

# Model Data Panel Statis

oleh

Bambang Juanda

**Departemen Ilmu Ekonomi, FEM IPB**

<https://bambangjuanda.com/>

Ilustrasi Data Panel dari <http://dss.princeton.edu/training/Panel101.dta> (Panel Data Analysis: Fixed & Random Effects using Stata(v.4.2) by Oscar Torres-Reyna, 2007)

	country	year	y	y_bin	x1	x2	x3	opinion
1	C	1990	-1.292e+09	0	1.312561	-1.293136	.204083	Agree
2	A	1990	1.343e+09	1	.2779036	-1.107956	.2825536	Str agree
3	E	1990	1.343e+09	1	.4528672	1.728403	.5970579	Str disag
4	F	1990	1.343e+09	1	-.5675749	-.346671	1.258419	Str agree
5	B	1990	-5.935e+09	0	-.08185	1.42512	.0234281	Agree
6	D	1990	1.883e+09	1	-.3139127	1.736656	.646637	Disag
7	G	1990	1.343e+09	1	.9448817	-1.515015	1.452657	Str disag
8	E	1991	2.296e+09	1	.4190403	1.70684	.7931353	Str agree
9	B	1991	-7.116e+08	0	.10616	1.649602	.2603625	Str agree
10	C	1991	-3.416e+09	0	1.177484	-1.344218	.2839719	Str agree
11	F	1991	3.560e+09	1	.1597458	-.4641182	.326653	Str disag
12	A	1991	-1.900e+09	0	.3206847	-.94872	.4925385	Disag
13	D	1991	6.038e+09	1	.360091	2.131864	1.099942	Disag
14	G	1991	-1.519e+09	0	1.098728	-1.461472	1.439645	Agree
15	D	1992	10244189	1	.0518877	1.681677	.9697682	Str agree
16	E	1992	1.738e+09	1	.3852135	1.685277	.9892128	Agree
17	A	1992	-11234363	0	.3634657	-.789484	.7025234	Disag
18	B	1992	-1.933e+09	0	.3537852	1.593719	-.2343988	Agree
19	C	1992	-3.558e+08	0	1.256408	-1.259951	.3733927	Agree
20	F	1992	3.192e+09	1	.8870664	-.5815655	-.6051133	Agree

## Penggabungan Data *Cross-Section* dan *Time Series* (modelnya?)

Data untuk beberapa unit individu (misalnya perusahaan, daerah atau negara) selama periode waktu tertentu (misalnya beberapa bulan atau tahun). Penggabungan Data *Cross-Section* dan *Time Series* ini biasa dikenal dengan *panel data* atau *pooled data*, *longitudinal data*, *event history analysis*, ataupun *cohort analysis*. Misalnya kita punya data total pendapatan (X) dan total konsumsi (Y) untuk N daerah kabupaten selama T tahun. Misalkan kita mengkaji model:

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it}, \text{ untuk } i = 1, 2, \dots, N; t = 1, 2, \dots, T \quad (5.6)$$

Jika asumsi klasik  $\varepsilon_{it}$  dipenuhi dan N,T besar maka **kita dapat menduga model regresi secara terpisah**, misalnya untuk tiap waktu t dengan melibatkan N pengamatan, modelnya untuk **t=1** (ada T persamaan ) adalah sebagai berikut:

$$Y_{i1} = \alpha_1 + \beta_1 X_{i1} + \varepsilon_{i1}, \text{ untuk } i = 1, 2, \dots, N;$$

Kita juga dapat menduga N model regresi dengan melibatkan T pengamatan.

**Jika  $\alpha$  dan  $\beta$  keduanya konstan sepanjang waktu dan juga untuk antar daerah,** maka dugaan yang lebih efisien adalah **model (5.6) dengan data besar (pooled)** sebanyak NT pengamatan, dan akan ada NT-2 derajat bebas sisaan.

# OLS Regression

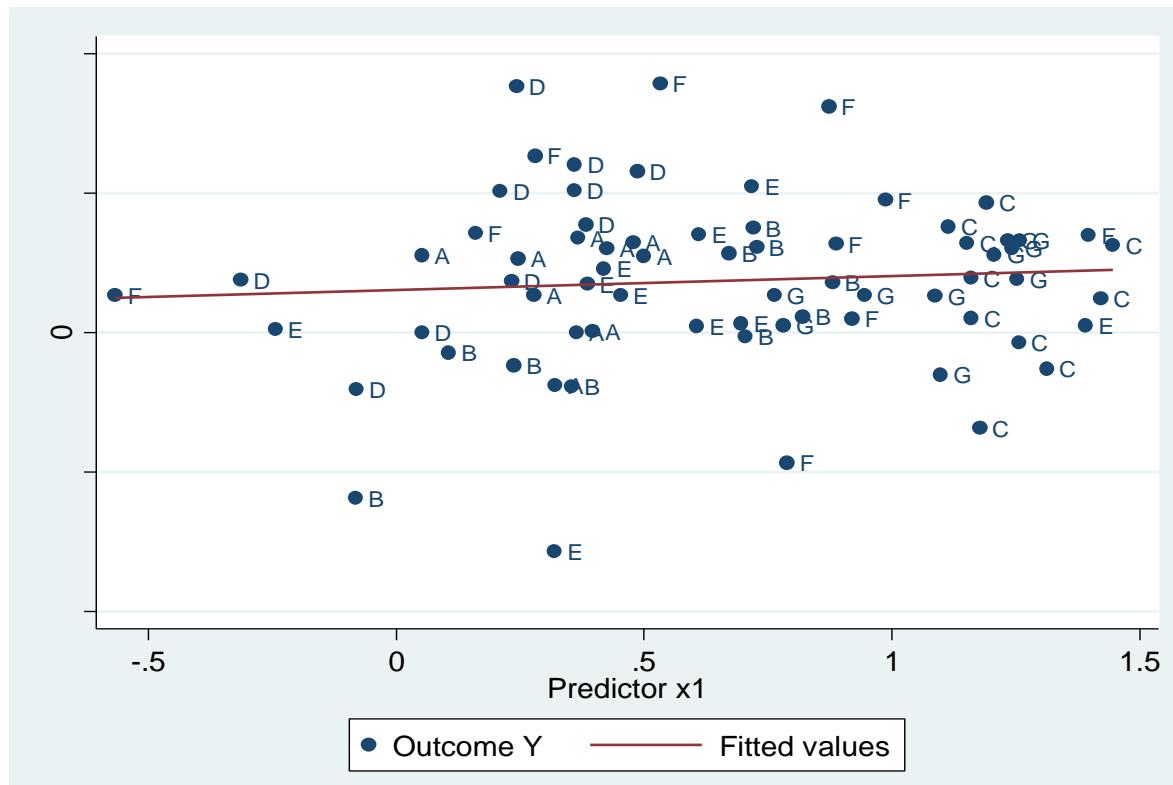
Source	SS	df	MS	Number of obs	=	70
Model	3.7039e+18	1	3.7039e+18	F(1, 68)	=	0.40
Residual	6.2359e+20	68	9.1705e+18	Prob > F	=	0.5272
				R-squared	=	0.0059
				Adj R-squared	=	-0.0087
Total	6.2729e+20	69	9.0912e+18	Root MSE	=	3.0e+09

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	4.95e+08	7.79e+08	0.64	0.527	-1.06e+09 2.05e+09
_cons	1.52e+09	6.21e+08	2.45	0.017	2.85e+08 2.76e+09

. regress y x1

. twoway scatter y x1,  
mlabel(country) || lfit y x1,  
clstyle(p2)



## Kentungan Menggunakan Data Panel

- **Jumlah observasi menjadi lebih banyak.** Jika menggunakan satu entitas, misalnya negara maka tidak dapat dilihat **perbedaan antar wilayah**. Jika diamati beberapa entitas, misalnya 34 provinsi hanya pada satu waktu maka tidak dapat dilihat **perkembangan tiap wilayah (34 provinsi)** tsb.
- Memberikan **data yang lebih informatif**, lebih bervariasi, mengurangi kolinearitas antarpeubah (dapat menggunakan *first difference*), memperbesar derajat kebebasan, dan lebih efisien.
- Dapat mengontrol *unobserved heterogeneity*
- Dapat dipelajari suatu bentuk perubahan yang dinamis.
- Dapat mendeteksi dan **mengukur efek suatu variabel** pada variabel lainnya dengan lebih baik (misalnya utk tiap entitas dan tiap waktu)
- Dapat digunakan untuk mempelajari model prilaku (*behavioral model*) yang lebih kompleks

## Masalah (Asumsi) Model Data Panel

Kerangka Umum Data Panel Satu Peubah Bebas

i	t	Y <sub>it</sub>	X <sub>it</sub>
1	1	Y <sub>11</sub>	X <sub>11</sub>
.	.	.	.
1	T	Y <sub>1T</sub>	X <sub>1T</sub>
.	.	.	.
.	.	.	.
N	1	Y <sub>N1</sub>	X <sub>N1</sub>
.	.	.	.
N	T	Y <sub>NT</sub>	X <sub>NT</sub>

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it}, \text{ untuk } i = 1, 2, \dots, N; t = 1, 2, \dots, T \quad (5.6)$$

Masalah model (5.6) adalah bahwa **asumsi  $\alpha$  dan  $\beta$  konstan mungkin tidak realistik**. Misalnya apakah *Marjinal Propensity to Consume* ( $\beta$ ) antara Kabupaten Bogor dengan Kabupaten Sorong sama? Begitu juga tentang intersep  $\alpha$ . Untuk mengatasi masalah ini dapat menggunakan FEM dan REM, **selain model Dummy interaksi**.

Sumber: Pornchaiwiseskul, 2004

Analisis data panel memungkinkan kita mengontrol faktor penyebab heterogenitas **antar individu** yang tidak dapat kita amati atau ukur; atau faktor yang merubah **antar waktu** tetapi bukan antar individu (entitas). Yang jarang diungkapkan dalam bukuteks adalah **interaksi antara keduanya**, seperti yang dibahas dalam masalah *Dummy Variable* di **Bab 5** <https://www.youtube.com/watch?v=xJNibcIOaRo&list=PL5cvP3A0cz7b1Ng6Zhkc0KfRU6vHyke9u&index=6>

Dengan data panel kita dapat memasukkan variabel di berbagai tingkatan analisis (misal Provinsi, Kab/Kota, K/L, siswa, sekolah) yang sesuai untuk pemodelan bertingkat atau hierarki.

