KAJIAN ANALISIS REGRESI DENGAN DATA PANEL

I Gede Nyoman Mindra Jaya
Neneng Sunengsih

Staf Pengajar Jurusan Statistika Universitas Padjadjaran

Abstrak

Data set yang merupakan kombinasi data *cross section* dan *time series* sering sekali ditemukan dalam kajian ekonomi, disebut sebagai data panel. Dalam pemodelan data panel kita dihadapkan pada komponen stokastik yang relatif kompleks. Analisis regresi data panel adalah analisis regresi dengan struktur data merupakan data panel. Umumnya pendugaan parameter dalam analisis regresi dengan data cross section dilakukan menggunakan pendugaan metode kuadrat terkecil (MKT). Metode ini akan memberikan hasil pendugaan yang bersifat *Best Linear Unbiased Estimation* (BLUE) jika semua asumsi Gauss Markov terpenuhi diantaranya adalah *non-autcorrelation*. Kondisi terakhir ini tentunya sulit terpenuhi pada saat kita berhadapan dengan data panel. Sehingga pendugaan paramter tidak lagi bersifat BLUE. Jika data panel dianalisis dengan pendekatan model-model time series seperti fungsi transfer, maka ada informasi keragaman dari unit *cross section* yang diabaikan dalam pemodelan. Salan satu keuntungan dari analisis regresi data panel adalah mempertimbangkan keragamaan yang terjadi dalam unit *cross section*. Dalam penelitian ini penulis menerapkan analisis regresi data panel untuk melakukan pemodelan penyerapan tenaga kerja industri kecil di wilayah Jawa Barat dan mengetahui faktor-faktor apa saja yang mempengaruhi penyerapan tenaga kerja di Jawa Barat.

Kata Kunci : Regresi Data Panel, *Fixed Model*, *Random Model*

PENDAHULUAN

Data set yang merupakan kombinasi data cross section dan time series sering sekali kita temukan dalam kajian ekonomi, disebut sebagai data panel. Dalam pemodelan data panel kita dihadapkan pada komponen stokastik yang relatif kompleks.

Analisis regresi data panel adalah analisis regresi dengan struktur data merupakan data panel. Umumnya pendugaan parameter dalam analisis regresi dengan data cross section dilakukan menggunakan pendugaan metode kuadrat terkecil (MKT). Metode ini akan memberikan hasil pendugaan yang bersifat Best Linear Unbiased Estimation (BLUE) jika semua asumsi Gauss Markov terpenuhi diantaranya adalah *non-autcorrelation*. Kondisi terakhir ini tentunya sulit terpenuhi pada saat kita berhadapan dengan data panel. Sehingga pendugaan paramter tidak lagi bersifat BLUE. Jika data panel dianalisis dengan pendekatan model-model time series seperti fungsi transfer, maka ada informasi keragaman dari unit cross section yang diabaikan dalam pemodelan. Salan satu keuntungan dari analisis regresi data panel adalah mempertimbangkan keragamaan yang terjadi dalam unit *cross section*.

Dalam penelitian ini penulis menerapkan analisis regresi data panel untuk melakukan pemodelan penyerapan tenaga kerja industri kecil di wilayah Jawa Barat dan mengetahui faktor-faktor apa saja yang mempengaruhi penyerapan tenaga kerja di Jawa Barat.

