349
country						
denmark	-.114456	.0677974	-1.69	0.092	-.2477271	.018815
germany	.025113	.0603319	0.42	0.677	-.0934829	.143709
greece	.0118451	.0749705	0.16	0.875	-.1355263	.1592165
italy	.2280151	.072273	3.15	0.002	.0859463	.3700838
spain	.0626802	.0612996	1.02	0.307	-.0578179	.1831784
france	.1974961	.0749061	2.64	0.009	.0502513	.3447409
ireland	.32869	.0755164	4.35	0.000	.1802456	.4771344
luxembourg	.2025788	.0714387	2.84	0.005	.06215	.3430075
netherlands	.2234786	.0755164	2.96	0.003	.0750342	.371923
portugal	.2311469	.0678625	3.41	0.001	.0977479	.3645459
great britain	-.0431905	.0612418	-0.71	0.481	-.163575	.077194
finland	-.2882814	.0654585	-4.40	0.000	-.4169548	-.1596079
sweden	-.4636473	.0805415	-5.76	0.000	-.6219697	-.305325
austria	.1766497	.0673974	2.62	0.009	.044165	.3091344

Fixed Effect model n entity-specific intercepts (using xtreg , FE stata option)

xtreg lprice ERpo i.country, fe

R-sq: variance of Y explained by X

Within: "How much variance within the panel units"

Between: "How much variance between seperate panel units"

overall "weighted average of these two"

```
Fixed-effects (within) regression               Number of obs      =       429
Group variable: country                         Number of groups   =        15

R-sq:                                           Obs per group:
     within  = 0.1861                                         min =        20
     between = 0.2305                                         avg =      28.6
     overall = 0.1831                                         max =        30

                                                F(1,413)           =     94.45
corr(u_i, Xb)  = -0.2183                        Prob > F           =    0.0000
```

lEAprinet_kw	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ERpo	-.0986983	.0101557	-9.72	0.000	-.1186617	-.0787349
_cons	-2.016872	.0464356	-43.43	0.000	-2.108152	-1.925592
sigma_u	.21580801					
sigma_e	.21691521					
rho	.49744134	(fraction of variance due to u_i)				

F test that all u i=0: F(14, 413) = 24.25 Prob > F = 0.0000

Unbalanced sample

The errors ui correlated with the regressors in FE model

Coefficients of the regressor (Slope): *Y* changes when *X* increases by one unit.

'rho': 49% of the variance is due to differences across panels.

$$rho = \frac{(sigma\_u)^2}{(sigma\_u)^2 + (sigma\_e)^2}$$

sigma_u = sd of residuals within groups u_i

sigma_e = sd of residuals (overall error term) e_i

deviation of the distribution from the mean

t-values test coeff ≠0. reject if t-value > |1.96| (95% confidence). X has a significant influence on Y

P-values test coeff ≠0. Reject if < 0.05 (95% confidence). X has a significant influence on Y

test (F) to see whether coeff different than zero. Ok If < 0.05

Fixed Effect model

Variable	ols	ols_dum
ERpo	-.07492021***	-.0986983***
_Icountry_2		-.11445602
_Icountry_3		.02511303
_Icountry_4		.0118451
_Icountry_5		.22801505**
_Icountry_6		.06268022
_Icountry_7		.19749612**
_Icountry_8		.32868997***
_Icountry_10		.20257876**
_Icountry_11		.22347857**
_Icountry_12		.2311469***
_Icountry_13		-.0431905
_Icountry_16		-.28828135***
_Icountry_17		-.46364735***
_Icountry_18		.17664973**
_cons	-2.1227926***	-2.082159***
N	429	429

legend: * p<0.05; ** p<0.01; *** p<0.001

Same Slope for all the countries

Same Slope for all the countries, but different once we add new explanatory variables