Beberapa kelemahan adalah masalah pengumpulan data (misal *sampling desain*, cakupan), *non-respons* dalam kasus panel-mikro atau ketergantungan lintas wilayah dalam kasus panel-makro (misal korelasi antar wilayah)

**Note:** Kelebihan dan kekurangan analisis data panel dapat lihat di Baltagi, *Econometric Analysis of Panel Data* (chapter 1).

# Pendekatan dalam Model Regresi Data Panel

- Estimasi Model tergantung asumsi intersep, slope dan sisaan  $u_{it}$
- Terdapat beberapa kemungkinan :
  - Intersep dan slope adalah konstan menurut waktu dan individu sedangkan sisaan berbeda antar waktu dan individu
  - Slope tetap tetapi intersep berbeda antar individu
  - Slope tetap tetapi intersep berbeda antar individu antar waktu
  - Semua koefisien (slope dan intersep) berbeda antar individu
  - Semua koefisien berbeda antar individu dan antar waktu
- Berdasarkan variasi asumsi tsb, terdapat 3 pendekatan model yaitu:
  1. Metode *Common-Constant* (*The Pooled OLS Method=PLS*)
  2. Metode *Fixed Effect* (FEM)
  3. Metode *Random Effect* (REM)

# Pendekatan dalam Model Regresi Data Panel

## Metode Common-Constant (Pooled Ordinary Least Square = PLS)

- ◎ Menggunakan metode OLS biasa.
- ◎ Diasumsikan setiap unit individu memiliki intersep dan slope yang sama (tidak ada perbedaan pada dimensi *antar waktu*).
- ◎ Regresi panel data yang dihasilkan berlaku untuk *setiap individu*.

## Metode Fixed Effect (Fixed Effect Model=FEM)

- ◎ Intersep dibedakan antar individu.  $\text{Corr}(u_i, X_{it}\beta)$  atau  $E(u_i, X_{it}\beta) \neq 0$
- ◎ Dalam membedakan intersepnya dapat digunakan peubah *dummy*.
- ◎ Metode ini dikenal dgn model *Least Square Dummy Variable* (LSDV).

## Metode Random Effect (Random Effect Model=REM)

- ◎ Intersep tidak dianggap konstan, namun dianggap sebagai peubah *random* dengan suatu **nilai rata-rata**.  $\text{Corr}(u_i, X_{it}\beta)$  atau  $E(u_i, X_{it}\beta) = 0$  dengan  $u_i \sim N(0, \sigma_u^2)$
- ◎ Metode random dikenal dgn sebutan *Error Components Model* (ECM)

## Dua teknik standar analisis data panel: FEM & REM

**xtreg**: perintah Stata untuk menduga FEM & REM

**xtset**: untuk *setting* data panel sebelum perintah xtreg, misal:

. **xtset country year**

```
panel variable: country (strongly balanced)
time variable: year, 1990 to 1999
delta: 1 unit
```

Dalam hal ini "country" mewakili entitas atau panel (*i*) dan "year" mewakili Variabel waktu (*t*).

Catatan "(strongly balanced)" mengacu pada fakta bahwa semua negara memiliki data untuk semua tahun. Jika, misalnya, satu negara tidak memiliki data selama satu tahun maka data tersebut *unbalanced*.

Idealnya kita ingin memiliki dataset yang *balanced* tetapi ini meskipun *unbalanced* kita masih dapat menduga model.

Catatan: Jika ada *error* setelah menggunakan **xtset**, misalnya

```
varlist: country: string variable not allowed
```

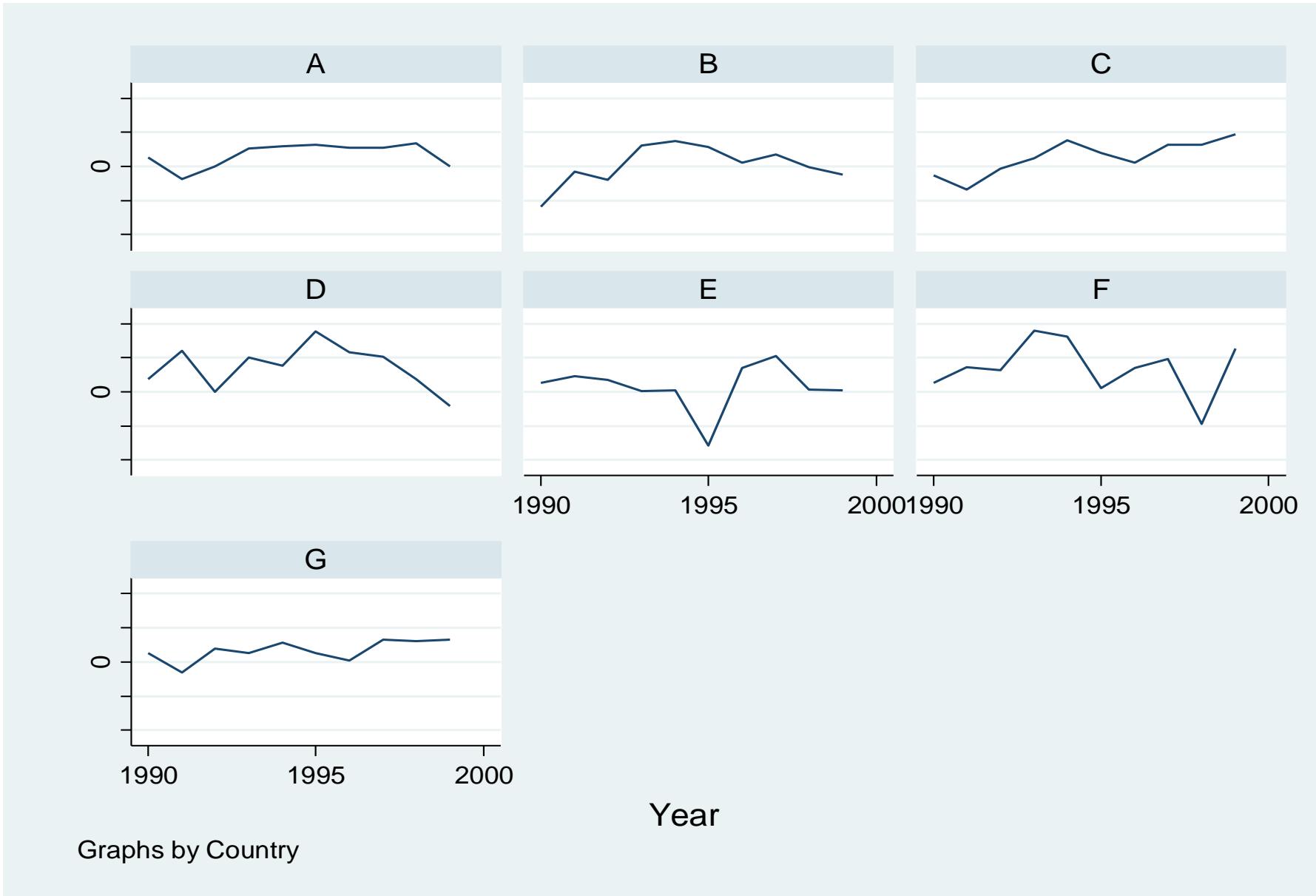
Kita perlu mengkonversi 'country' ke numerik, misalnya

```
encode country, gen(country1)
```

Gunakan 'country1' sebagai pengganti 'country' dalam perintah **xtset**

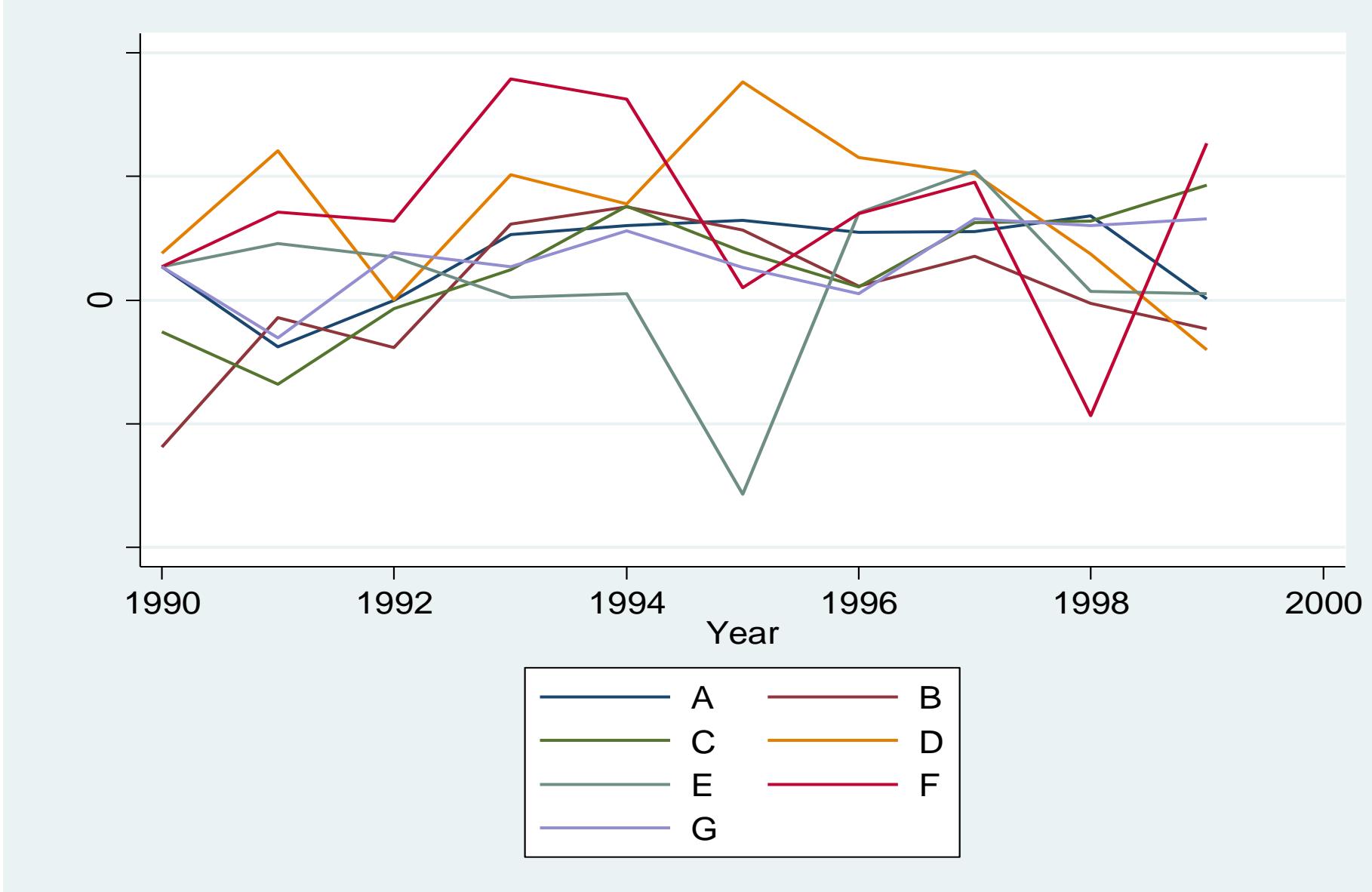
# Eksplorasi Data Panel

Xtline y /\*eksplorasi perkembangan data y utk masing2 country panel\_ID\*/



`xtline y, overlay`

\*Graphics>Panel-data linePlots: overlay each panel on same graph\*



# 1. Fixed Effects (Covariance) Model

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

Gunakan FEM jika kita hanya tertarik untuk menganalisis dampak variabel yang bervariasi dari waktu ke waktu.