Analisis Regresi Data Panel

Data panel adalah gabungan antara data *cross section* dan data *time series*, dimana unit *cross section* yang sama diukur pada waktu yang berbeda. Analissi regresi data panel adalah analissi regresi yang didasarkan pada data panel untuk mengamati hubungan antara satu variabel terikat (dependent variabel) dengan satu atau lebih variabel bebas independen variabel.
Beberapa alternatif model yang dapat diselesaikan dengan data panel yaitu :
Model 1 : semua koefisien baik intersep maupun slop koefisien konstan

$$Y_{it} = \beta_1 + \sum_{K=2}^{K} \beta_k X_{kit} + \varepsilon_{it} \tag{1}$$

Model 2 : slop koefisien konstan, tetapi intersep berbeda akibat perbedaan unit *cross section.*

$$Y_{it} = \beta_{1i} + \sum_{K=2}^{K} \beta_k X_{kit} + \varepsilon_{it} \tag{2}$$

Model 3 : slop koefisien konstan, tetapi intersep berbeda akibat perbedaan unit *cross section* dan berubahnya waktu.

$$Y_{it} = \beta_{1it} + \sum_{K=2}^{K} \beta_k X_{kit} + \varepsilon_{it} \tag{3}$$

Model 4 : intersep dan slope koefisien berbeda akibat perbedaan unit cross section.

$$Y_{it} = \beta_{1i} + \sum_{K=2}^{K} \beta_{ki} X_{kit} + \varepsilon_{it} \tag{4}$$

Model 5 : intersep dan slope koefisien berbeda akibat perbedaan unit cross section dan berubahanya waktu.

$$Y_{it} = \beta_{1it} + \sum_{K=2}^{K} \beta_{kit} X_{kit} + \varepsilon_{it} \tag{5}$$

dengan *i* =1,2,…,N
t = 1,2,…,T
N : Banyak unit cross section
T : Banyak data time series
Y_{it} : Nilai variabel terikat cross section ke-i time series ket-t
X_{kit} : Nilai vairabel bebas ke-k untuk cross section ke-i tahun ke-t
β_{kit} : paramter yang ditaksir
ε_{it} : unsur gangguan populasi
K :Banyak parameter regresi yang akan ditaksir

Model dengan Slope Koefisien Konstan dan Intersep Bervariasi Akibat Perbedaan Unit Coress Section

Model ini memperhatikan efek penggabungan unit cross section yang berbeda sedangkan berubanya waktu (tahun) hanya sebagai transisi.

$$Y_{it} = \beta_{1i} + \sum_{K=2}^{K} \beta_{kit} X_{kit} + \varepsilon_{it} \quad ; i=1,2,\ldots,N \; ; t=1,2,..,T \tag{6}$$

Persamaan di atas dapat ditulis sebagai berikut :

$$Y_{it} = (\alpha + \mu_i) + \sum_{K=2}^{K} \beta_k X_{kit} + \varepsilon_{it} \tag{7}$$

Dengan $\beta_{1i} = (\alpha + \mu_i)$

$$\alpha = \frac{\sum_{i=1}^{N} \beta_{1i}}{N}$$

Y_{it} : nilai variabel terikat wilaya ke-*i* tahun ke-*t*
β_{it} : intersep wilayah ke-*i*
α : rata-rata instersep (konstanta)
μ_i : perbedaan rata-rata intersep dengan intersep wilayah ke-*i*
μ_i : perbedaan rata-rata intersep dengan intersep wilayah ke-*i*
β_k : *slope coefficient* variabel ke-*k*
X_{kit} : nilai variabel bebas ke-k untuk wilayah ke-i tahun ke-t

ε_{it} : unsur gangguan populasi

Prosedur pendugaan persamaan di atas tergantung apakah μ_i diasumsikan *fixed* atau *random*. Jika μ_i diasumsikan *fixed*, μ_i merupakan bagian dari intersep maka persamaan di atas adalah *Fixed Effect Model* (FEM). Jika μ_i diasumsikan random, μ_i sebagai bagian dari unsur gangguan atau error maka persamaan di atas adalah *Random Effect Model* (REM)

Fixed Effect Model dan Random Effect Model

Fixed Effect Model (FEM) : $$Y_{it} = (\alpha + \mu_i) + \sum_{K=2}^{K} \beta_k X_{kit} + \varepsilon_{it} \quad (8)$$