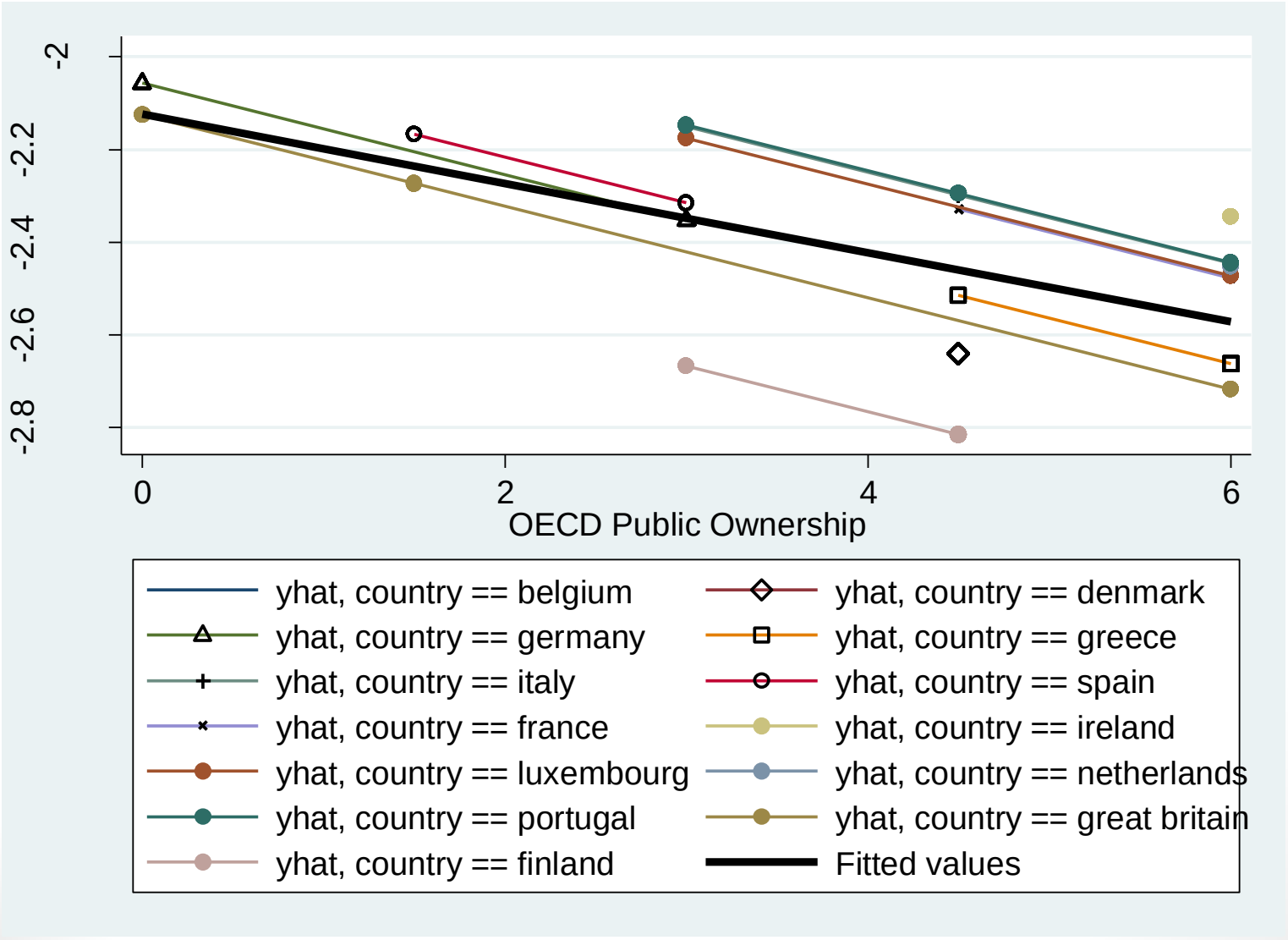
Same Intercept for all the countries

Different intercept for each country = _cons+_Icontry_i

Constant "alone" refers to the intercept of the omitted country 1

Fixed Effect model

- Different intercepts same slope



Fixed Effect model

- **Conditions for efficiency**
- All x_{it} are independent of all ε_{it}
- x_{it} uncorrelated with ε_{it}
- $\overline{x}_i$ has no correlation with the error term
- x_i strictly exogenous and does not depend on values of the error terms