FEM mengeksplorasi hubungan antara variabel penjelas (x) dan hasil (y) dalam suatu entitas. Setiap entitas memiliki karakteristik masing-masing yang mungkin (atau tidak) mempengaruhi pengaruh variabel penjelas (lihat penjelasan **interaksi Dummy dengan variabel x**).

Kita berasumsi bahwa sesuatu dalam diri individu **dapat berdampak atau membiaskan variabel penjelas atau hasil**, dan kita perlu mengendalikannya. Inilah alasan di balik **asumsi korelasi antara error komponen cross section atau entitas**; (dengan variabel penjelas).  $\text{Corr}(u_i, X_{it}\beta) \neq 0$  atau  $E(u_{it}, u_{i't}) \neq 0$

FEM dapat **menghilangkan pengaruh dari karakteristik waktu (yg tetap)** sehingga kita dapat menilai pengaruh bersih dari penjelas (X) terhadap variabel hasil (Y).

Asumsi penting lainnya adalah bahwa karakteristik waktu (yg tetap) itu **unik untuk individu dan tidak boleh dikorelasikan dengan karakteristik individu** lainnya.

Masing-masing entitas berbeda sehingga **komponen error entitas dan konstanta (Dummy yang merefleksikan karakteristik individu)** seharusnya **tidak berkorelasi dengan variabel lainnya**.

**Jika komponen error berkorelasi, maka FEM tidak cocok** karena kesimpulan mungkin tidak benar dan kita perlu memodelkan hubungan itu (mungkin menggunakan REM), inilah alasan utama untuk **uji Hausman**.

Pers FEM umumnya Jika hanya intersep  $\alpha$  yang berubah sepanjang waktu dan berbeda antar daerah, maka dapat diatasi dengan memasukkan peubah dummy untuk memungkinkan perbedaan intersep  $\alpha$ . Model ini disebut model efek tetap atau (*Covariance Model*) :

$$Y_{it} = \alpha + \beta X_{it} + \gamma_2 W_{2t} + \gamma_3 W_{3t} + \dots + \gamma_N W_{Nt} \\ + \delta_2 Z_{i2} + \delta_3 Z_{i3} + \dots + \delta_T Z_{iT} + \varepsilon_{it}, \quad (5.8)$$

dimana

$$W_{it} = \begin{cases} 1 & \text{untuk individu daerah ke-}i; \quad i = 2, 3, \dots, N \\ 0 & \text{selainnya.} \end{cases}$$

$$Z_{it} = \begin{cases} 1 & \text{untuk periode waktu ke-}t; \quad t = 2, 3, \dots, T \\ 0 & \text{selainnya.} \end{cases}$$

FEM diatas sebenarnya merupakan **model regresi peubah dummy dengan menambah (N+T-2) peubah bebas dummy**. Jika kita menduga model (5.8) dengan OLS akan menghasilkan dugaan semua parameter yang tak bias dan konsisten jika asumsi klasik lainnya dipenuhi. Koefisien *dummy* mengukur **perbedaan intersep antar daerah i dengan daerah pertama (i=1)**, dan **perubahan intersep antara waktu t dengan waktu pertama (t=1)**. Coba uraikan model tiap daerah, dan tiap waktu.

## 1a. Model Efek Tetap (*Covariance Model*): FEM

Untuk menguji apakah intersep  $\alpha$  konstan pada tiap daerah  $i$  dan tiap waktu  $t$ , maka digunakan statistik uji berikut yang menyebar menurut sebaran F dengan  $db1=(N+T-2)$  dan  $db2=NT-N-T$ .

$$F_{(N+T-2, NT-N-T)} = \frac{(JKS_1 - JKS_2)/(N + T - 2)}{(JKS_2)/(NT - N - T)} \quad (5.9)$$

$JKS_1$  (Jumlah Kuadrat Sisa dalam ANOVA) hasil model **OLS** dan  $JKS_2$  hasil model efek tetap. Jika nilai- $p < \alpha$  atau  $(F_{\text{hit}} > F_{\alpha(db1, db2)})$  maka terbukti bahwa asumsi (restriksi) semua intersep konstan tidak benar.

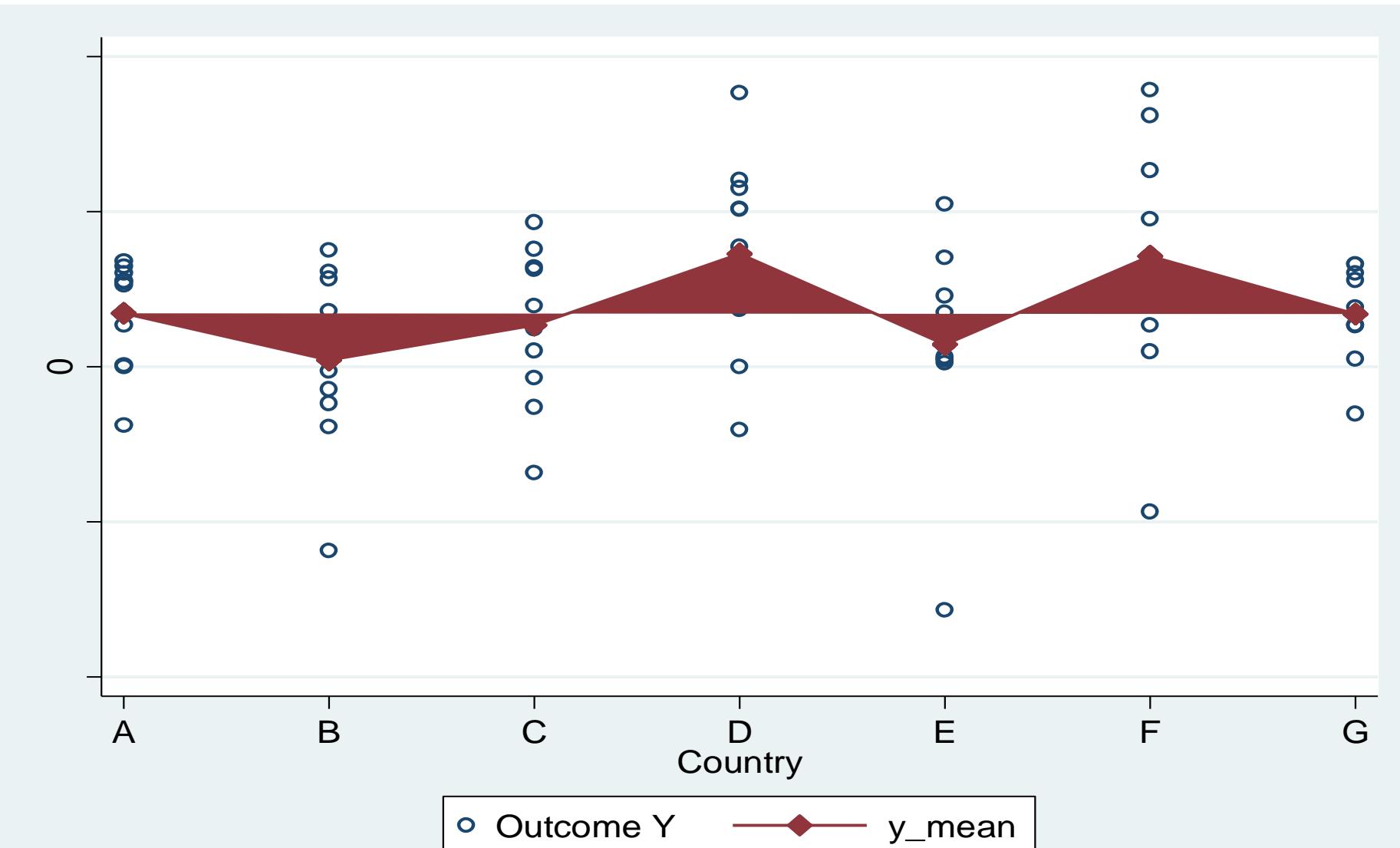
Ada beberapa masalah penggunaan model efek tetap. Pertama, penggunaan *dummy* tidak langsung mengidentifikasi apa yang menyebabkan pergeseran garis regresi sepanjang waktu dan antar daerah. Kedua, teknik *dummy* mengurangi derajat bebas  $(N+T-2)$  cukup besar.

