



IPB University
— Bogor Indonesia —

Study Program
Statistics and Data Science
Department of Statistics

Econometric Analysis of Panel Data

Deri Siswara

Badi H. Baltagi

<https://link.springer.com/book/10.1007/978-3-030-53953-5>

PENDAHULUAN

■ Apa itu data panel?

person	year	income	age	sex
1	2003	1500	27	1
1	2004	1700	28	1
1	2005	2000	29	1
2	2003	2100	41	2
2	2004	2100	42	2
2	2005	2200	43	2

Data Panel: Pengamatan pada kumpulan individu dalam beberapa waktu tertentu.

*Micro-panels** = Large N, Small T
(30-100 N dan max 20 T)

Macro-panels = Large N and T
(T hingga 60 atau lebih)

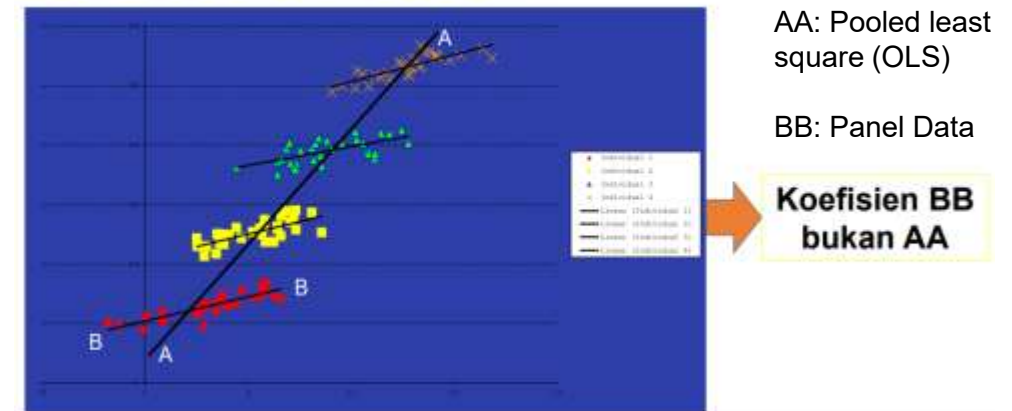
*Model umum dan sederhana



Kelebihan regresi data panel

1. Dapat menangkap *individual heterogeneity*.

Bias dari Estimasi OLS



2. Panel data memberikan *informative data*, *more variability*, *less collinearity* antar perubah, *more degrees of freedom*, dan lebih efisien (Tidak bias dan ragam minimum).

PENDAHULUAN



Kelebihan regresi data panel

3. Dapat menangkap *dynamics of adjustment*.

Pengaruh variabel independen dapat berubah dari waktu ke waktu atau dari individu ke individu dalam panel data.

4. Panel data are better able to identify and measure effects that are simply not detectable in pure cross-section or pure time-series data.
5. Panel data models allow us to construct and test more complicated behavioral models than purely cross-section or time-series data.