- characteristics are **unique to the group and not correlated with other group characteristics**.
- If characteristics are common to groups → the group's error term and the constant (captures group characteristics) are correlated with others.
- If the error terms are correlated, then FE is no suitable

Random Effect model

- FE exploits the within dimension of the data (differences within individuals, deviation from individual mean)
- RE exploits the between dimension of the data (differences between individuals → regression of individual averages of y on individual averages of x
- Difference between FE and RE lies in the role of dummies
- FE: individual and time-invariant effects → captured in the intercept (one for each group)
- RE: individual effect is an error component → intercepts may be different but due to an error component
- The intercept and slopes of regressors are the same across individuals. Difference among individuals (or time periods) lies in their individual specific errors, not in their intercepts.

Random Effect model

$$y_{it} = \alpha + X'_{it}\beta + u_i + \varepsilon_{it},$$

- Error term consists in two components
  - u_i: time-invariant and individual-specific and random component.
  - ε_{it}: component uncorrelated over time
  - All correlation of the error terms over time is attributed to the individual effects u_i
  - u_i and ε_{it} are mutually independent and independent of x_{js}
- Group effects assumed to be random variables with:
  - mean value common to all groups α;
  - intercept value for group i: $\alpha_i = \alpha + u_i$,

- Distinction between FE and RE is whether unobserved individual effect embodies elements that are correlated with the regressor
  - FE assumes time-invariant effects correlated the regressor
  - RE assume individual effects to be random and uncorrelated with regressor

Random Effect model

- RE allow to include time invariant variables in the model that in FE are absorbed by the intercept.
- RE assume that the entity's error term is not correlated with the predictors which allows for time-invariant variables to play a role as explanatory variables.
- In random-effects you need to specify those individual characteristics that may or may not influence the predictor variables.
- These variables may not be available therefore leading to omitted variable bias in the model.

Random Effect model

Differences across units uncorrelated with the regressors

test (F) to see whether coeff different than zero. Ok If < 0.05

xtreg lEAprinet_kw ERpo, re

```
corr(u_i, X)    = 0 (assumed)                 Wald chi2(1)      =      97.99
                                              Prob > chi2       =     0.0000
```

lEAprinet_kw	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ERpo	-.0969144	.0097903	-9.90	0.000	-.116103	-.0777258
_cons	-2.037612	.071535	-28.48	0.000	-2.177818	-1.897406
sigma_u	.21624464					
sigma_e	.21691521					
rho	.49845192	(fraction of variance due to u_i)				

Coeff similar to FE, but interpretation differs, as include both the within-group and between-group effects. average effect of X over Y when X changes across time and between countries by one unit.

xtreg lEAprinet_kw ERpo, fe

```
corr(u_i, Xb)   = -0.2183                     Prob > F          =     0.0000
```

lEAprinet_kw	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ERpo	-.0986983	.0101557	-9.72	0.000	-.1186617	-.0787349
_cons	-2.016872	.0464356	-43.43	0.000	-2.108152	-1.925592
sigma_u	.21580801					
sigma_e	.21691521					
rho	.49744134	(fraction of variance due to u_i)				

The errors ui correlated with the regressors

Choose a proper panel data model

- Should we opt for FIXED or RANDOM EFFECT?
- It may depend on whether or not the individuals can be viewed as a random sample from a large population
- Our sample is randomly chosen?
- Is representative of the whole population?
- Can we extend results to the whole population
- Our sample shows some specific characteristics which are peculiar to the group and cannot extended to the world population?
- Do results strictly depend on the type of individual analyzed?