## FEM: Keheterogenan antar entitas (*country*)

```

bysort country: egen y_mean=mean(y)
twoway scatter y country, msymbol(circle_hollow) || connected y_mean country,
msymbol(diamond) ||, xlabel(1 "A" 2 "B" 3 "C" 4 "D" 5 "E" 6 "F" 7 "G")

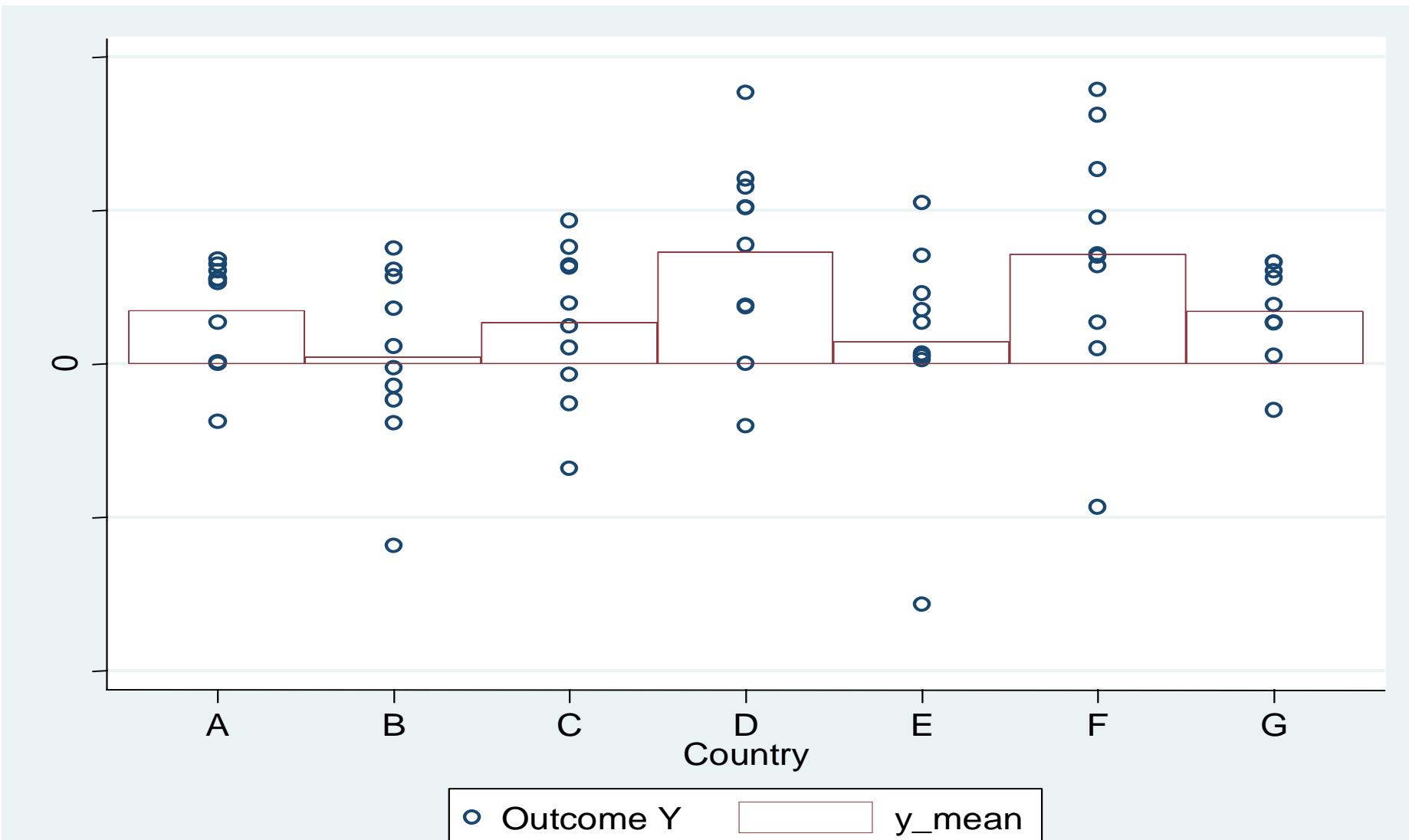
```



**Keheterogenan:** variabel2 yg tidak teramati yang tidak berubah sepanjang waktu

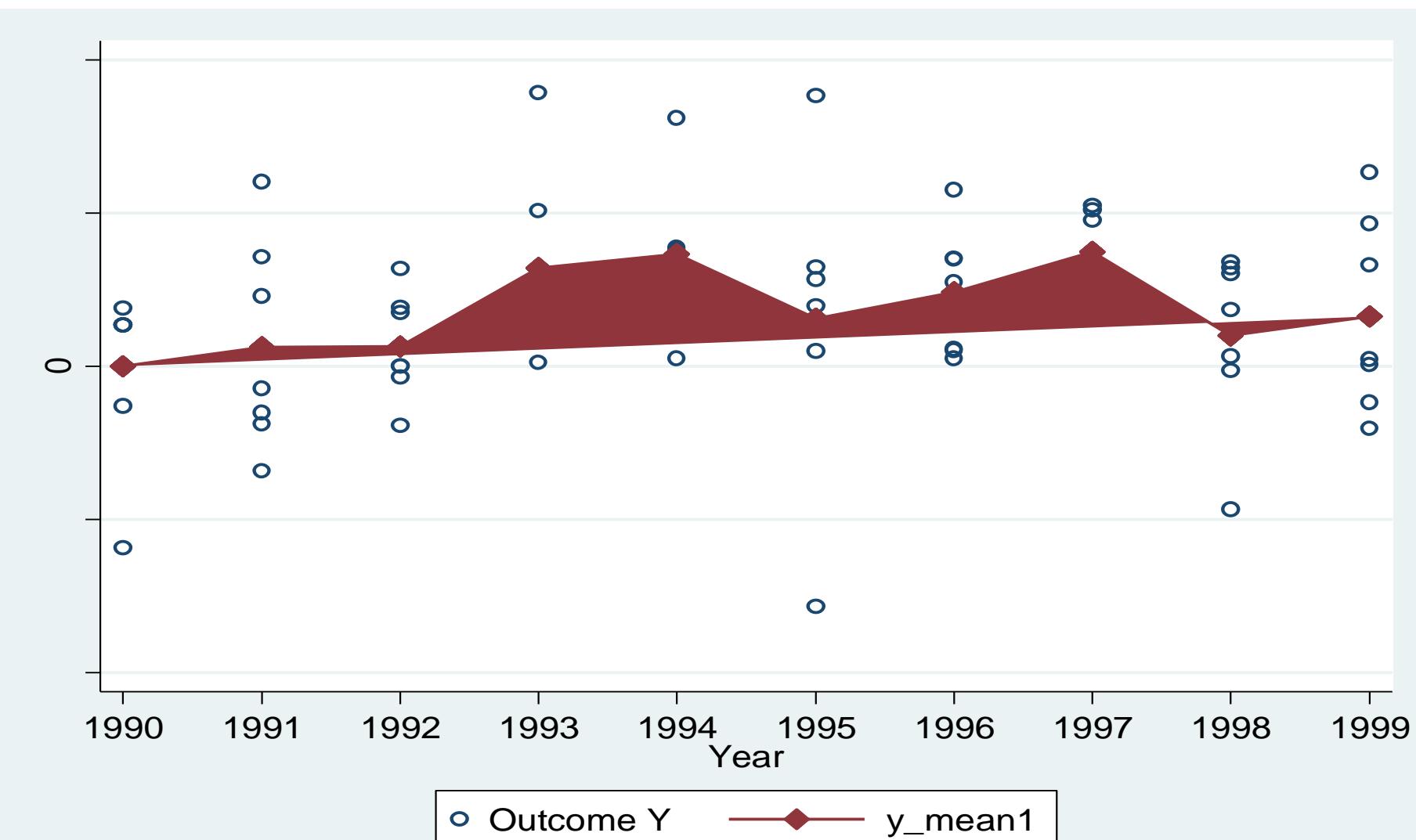
## FEM: Keheterogenan antar entitas (country)

```
twoway scatter y country, msymbol(circle_hollow) || bar y_mean country,  
fcolor(none) ||, xlabel(1 "A" 2 "B" 3 "C" 4 "D" 5 "E" 6 "F" 7 "G")
```



# FEM: Keheterogenan antar Tahun (year)

```
by sort year: egen y_mean1=mean(y)  
twoway scatter y year, msymbol(circle_hollow) || connected y_mean1 year,  
msymbol(diamond) ||, xlabel(1990(1)1999)
```



# OLS Regression

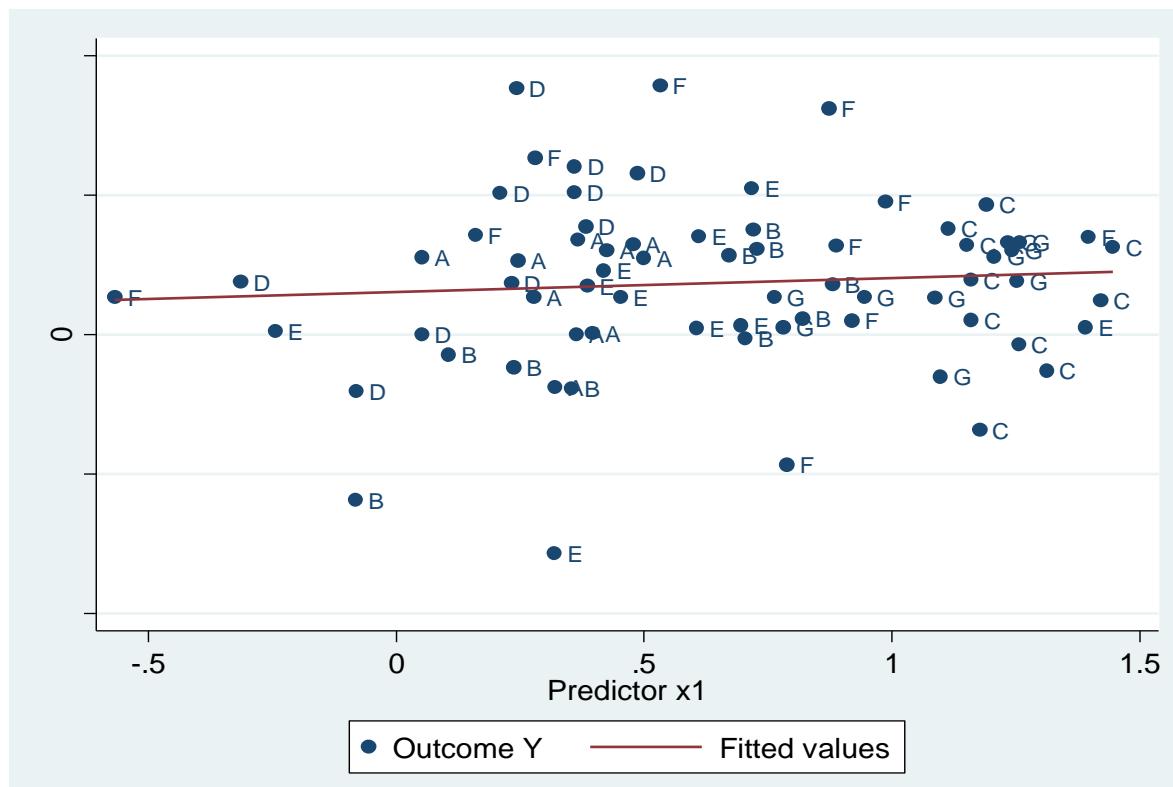
Source	SS	df	MS	Number of obs	=	70
Model	3.7039e+18	1	3.7039e+18	F(1, 68)	=	0.40
Residual	6.2359e+20	68	9.1705e+18	Prob > F	=	0.5272
				R-squared	=	0.0059
				Adj R-squared	=	-0.0087
Total	6.2729e+20	69	9.0912e+18	Root MSE	=	3.0e+09