Limitasi regresi data panel

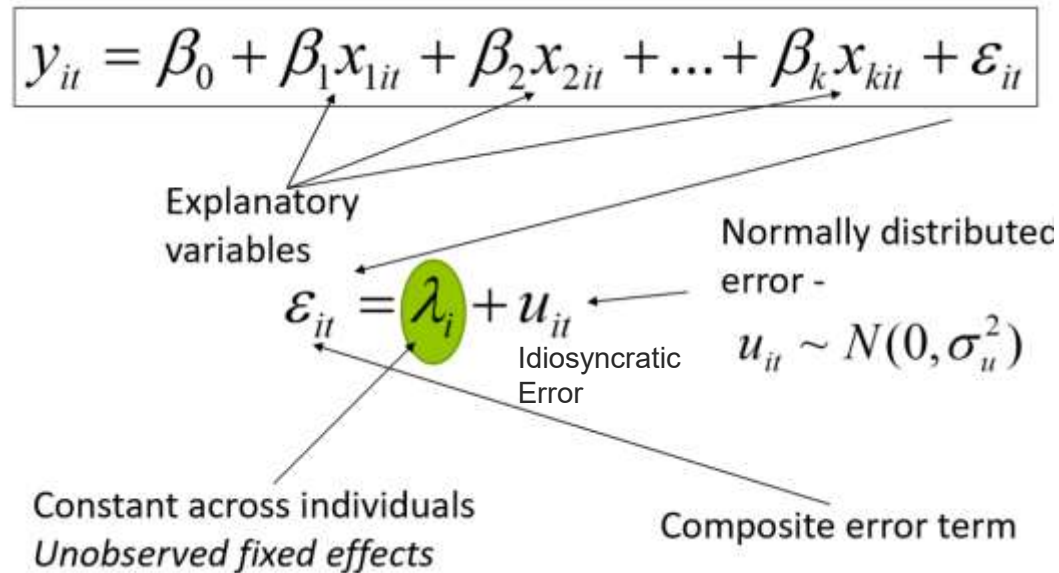
1. Design and data collection problems
2. Distortions of measurement errors

Ketidakakuratan dalam pengukuran dalam pengumpulan data yang dapat menyebabkan kesalahan dalam analisis regresi panel yang menyebabkan distorsi atau bias pada hasil analisis regresi.

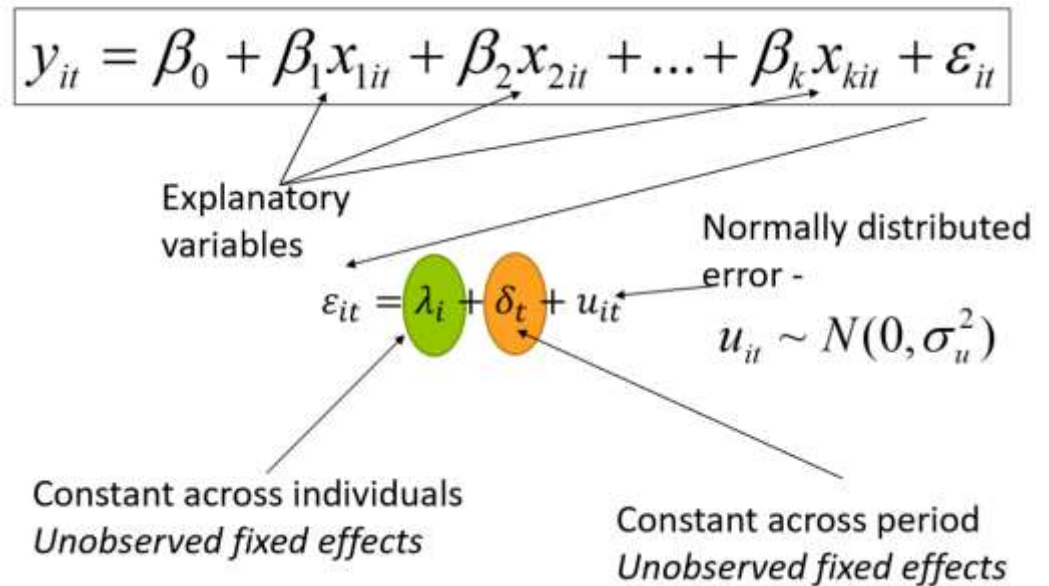
3. Informasi yang hilang dalam periode atau individu tertentu dalam proses pengambilan data.
4. *Cross-section dependence* - Cross-individual section dependence may lead to misleading inference

MODEL DAN CARA ESTIMASI

The One-Way Error Component Model



The Two-Way Error Component Model



Unobserved fixed effects?

Dalam konteks individu adalah sifat konstan/tetap individu yang biasa disebut **unobserved individual heterogeneity**.

Ex. Variasi regulasi antar negara, perbedaan tata kelola perusahaan, dll.