- Intuitively, if sample is random go for RE, if sample has fixed-specific characteristics go for FE

Choose a proper panel data model

- Group effect is correlated with the regressor (part of the intercept) → then use FE
- is it a random component not correlated with the regressor (part of the error term)? → use RE
  - for small T and large N, RE estimators more efficient than FE
- **Hausman Test**
- Allow to decide between FE or RE
- It basically tests whether the unique errors (ui) are correlated with the regressors, the null hypothesis is they are not
- null hypothesis is that the preferred model is random effects vs. the alternative the fixed effects.
- If $p < 0.05$ (i.e. significant), we reject the null and use fixed effects.

Choose a proper panel data model

- What is the impact of ownership on electricity prices? (Florio and Fiorio, *EnEc* 2013)

$$P_{it} = R'_{it}\beta + X'_{it} + \delta_t + \xi_i + \varepsilon_{it}$$

- R_{it}: Vector of regulatory variables scoring the intensity of regulation in electricity industry
- X_{it}: set of control variables likely to affect prices
  - Nominal GDP per capita (Source: WB).
  - Amount of electricity produced by using combustible fuel (Source: IEA)
  - Amount of electricity import (Source: IEA)
  - Amount of electricity residential consumtion (GWh) (Source: IEA)
  - Consumer price index (Source: WB).

- δ_t: Year time fixed-effects to capture the common trend across the EU and common cyclical components in prices, market reforms and other macroeconomic variables

- ξ_i: country fixed effects to control for country-specific characteristics and time-invariant differences among countries

- Prices and control variables are log-transformed to interpret coefficients as elasticities

Choose a proper panel data model

xtreg `dip' `indio var',fe
estimates store fe

xtreg `dip' `indio var‘, re
estimates store re

hausman fe re

	Coefficients (b) fe	(B) re	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
lEAscmbf	-.095354	-.0480685	-.0472855	.0069841
lEAimports	-.006512	-.0050933	-.0014187	.
lEArescons	-.3028592	-.0436712	-.259188	.059543
lMWgdppc	.2099768	.1330524	.0769244	.0139708
lMWcpi	.3699643	.2846999	.0852644	.0187848
ERpo_d	-.0823168	-.0845201	.0022032	.
ERvi_d	.0522291	.0056029	.0466263	.0035151
ERen2_d	.0062585	.0315378	-.0252792	.
ERen3_d	-.0596546	-.0378938	-.0217608	.

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(9) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 44.71
Prob>chi2 = 0.0000
(V_b-V_B is not positive definite)

P < 0.000 → we reject the null and opt for FE

VARIABLES	(1) FE	(2) FE	(3) FE
lEAscmbf	0.026	0.011	0.011
	(0.161)	(0.567)	(0.555)
lEAimports	-0.014***	-0.012***	-0.012***
	(0.000)	(0.000)	(0.000)
lEArescons	-0.263***	-0.261***	-0.255***
	(0.000)	(0.000)	(0.000)
lMWgdppc	0.557***	0.515***	0.504***
	(0.000)	(0.000)	(0.000)
ER	-0.032***		
	(0.000)		
ERpo_d		-0.124***	-0.131***
		(0.000)	(0.000)
ERvi_d		0.035	0.039
		(0.212)	(0.168)
ERen_d		-0.083***	
		(0.004)	
ERen1_d (conditions of TPA)			-0.039*
			(0.085)
ERen2_d (liberalised wholesale market))			0.001
			(0.953)
ERen3_d (consumers can switch providers)			-0.081***
			(0.001)
Constant	2.168***	2.025***	1.956***
	(0.000)	(0.000)	(0.000)
Year Dummy	YES	YES	YES
Observations	417	417	417
R-squared	0.835	0.839	0.842
Number of country	15	15	15
pval in parentheses	*** p<0.01, ** p<0.05, * p<0.1		