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	4.95e+08	7.79e+08	0.64	0.527	-1.06e+09 2.05e+09
_cons	1.52e+09	6.21e+08	2.45	0.017	2.85e+08 2.76e+09

. regress y x1

. twoway scatter y x1,  
mlabel(country) || lfit y x1,  
clstyle(p2)



# FEM menggunakan Least Squares Dummy Variable (LSDV) model

Source	SS	df	MS	Number of obs	=	70
Model	1.4276e+20	7	2.0394e+19	F(7, 62)	=	2.61
Residual	4.8454e+20	62	7.8151e+18	Prob > F	=	0.0199
Total	6.2729e+20	69	9.0912e+18	R-squared	=	0.2276
				Adj R-squared	=	0.1404
				Root MSE	=	2.8e+09

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	2.48e+09	1.11e+09	2.24	0.029	2.63e+08 4.69e+09
country					
B	-1.94e+09	1.26e+09	-1.53	0.130	-4.47e+09 5.89e+08
C	-2.60e+09	1.60e+09	-1.63	0.108	-5.79e+09 5.87e+08
D	2.28e+09	1.26e+09	1.81	0.075	-2.39e+08 4.80e+09
E	-1.48e+09	1.27e+09	-1.17	0.247	-4.02e+09 1.05e+09
F	1.13e+09	1.29e+09	0.88	0.384	-1.45e+09 3.71e+09
G	-1.87e+09	1.50e+09	-1.25	0.218	-4.86e+09 1.13e+09
_cons	8.81e+08	9.62e+08	0.92	0.363	-1.04e+09 2.80e+09

`predict yhat`

`separate y, by(country)`

`separate yhat, by(country)`

`twoway connected yhat1-yhat7`

`x1, symbol(nonediamond_hollow`

`triangle_hollow`

`square_hollow +`

`circle_hollow x)`

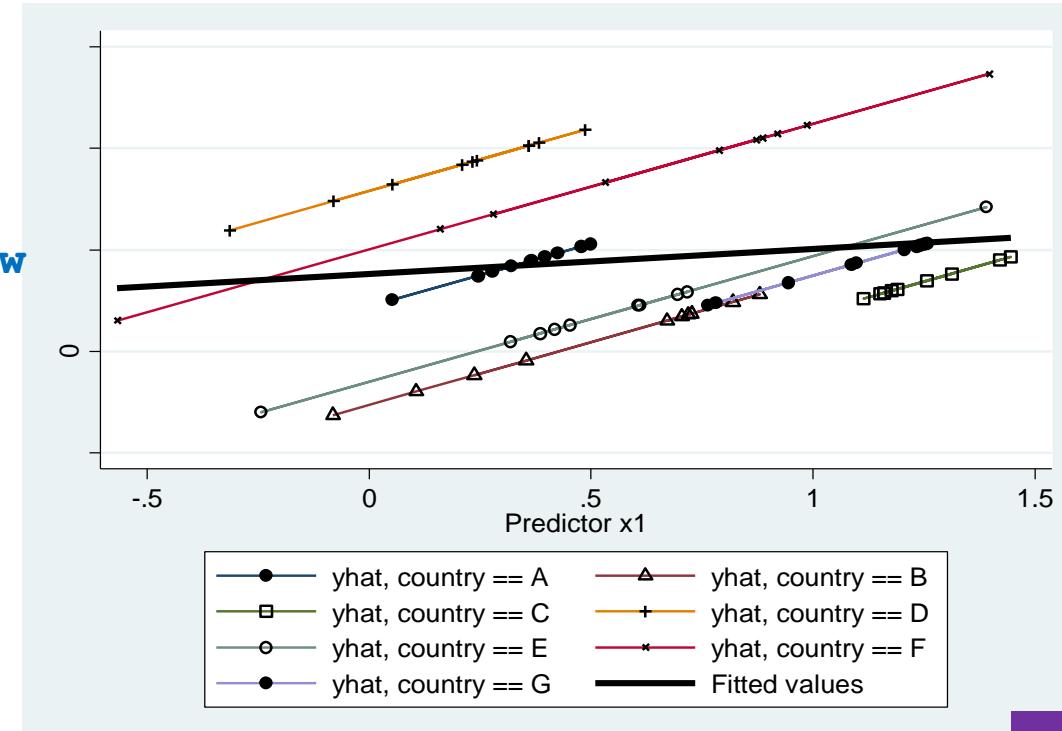
`msize(medium) mcolor(black`

`black black black black`

`black black) || lfit y x1,`

`clwidth(thick)`

`clcolor(black)`



## FEM: Keheterogenan antar entitas (country)

Model LSDV memberikan cara yang baik untuk memahami FEM.

Pengaruh  $x_1$  dimediasi oleh perbedaan antar negara.

Dengan menambahkan *dummy* untuk masing2 negara, kita menduga pengaruh murni  $x_1$  (dengan mengendalikan heterogenitas yang tidak teramat).

Setiap *dummy* merefleksikan pengaruh tertentu untuk masing-masing negara.

Variable	ols	ols_dum
x1	4.950e+08	2.476e+09*
country		
B		-1.938e+09
C		-2.603e+09
D		2.282e+09
E		-1.483e+09
F		1.130e+09
G		-1.865e+09
_cons	1.524e+09*	8.805e+08
N	70	70
r2	.00590459	.22757701
r2_a	-.00871446	.14036796

`regress y x1  
estimates store ols`

`regress y x1 i.country  
estimates store ols_dum`

`estimates table ols ols_dum,  
star stats(N r2 r2_a)`

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

## FEM: n intersep spesifik-entitas menggunakan xtreg

**Xtreg y x1, fe**

Fixed-effects (within) regression Group variable: country	Number of obs = 70 Number of groups = 7
R-sq:	Obs per group:
within = 0.0747	min = 10
between = 0.0763	avg = 10.0
overall = 0.0059	max = 10
corr(u_i, xb) = -0.5468	F(1, 62) = 5.00 Prob > F = 0.0289
-----	
y	Coef. Std. Err. t P> t  [95% Conf. Interval]
x1	2.48e+09 1.11e+09 2.24 0.029 2.63e+08 4.69e+09
_cons	2.41e+08 7.91e+08 0.30 0.762 -1.34e+09 1.82e+09
sigma_u	1.818e+09
sigma_e	2.796e+09
rho	.29726926 (fraction of variance due to u_i)

F test that all u\_i=0: F(6, 62) = 2.97 Prob > F = 0.0131

**regress y x1 i.country**

**NOTE: Add the option 'robust' to control for heteroskedasticity**

Source	SS	df	MS	Number of obs = 70
Model	1.4276e+20	7	2.0394e+19	F(7, 62) = 2.61
Residual	4.8454e+20	62	7.8151e+18	Prob > F = 0.0199
Total	6.2729e+20	69	9.0912e+18	R-squared = 0.2276 Adj R-squared = 0.1404 Root MSE = 2.8e+09

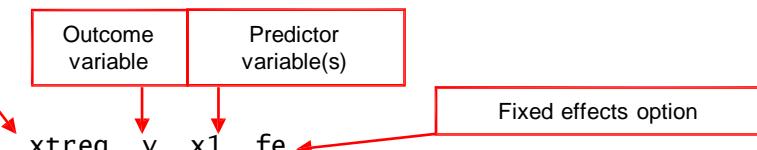
**FEM**  
**menggunakan**  
**dummy OLS**

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	2.48e+09	1.11e+09	2.24	0.029	2.63e+08 4.69e+09
country					
B	-1.94e+09	1.26e+09	-1.53	0.130	-4.47e+09 5.89e+08
C	-2.60e+09	1.60e+09	-1.63	0.108	-5.79e+09 5.87e+08
D	2.28e+09	1.26e+09	1.81	0.075	-2.39e+08 4.80e+09
E	-1.48e+09	1.27e+09	-1.17	0.247	-4.02e+09 1.05e+09
F	1.13e+09	1.29e+09	0.88	0.384	-1.45e+09 3.71e+09
G	-1.87e+09	1.50e+09	-1.25	0.218	-4.86e+09 1.13e+09
_cons	8.81e+08	9.62e+08	0.92	0.363	-1.04e+09 2.80e+09

# Fixed effects: $n$ entity-specific intercepts (using xtreg)

$$Y_{it} = \beta_1 X_{1it} + \dots + \beta_k X_{kit} + \alpha_i + e_{it} \quad [\text{see eq.1}]$$

**NOTE:** Add the option 'robust' to control for heteroskedasticity



Fixed-effects (within) regression  
Group variable: country

R-sq: within = 0.0747  
between = 0.0763  
overall = 0.0059

$$\text{corr}(u_i, x_b) = -0.5468$$

The errors  $u_i$  are correlated with the **regressors** in the fixed effects model

Total number of cases (rows)	
Number of obs	= 70
Number of groups (entities)	= 7
Obs per group: min	= 10
avg	= 10.0
max	= 10
F(1, 62)	= 5.00
Prob > F	= 0.0289

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.

Coefficients of the regressors. Indicate how much Y changes when X increases by one unit.	y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	x1	2.48e+09	1.11e+09	2.24	0.029	2.63e+08 4.69e+09
	_cons	2.41e+08	7.91e+08	0.30	0.762	-1.34e+09 1.82e+09

sigma\_u  
sigma\_e  
rho

29.7% of the variance is due to differences across panels.  
'rho' is known as the intraclass correlation

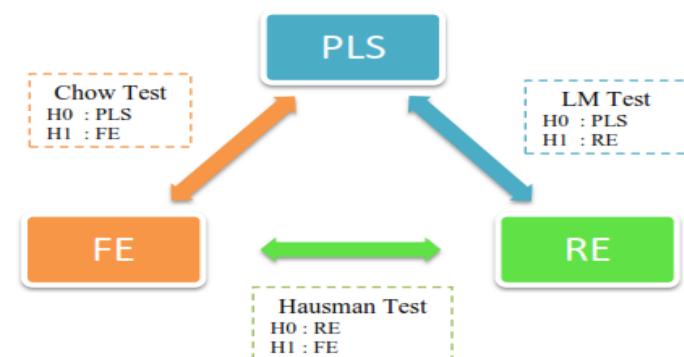
F test that **all  $u_i=0$** :  $F(6, 62) = 2.97$  Prob > F = 0.0131

$$\rho = \frac{(\sigma_u)^2}{(\sigma_u)^2 + (\sigma_e)^2}$$

$\sigma_u$  = sd of residuals **within groups  $u_i$**   
 $\sigma_e$  = sd of residuals (overall error term)  $e_i$

**Chow Test: H<sub>0</sub>: PLS vs H<sub>1</sub>: FEM**

For more info see Hamilton, Lawrence,  
*Statistics with STATA*.