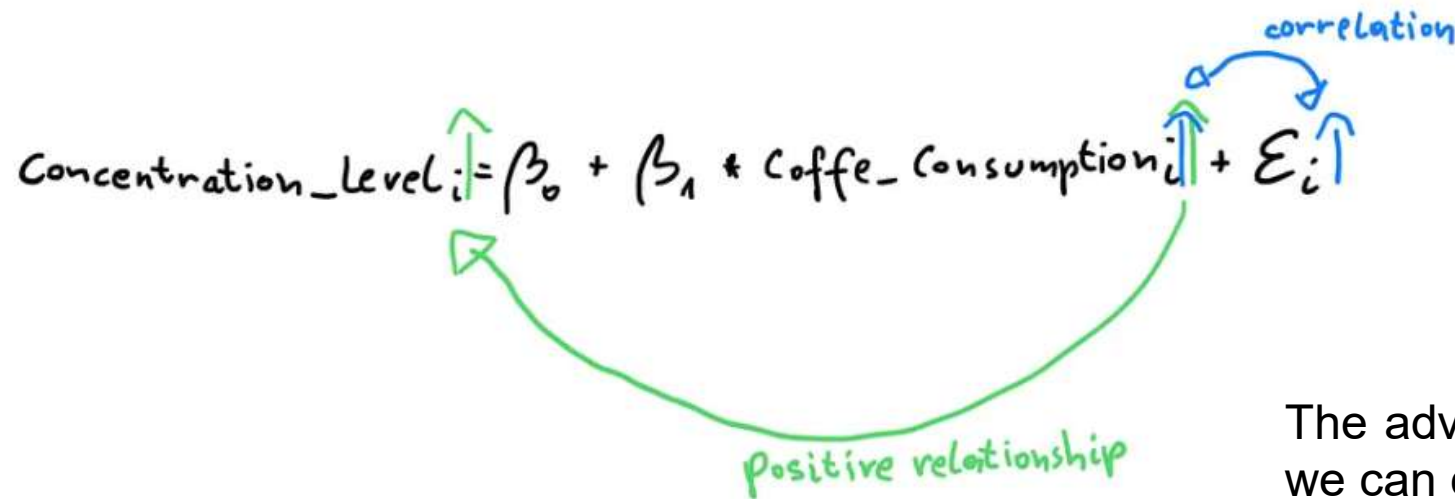
The One or Two-Way Error Component Model merupakan fokus dari prosedur estimasi persamaan regresi data panel

Memilih *one or two-way*
-> **LM Test**
Menguji adanya efek individu, waktu atau keduanya

MODEL DAN CARA ESTIMASI

Cara Estimasi

Peneliti fokus pada hubungan ***unobserved fixed effects (heterogeneity)*** dengan **variabel bebas**, biasa disebut masalah ***endogeneity***.


$$\text{Concentration_Level}_i = \beta_0 + \beta_1 * \text{Coffe_Consumption}_i + \epsilon_i$$

The advantage of panel data is that we can control **heterogeneity** in our regression model by acknowledge heterogeneity as **fix** or **random**

Unobserved fixed effects dalam komponen error dapat berpengaruh atau memiliki hubungan dengan variabel bebas sehingga model **over** atau **under estimate (bias)**.

MODEL DAN CARA ESTIMASI

■ Estimasi dengan Pendekatan Fixed-Effects (FE)

Model Fixed-Effects

$$Pfit_{it} = \beta_0 + \beta_1 Raw_{it} + \beta_2 Labor_{it} + \lambda_i + u_{it}$$

$$Pfit_{it} = (\beta_0 + \lambda_i) + \beta_1 Raw_{it} + \beta_2 Labor_{it} + u_{it}$$

$$u_{it} \sim N(0, \sigma_u^2)$$

Cara estimasi model FE dapat dengan beberapa langkah, fokusnya yaitu menghilangkan λ_i atau λ_t . **Langkah pertama** adalah merubah bentuk model"