## Cara lain menduga FEM: n intersep spesifik-entitas menggunakan areg

Meskipun outputnya kurang informatif dibandingkan regresi dgn variabel dummy, **areg** memiliki dua keunggulan. Pertama, **mempercepat hasil eksplorasi & memberikan umpan balik** tentang apakah pendekatan variabel dummy bermanfaat. Kedua, ketika variabel (jumlah entitas) yang dikaji banyak, membuat *dummy* untuk masing-masing entitas menyebabkan terlalu banyak variabel atau modelnya terlalu besar

. areg y x1, absorb(country)

Hide the binary variables for each entity

NOTE: Add the option 'robust' to control for heteroskedasticity

Linear regression, absorbing indicators

Number of obs = 70  
F( 1, 62) = 5.00  
Prob > F = 0.0289  
R-squared = 0.2276  
Adj R-squared = 0.1404  
Root MSE = 2.8e+09

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	2.48e+09	1.11e+09	2.24	0.029	2.63e+08 4.69e+09
_cons	2.41e+08	7.91e+08	0.30	0.762	-1.34e+09 1.82e+09
country		F(6, 62) = 2.965		0.013	(7 categories)

Fixed-effects (within) regression  
Group variable: country

Number of obs = 70  
Number of groups = 7

**xtreg y x1, fe**

R-sq:  
within = 0.0747  
between = 0.0763  
overall = 0.0059

Obs per group:  
min = 10  
avg = 10.0  
max = 10

corr(u\_i, Xb) = -0.5468  
F(1, 62) = 5.00  
Prob > F = 0.0289

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	2.48e+09	1.11e+09	2.24	0.029	2.63e+08 4.69e+09
_cons	2.41e+08	7.91e+08	0.30	0.762	-1.34e+09 1.82e+09
sigma_u	1.818e+09				
sigma_e	2.796e+09				
rho	.29726926				(fraction of variance due to u_i)

F test that all u\_i=0: F(6, 62) = 2.97

Prob > F = 0.0131

# Membandingkan 3 Metode Pendugaan FEM (Semuanya memberikan hasil yang sama)

```
xtreg y x1 x2 x3, fe
estimates store fixed
regress y x1 x2 x3 i.country
estimates store ols_dum
areg y x1 x2 x3, absorb(country) /*Dalam perbandingan R2 hanya bermakna utk regress & areg*/
estimates store areg
estimates table fixed ols_dum areg, star stats(N r2 r2_a)
```

Variable	fixed	ols_dum	areg
x1	2.425e+09*	2.425e+09*	2.425e+09*
x2	1.823e+09	1.823e+09	1.823e+09
x3	3.097e+08	3.097e+08	3.097e+08
country			
2		-5.961e+09	
3		-1.598e+09	
4		-2.091e+09	
5		-5.732e+09	
6		8.026e+08	
7		-1.375e+09	
_cons	-2.060e+08	2.073e+09	-2.060e+08
N	70	70	70
r2	.10092442	.24948198	.24948198
r2_a	-.03393692	.13690428	.13690428

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

## Testing for time-fixed effects

**xtreg y x1 i.year, fe**

Fixed-effects (within) regression  
Group variable: country

R-sq:

within = 0.2323  
between = 0.0763  
overall = 0.1395

corr(u\_i, Xb) = -0.2014

Number of obs = 70  
Number of groups = 7  
  
Obs per group:  
min = 10  
avg = 10.0  
max = 10  
  
F(10,53) = 1.60  
Prob > F = 0.1311

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	1.39e+09	1.32e+09	1.05	0.297	-1.26e+09 4.04e+09
year					
1991	2.96e+08	1.50e+09	0.20	0.844	-2.72e+09 3.31e+09
1992	1.45e+08	1.55e+09	0.09	0.925	-2.96e+09 3.25e+09
1993	2.87e+09	1.50e+09	1.91	0.061	-1.42e+08 5.89e+09
1994	2.85e+09	1.66e+09	1.71	0.092	-4.84e+08 6.18e+09
1995	9.74e+08	1.57e+09	0.62	0.537	-2.17e+09 4.12e+09
1996	1.67e+09	1.63e+09	1.03	0.310	-1.60e+09 4.95e+09
1997	2.99e+09	1.63e+09	1.84	0.072	-2.72e+08 6.26e+09
1998	3.67e+08	1.59e+09	0.23	0.818	-2.82e+09 3.55e+09
1999	1.26e+09	1.51e+09	0.83	0.409	-1.77e+09 4.29e+09
_cons	-3.98e+08	1.11e+09	-0.36	0.721	-2.62e+09 1.83e+09
sigma_u	1.547e+09				
sigma_e	2.754e+09				
rho	.23985725	(fraction of variance due to u_i)			

F test that all u\_i=0: F(6, 53) = 2.45

Prob > F = 0.0362

Tidak ada alasan menolak Ho bahwa koefisien untuk semua tahun secara bersama-sama, sama dgn nol. Oleh karena itu **tidak diperlukan FE waktu** dalam kasus ini

Untuk melihat apakah **FE waktu** diperlukan ketika menduga FEM, gunakan command **testparm**. Ini adalah tes untuk melihat apakah *dummy* untuk semua tahun, sama dengan 0, jika ya, maka FE waktu tidak diperlukan)

**testparm i.year**

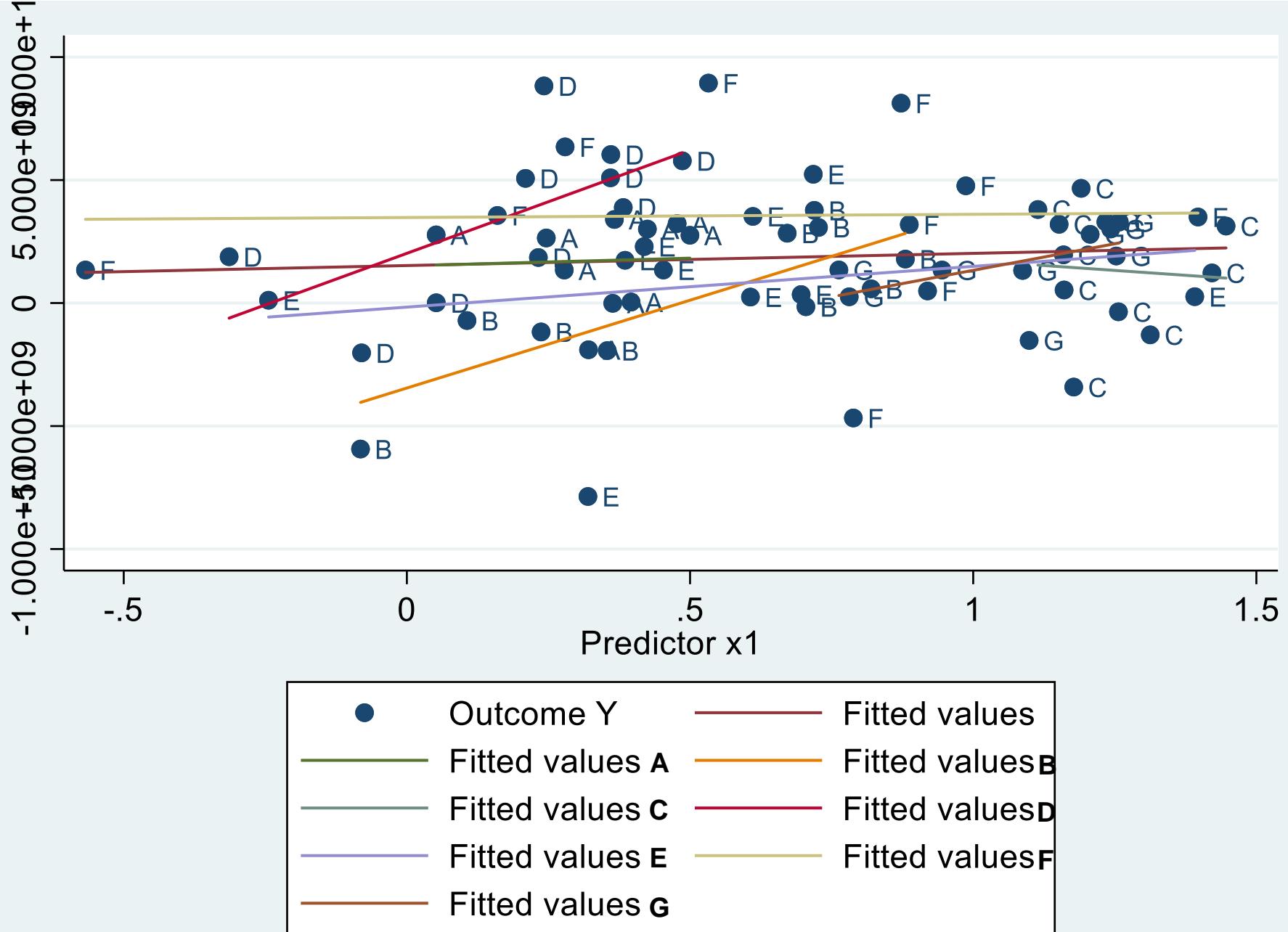
- ( 1 ) 1991.year = 0
- ( 2 ) 1992.year = 0
- ( 3 ) 1993.year = 0
- ( 4 ) 1994.year = 0
- ( 5 ) 1995.year = 0
- ( 6 ) 1996.year = 0
- ( 7 ) 1997.year = 0
- ( 8 ) 1998.year = 0
- ( 9 ) 1999.year = 0

F(9,53) = 1.21  
Prob > F = 0.3094

## Catatan Model Efek Tetap (FEM) yang Standar

- FEM mengendalikan semua **perbedaan antar individu** (bukan karena variasi waktu shg intersepnya tidak bervariasi tiap waktu) sehingga diharapkan dugaan koefisiennya tidak bias karena karakteristiknya (budaya, agama, Jenis Kelamin, RAS dll) dihilangkan.
- Kelemahan FEM adalah mengasumsikan pengaruh variabel penjelas X sama saja untuk tiap individu (**realistik?**)  
→ **dipecahkan dgn dummy interaksi (Juanda, 2009)**

*Technically, time-invariant characteristics of the individuals are perfectly collinear with the person [or entity] dummies. Substantively, however, fixed-effects models are designed to study the causes of changes within a person [or entity]. A time-invariant characteristic cannot cause such a change, because it is constant for each person." (Underline is mine) Kohler, Ulrich, Frauke Kreuter, Data Analysis Using Stata, 2<sup>nd</sup> ed., p.245*



```
twoway (scatter y x1,
mlabel(country)) (lfit y x1)
(lfit y x1 if country==1)
(lfit y x1 if country==2)
(lfit y x1 if country==3)
(lfit y x1 if country==4)
(lfit y x1 if country==5)
(lfit y x1 if country==6)
(lfit y x1 if country==7)
```

Mudah memahami  
keheterogenan  
**pengaruh x1 utk**  
**masing2 country**  
dapat berbeda  
**slopenya**

## 2. Model Efek Acak (*Error-Components Model*): REM

Memasukkan peubah *dummy* menggambarkan kekurangan pengetahuan mengenai model, maka cukup logis menggambarkan kekurangan pengetahuan ini melalui komponen sisaan. Kita dapat mengasumsikan data panel tersebut mempunyai karakteristik komponen sisaannya berkorelasi (bervariasi) antar waktu dan antar individu daerah. Model komponen sisaan (*error*) ini disebut juga model efek acak, misalnya dengan model berikut:

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it}, \\ \varepsilon_{it} = u_i + v_t + w_{it}, \quad \text{untuk } i = 1, 2, \dots, N; t = 1, 2, \dots, T \quad (5.10a)$$

dimana  $u_i \sim N(0, \sigma_u^2)$  : komponen sisaan data *cross-section*

$v_t \sim N(0, \sigma_v^2)$  : komponen sisaan data *time series*

$w_{it} \sim N(0, \sigma_w^2)$  : komponen sisaan gabungan.

Lihat P.Balestra dan M.Nerlove, “Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: the Demand for Natural Gas” dalam *Econometrica*, Vol 34(4):585-612, 1966.

## 2. Model Efek Acak (*Error-Components Model*): REM

Hubungan antara *error-components model* dan *covariance model* dapat dilihat dengan memperlakukan komponen-komponen intersep dalam *covariance model* sebagai **2 peubah acak**, satu peubah deret waktu dan satunya peubah *cross-section*. Jika kedua peubah acak ini menyebar normal, **derajat bebas sisaan dapat dihemat** sehingga kita hanya memperhatikan rataan dan ragam dari masing-masing komponen sisaan. Formulasi *error-components* diperoleh dari *covariance model* dengan mengasumsikan bahwa rataan efek acak dari *time series* dan *cross-section* dimasukkan dalam komponen intersep, dan deviasi acak disekitar rataannya disamakan dengan masing-masing *error-components*  $u_i$  dan  $v_t$ . Dari (5.10b), ragam sisaan  $\varepsilon_{it}$  dapat diperoleh dengan rumus:

$$\text{Var} (\varepsilon_{it}) = \sigma_u^2 + \sigma_v^2 + \sigma_w^2 \quad (5.11)$$

Model OLS yang diterapkan pada data panel, mempunyai ragam sisaan  $\varepsilon_{it}$  dengan rumus:  $\text{Var} (\varepsilon_{it}) = \sigma_w^2$ . Oleh karena itu model efek acak dapat diduga dengan OLS jika  $\sigma_u^2 = \sigma_v^2 = 0$ . Jika tidak demikian, umumnya diduga dengan metode GLS (*Generalized Least Squares*) yang akan dibahas dalam Bab 9 dan 12.

- **Keragaman antar entitas** diasumsikan **acak dan tidak berkorelasi dengan variabel penjelas (prediktor)** yang ada dalam model.  
 $\text{Corr}(u_i, X_{it}\beta) = 0$  atau  $E(u_{it}, u_{i't'}) = 0$
- Perbedaan utama FE dgn RE adalah **apakah pengaruh individu yang tak teramat memasukkan unsur-unsur yang berkorelasi dengan variabel penjelas** dalam model, bukan apakah pengaruh ini stokastik atau tidak "[Green, 2008, p.183]
- Keuntungan dari REM adalah bahwa Anda dapat memasukkan variabel (misal jenis kelamin) yang tetap antar waktu; serta inferensi lebih luas dari samplenya. **Dalam model FEM variabel-variabel ini direfleksikan oleh intersep.**
- Perlu **menentukan karakteristik individu yang dapat mempengaruhi variabel prediktor**. Masalahnya disini adalah bahwa beberapa variabel mungkin tidak tersedia sehingga mengarah ke bias *omitted variable* dalam model.

# Random Effects Model

`xtreg y x1, re`

Random-effects GLS regression  
Group variable: country

outcome variable

R-sq:

within = 0.0747  
between = 0.0763  
overall = 0.0059

corr(u\_i, X) = 0 (assumed)

Number of obs = 70  
Number of groups = 7

Obs per group:  
min = 10  
avg = 10.0  
max = 10  
d chi2(1) = 1.91  
Prob > chi2 = 0.1669

Perbedaan antar unit tidak berkorelasi dengan regressor  
(Lihat Juanda, 2009)

uji model

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
x1	1.25e+09	9.02e+08	1.38	0.167	-5.21e+08 3.02e+09
_cons	1.04e+09	7.91e+08	1.31	0.190	-5.13e+08 2.59e+09
sigma_u	1.065e+09				
sigma_e	2.796e+09				
rho	.12664193				(fraction of variance due to u_i)

Interpretasi dari koefisien X1 agak rumit karena mencakup pengaruh di dalam entitas dan antar entitas. Jadi merepresentasikan rata-rata pengaruh X1 terhadap Y ketika X1 berubah antar waktu dan antar entitas sebesar satu unit.

# Hausman Test: Fixed or Random

Uji Hausman untuk menentukan FE atau RE.

$H_0$ : REM vs  $H_1$ : FEM (Green, 2008). Pada hakikatnya menguji apakah **unique errors ( $u_i$ )** berkorelasi dengan regressor (Ho: tidak berkorelasi)

```
. xtreg y x1, fe
. estimates store fixed
. xtreg y x1, re
. estimates store random
. hausman fixed random
```

	Coefficients			
	(b) fixed	(B) random	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
x1	2.48e+09	1.25e+09	1.23e+09	6.41e+08

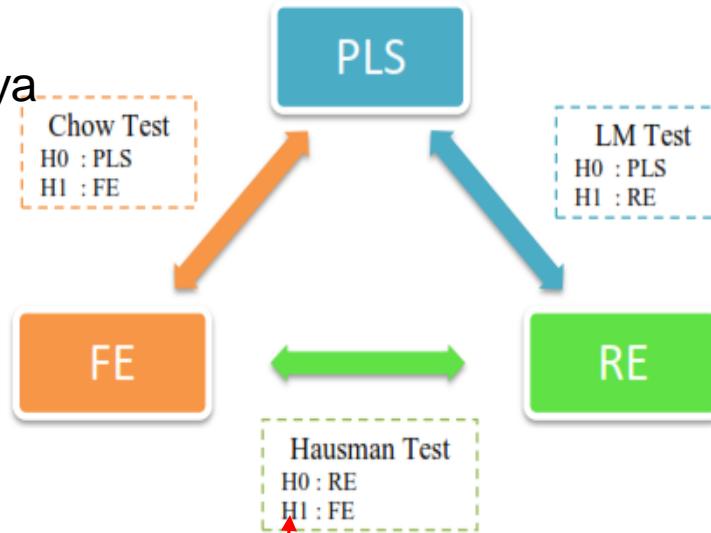
b = consistent under  $H_0$  and  $H_a$ ; obtained from xtreg

B = inconsistent under  $H_a$ , efficient under  $H_0$ ; obtained from xtreg

Test:  $H_0$ : difference in coefficients not systematic REM

$$\begin{aligned} \text{chi2}(1) &= (b-B)' [ (V_b-V_B)^{-1} ] (b-B) \\ &= 3.67 \end{aligned}$$

Prob>chi2 = 0.0553 > \alpha maka terima  $H_0$ :REM

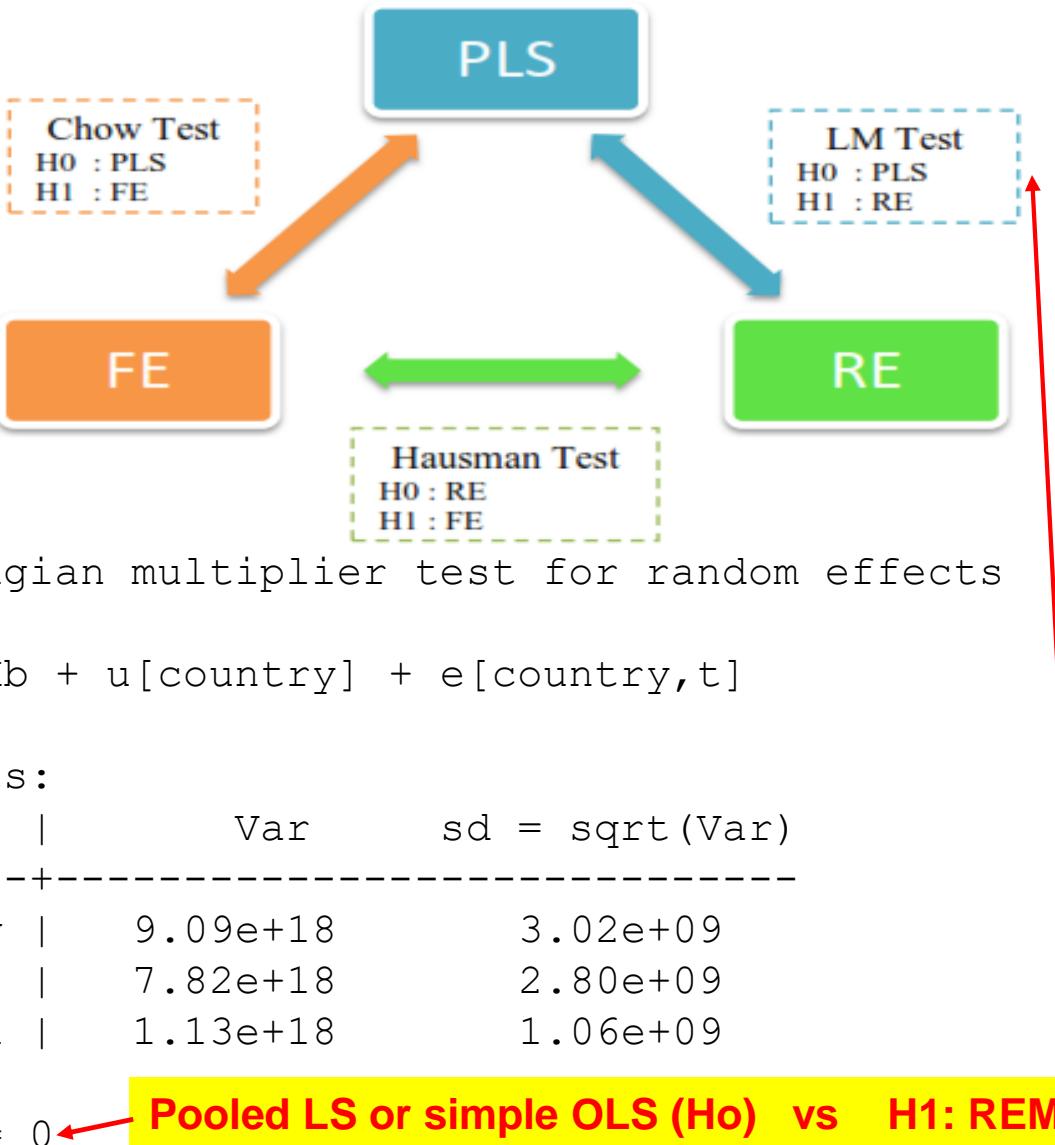


# Testing for REM: Breusch-Pagan Lagrange Multiplier (LM)

Uji LM untuk menentukan *REM* atau *simple OLS* (PLS).