1. Within Estimator

$$(y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}_{it}) = \beta' (x_{it} - \bar{x}_i - \bar{x}_t + \bar{x}_{it}) + (u_{it} - \bar{u}_i - \bar{u}_t + \bar{u}_{it})$$

$$\tilde{y}_{it} = \beta_1 \tilde{x}_{1it} + \dots + \beta_k \tilde{x}_{kit} + \tilde{u}_{it}$$

λ_i diamsusikan sebagai parameter yang fiks/tetap, sehingga dapat diletakan sebagai konstanta atau intersep.

2. Least square dummy variable (LSDV)

$$y_{it} = \sum_{j=1}^N \alpha_j \mathcal{D}_{jit} + \sum_{\tau=1}^N \lambda_{\tau} \mathcal{B}_{\tau it} + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + u_{i,t},$$

where $u_{i,t} \sim \mathcal{N}(0, \sigma_u^2)$.

Langkah kedua

Baik within / LSDV, estimasinya bisa dengan cara OLS (Meminimumkan jumlah kuadrat error) atau ML (Metode kemungkinan maksimum).

MODEL DAN CARA ESTIMASI

■ Estimasi dengan Pendekatan Fixed-Effects (FE)

Within Estimator

$$y_{it} = X_{it}\beta + \alpha_i + u_{it} \quad \text{for } t = 1, \dots, T \text{ and } i = 1, \dots, N$$

$\text{Cov}(X_{it}, \alpha_i) \neq 0$

Ada korelasi antara variabel bebas dengan *unobserved fixed effects*

$$y_{it} - \bar{y}_i = \beta(X_{it} - \bar{X}_i) + \underbrace{(\alpha_i - \bar{\alpha}_i)}_{=0} + (u_{it} - \bar{u}_i)$$

Tujuan estimasi adalah mencari koefisien dugaan (Beta) yang menempel pada variabel bebas. Jika masih terdapat *unobserved fixed effects*, menyebabkan masalah *endogeneity*.

Least square dummy variable (LSDV)

$$y_{it} = \alpha + \gamma_1 D_1 + \gamma_2 D_2 + \gamma_3 D_3 + \beta x_{it} + \varepsilon_{it}$$

Keterangan:

y_{it} = variabel terikat individu ke i periode ke t

α = intersep gabungan

β = koefisien regresi atau slope

x_{it} = variabel penjelas individu ke i periode ke t

γ_i = intersep individu i

D_i = variabel dummy untuk individu ke i

ε_{it} = galat individu ke i period ke t (*idiosyncratic term*)

$i = 1, 2, 3, 4; t = 1, 2, \dots, T$

MODEL DAN CARA ESTIMASI

■ Estimasi dengan Pendekatan Random-Effects (RE)

λ_i diamsuikan sebagai bagian dari **galat/random elements**:

$$Pfit_{it} = \beta_0 + \beta_1 Raw_{it} + \beta_2 Labor_{it} + \lambda_i + u_{it}$$

$Cov(\lambda_i, X_{it}) = 0$

Tidak ada korelasi antara variabel bebas dengan *unobserved fixed effects*

• **GLS:** $\Rightarrow \mathbf{b}_{GLS} = (\mathbf{X}'(\Omega)^{-1} \mathbf{X})^{-1} \cdot \mathbf{X}'(\Omega)^{-1} \mathbf{Y}$

Beda dengan **OLS** $\hat{\beta} = (X'X)^{-1}X'y$

Cara estimasi model RE berbeda dengan FE. Karena *unobserved fixed effects* tidak memberikan masalah *endogeneity* maka model RE bentuknya tidak diubah. Estimasi dengan OLS atau ML atau seringkali dengan GLS agar hasil lebih efisien.

PEMILIHAN MODEL TERBAIK

■ OLS VS FEM VS REM



OLS VS FEM

Chow Test

$$F_1 = \frac{(RRSS - URSS)/(N + T - 2)}{URSS/(N - 1)(T - 1) - K} \stackrel{H_0}{\sim} F_{(N+T-2), (N-1)(T-1)-K}$$

RRSS: residual sums of squares OLS

URSS: residual sums of squares FEM

Hipotesis

H0: OLS

H1: FEM



FEM VS REM

Hausman Test

$$H = \left(\hat{\beta}_{RE} - \hat{\beta}_{FE} \right)' \left[Var(\hat{\beta}_{RE}) - Var(\hat{\beta}_{FE}) \right]^{-1} \left(\hat{\beta}_{RE} - \hat{\beta}_{FE} \right)$$

Untuk menguji apakah model REM atau FEM yang tepat, maka diajukan hipotesis $H_0 : E(\lambda_{it}|x_{it}) = 0$ atau **H₀ : REM**

Jika tidak ada korelasi antara variabel independen dan efek individu, maka penduga FEM dan penduga REM **keduanya konsisten**, tetapi FE tidak efisien

Jika terdapat korelasi antara variabel independen dan efek individu, maka penduga FEM konsisten tetapi penduga REM tidak konsisten

Hipotesis

H0: REM

H1: FEM

Catatan:

- Efek individu = *unobserved fixed effects*
- Ada tidaknya korelasi λ_i dan x_{it} dapat didekati dengan melihat persamaan hausman yang mengacu pada kesamaan penduga FEM dan REM.

DIAGNOSTIC MODEL

■ Pengujian Residual Hasil Estimasi

Kondisi *Gauss-Markov* $u \sim N(0, \sigma^2 I_{NT})$

$V\{\mu_i\} = \sigma^2, i = 1, \dots, N$: homoskedastisitas

$Cov\{\mu_i, \mu_j\} = 0, i = 1, \dots, N, i \neq j$: autokorelasi

Hipotesis

H0: Homoskedastisitas
H1: Heteroskedastisitas

Hipotesis

H0: Tidak Terdapat Autokorelasi
H1: Terdapat Autokorelasi

■ Prasyarat

Multikolinearitas atau hubungan antar variabel bebas (independen).

Pengujian: Indikator VIF atau Korelasi

Multikolinearitas yang kuat dapat menyebabkan bias pada koefisien dugaan.

Tahapan Analisis

■ Tahapan Analisis

1. Mempersiapkan data
2. Eksplorasi data
3. Menguji adanya efek individu dan efek waktu
4. Menentukan dan memilih model terbaik antara PLS
vs Fem vs REM
5. Diagnostic Model
6. Interpretasi dan Visualisasi

CONTOH KASUS

■ Perbankan

Unobserved fixed effects

- **Risiko operasional:** Risiko operasional berkaitan dengan risiko yang timbul dari kegagalan sistem, proses, atau personel dalam operasional bank. Faktor-faktor unobserved heterogeneity yang mempengaruhi risiko operasional antara bank dapat termasuk sistem teknologi yang berbeda, proses pengendalian risiko yang berbeda, atau staf yang memiliki tingkat keahlian yang berbeda.
- **Karakteristik pasar:** Karakteristik pasar seperti tingkat persaingan dan regulasi dapat mempengaruhi strategi dan keputusan yang diambil oleh bank. Faktor-faktor unobserved heterogeneity seperti preferensi lokal, akses ke sumber daya, dan kemampuan untuk memahami lingkungan pasar dapat membedakan strategi bank.
- **Kapasitas manajemen risiko:** Kemampuan bank untuk mengelola risiko dapat dipengaruhi oleh banyak faktor, termasuk sumber daya manusia, sistem teknologi, dan kebijakan internal. Faktor-faktor unobserved heterogeneity seperti keterampilan manajemen risiko, budaya risiko, dan kebijakan pengendalian risiko dapat membedakan kemampuan bank untuk mengelola risiko.

Terima Kasih