$H_0$  adalah keragaman antar entitas sama dengan 0; atau **perbedaan antar entitas tidak signifikan** (Tidak ada efek panel).

- . **xtreg y x1, re**
- . **xttest0**



Breusch and Pagan Lagrangian multiplier test for random effects

$$y[\text{country}, t] = Xb + u[\text{country}] + e[\text{country}, t]$$

Estimated results:

	Var	sd = sqrt(Var)
-----+-----		
y	9.09e+18	3.02e+09
e	7.82e+18	2.80e+09
u	1.13e+18	1.06e+09

Test:  $\text{Var}(u) = 0$  **Pooled LS or simple OLS ( $H_0$ ) vs  $H_1$ : REM**

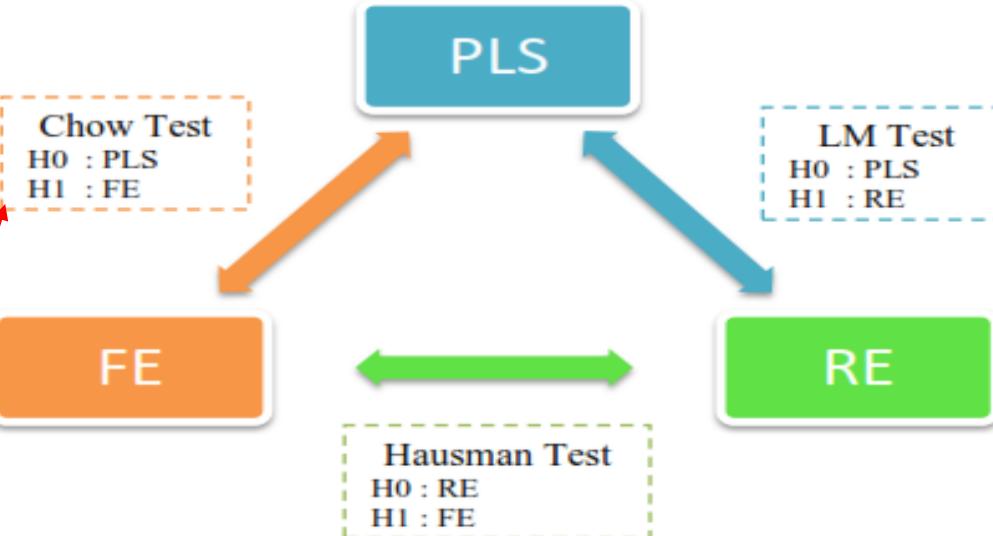
REM kurang tepat; tidak ada bukti adanya perbedaan signifikan antar entitas, sehingga *simple OLS* dipilih

chibar2(01) = 2.67  
Prob > chibar2 = 0.0512

# Testing for cross-sectional dependence/contemporaneous correlation using Breusch-Pagan Lagrange Multiplier (LM) test of independence

Ketergantungan *cross-sectional* adalah masalah di panel makro dengan seri waktu yang lama (lebih dari 20-30 tahun). Hal ini tidak banyak masalah dalam panel mikro (beberapa tahun dan banyak kasus).

`xtreg y x1, fe`  
`xttest2 alternatif`



Correlation matrix of residuals:

	e1	e2	e3	e4	e5	e6	e7
e1	1.0000						
e2	0.3615	1.0000					
e3	0.5391	0.4146	1.0000				
e4	0.3209	0.4660	-0.3015	1.0000			
e5	-0.2032	-0.3764	-0.3590	-0.3080	1.0000		
e6	-0.2572	0.2432	-0.0491	-0.1065	0.2321	1.0000	
e7	0.6403	0.0793	0.8206	-0.4358	-0.0818	0.0355	1.0000

Ho: Residual antar entitas tidak berkorelasi  
H1: Residual antar entitas berkorelasi

>  $\alpha$  maka  
terima Ho  
(beda  
hasil dgn  
Chow test)

Breusch-Pagan LM test of independence:  $\text{chi2}(21) = 28.914$ ,  $\text{Pr} = 0.1161$   
Based on 10 complete observations over panel units

# Testing for cross-sectional dependence/contemporaneous correlation using Pasaran CD test

Seperti disebutkan dalam slide sebelumnya, ketergantungan cross-sectional lebih merupakan masalah di panel makro dengan seri waktu yang lama (lebih dari 20-30 tahun) daripada di panel mikro. Tes Pesaran CD (cross-sectional dependence) digunakan untuk menguji apakah residu berkorelasi antar entitas \*. Ketergantungan cross-sectional dapat menyebabkan bias dalam hasil tes (juga disebut korelasi kontemporer)

Pesaran's test of cross sectional independence = 1.155, Pr = 0.2479  
Average absolute value of the off-diagonal elements = 0.316



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

- Suatu Uji heteroskedastisitas tersedia untuk FEM menggunakan perintah **xttest3**.
  - Ini adalah program yang ditulis pengguna, ketik: **ssc instal xttest3**
- xttest3**

**Modified Wald test for groupwise heteroskedasticity  
in fixed effect regression model**

H0:  $\sigma(i)^2 = \sigma^2$  for all i

chi2 (7) = 42.77  
Prob>chi2 = 0.0000

Hasil di atas, menolak nol dan menyimpulkan heteroskedastisitas.

**CATATAN:** Gunakan opsi ‘robust’ untuk mendapatkan standar errors dari heteroskedasticity-robust (juga dikenal sebagai Huber / White atau penaksir sandwich).

# Testing for Serial Correlation

- Uji korelasi serial berlaku untuk panel makro dengan seri waktu yang lama (lebih dari 20-30 tahun). Bukan masalah di panel mikro (yang waktunya sangat sedikit). Korelasi serial menyebabkan *standar error koefisien menjadi lebih kecil daripada yang sebenarnya*, dan juga R<sup>2</sup> lebih tinggi dari yang seharusnya.
- *Lagram-Multiplier dapat digunakan* untuk uji korelasi serial: **xtserial**.

**xtserial y x1**

Wooldridge test for autocorrelation in panel data

H0: no first-order autocorrelation

F( 1, 6) = 0.214

Prob > F = 0.6603

Tidak ada alasan Hipotesis menolak nol dan menyimpulkan tidak ada autokorelasi ordo-1.

# Tugas Mandiri

- Review apa yang dibahas dalam *synchronous e-learning*, dan gunakan data yang sama.
- Jika sudah faham, coba latihan dengan data panel perekonomian provinsi Jawa tengah yang sudah diberikan



# Faktor-faktor yang mempengaruhi belanja modal (lbmodalsp)

	Nasional	KBI	KTI
VARIABLES	Lbmodalsp	Lbmodalsp	Lbmodalsp
Ldausp	0.125	0.371**	-0.169
Ldbhsp	0.349***	0.293***	0.429***
Ldaksp	0.396***	0.181*	0.524***
Lgablainsp	-0.047	-0.011	-0.041
Lpadsp	0.192***	0.188***	0.261***
Constant	0.540	-0.304	1.511
Observations	160	79	81
P-Values in parentheses			
R-squared	0.902	0.910	0.941
*** p<0.01, ** p<0.05, * p<0.1			

- Hasil simulasi menunjukkan bahwa secara nasional DBH dan DAK memberikan pengaruh yang positif dan signifikan terhadap belanja modal. DAK merupakan variabel yang pengaruhnya paling besar, dimana kenaikan 1% DAK dapat meningkatkan belanja modal sebesar 0,396%.
- Simulasi antara Kawasan Barat Indonesia (KBI) dan Kawasan Timur Indonesia (KTI) menunjukkan hasil yang berbeda, dimana untuk KBI DAU, DBH dan DAK memberikan pengaruh yang positif dan signifikan terhadap belanja modal, dengan pengaruh terbesar berasal dari DAU. Sementara itu, KTI menunjukkan pola yang sama dengan nasional, dimana DBH dan DAK memberikan pengaruh yang positif dan signifikan terhadap belanja modal sedangkan pengaruh DAU tidak signifikan

# Latihan Kegiatan Mandiri

- Review apa yang dibahas dalam *synchronous e-learning*.
  1. Jelaskan tentang Data panel, dan Kentungannya menggunakan data panel.
  2. Jelaskan beberapa kemungkinan tentang intersep dan slope dari variable X antar entitas (individu) dan antar waktu.
  3. Dalam metode FEM, asumsi tentang intersep bagaimana? Jelaskan.
  4. Dalam metode REM, asumsi tentang intersep bagaimana? Jelaskan.
  5. Dalam model FEM dan/atau REM yang standar, pengaruh (*slope*) variable X dianggap konstan. Apakah asumsi ini realistik? Jelaskan
  6. Bagaimana memasukkan *technological improvement* dalam suatu model output (produksi) nasional (daerah)?
  7. Jelaskan tentang **Model Analisis Pengaruh/Dampak Kebijakan Desentralisasi Fiskal** dan berikan ilustrasinya.



Semoga bermanfaat  
Terima kasih  
**(Salam, BJ)**



**IPB University**  
— Bogor Indonesia —

**Departemen Ilmu Ekonomi  
Fakultas Ekonomi dan Manajemen  
Institut Pertanian Bogor**