Random Effect Model (REM) : $$Y_{it} = \alpha + \sum_{K=2}^{K} \beta_k X_{kit} + (\mu_i + \varepsilon_{it}) \quad (9)$$

Pada *Fixed Effect Model* tidak lengkapnya inormasi karena perbedaan unit cross section dicerminkan dengan perbedan intersep untuk masing-masing unit cross section. Sedangkan Random Effects Model tidak lengkapnya informasi pada model dipandang sebagai error.
Fixed Effect Model ditaksir dengan metode Least Square Dummy Variable (LSDV). Sedangkan Random Effects Model ditaksir dengan metode Generalized Least Square (GLS). Perbandingan Fixed Effects Model dan Random Effect Model.

Least Square Dummy Variable (LSVD)

Least Square Dummy Variable (LSVD) merupakan metode pendugaan pamater regresi data panel untuk *Fixed Effect Model* . Penyertaan variabel dummy dalam analisis regresi data panel untuk kasus *fixed effect* diharapkan mampu mewakili ketidaklengkapan informasi dalam pembuatan model.
Secara umum model data panel dengan pendekatan variabel dummy dapat ditulis sebagai berikut :

$$Y_{it} = \alpha + \sum_{k=1}^{K} \beta_k X_{kit} + \sum_{i=2}^{N} D_i + \varepsilon_{it} \ ; \quad (10)$$

Dengan D_i adalah variabel Dummy ke-i , i=2,3,…,N dan N adalah banyaknya unit cross section. Pendugaan parameter regresi data panel dilakukan dengan menggunakan *Ordinary Least Square*.

Uji Spesifikasi *Fixed Effect Model*s

Hipotesis nol menyatakan bahwa intersep setiap unit cross section adalah sama, artinya efek unit cross section secara keseluruhan tidak berarti dalam model dugaan. Sebagai alternatifnya bahwa efek unit cross section berarti dalam model dugaan. Jika H_0 diterima maka pooled model lebih baik daripada *Fixed Effect Model*. Sebaliknya jika H_0 ditolak maka *Fixed Effect Model* lebih baik dibandingkan pooled model.
Hipotesis statistik :
H_0 : $\mu_1=\mu_2=\ldots=\mu_I=0$ (efek unit cross section secara keseluruhan tidak berarti)
H_0 : sekurang-kurangnya ada sebuah $\mu_i \neq 0$ (efek wilayah berarti)
Statistik uji

$$F = \frac{(e`e_{pooled} - e`e_{FEM})/(N-1)}{e`e_{FEM}/(NT-N-K)} = \frac{(R^2_{FEM} - R^2_{pooled})/(N-1)}{(1-R^2_{FEM})/(NT-N-K)} \sim F_{(N-1),(NT-N-k)} \quad (11)$$

Tolak hipotesis nol jika F hitung lebih besar dari F tabel

Generalized Least Square (GLS)

Untuk Random Effect Model (REM), pendugaan paramternya dilakukan menggunakan Generalized Least Square jika matriks Ω diketahui, namun jika tidak diketahui dilakukan dengan FGLS yaitu menduga elemen matriks Ω. Pada REM ketidklengkapan informasi untuk setiap unit cross section dipandang sebagai error sehingga μ_i adalah bagian dari unsur gangguan.
Model REM dapat dituliskan dapat dituliskan sebagai berikut :

$$Y_{it} = \alpha + \sum_{k=1}^{K} \beta_k X_{kit} + (\mu_i + e_{it}) \tag{12}$$

Asumsi :

$$\mu_i \sim N(0,\sigma_\mu^2) \qquad E(\mu_i \mu_j) = 0;\ i \neq j$$

$$e_{it} \sim N(0,\sigma_e^2) \qquad E(\mu_i e_{it}) = 0$$

$$E(e_{it}e_{is}) = E(e_{it}e_{jt}) = E(e_{it}e_{js}) = 0 \ \ ; i \neq j; t \neq s$$

Untuk data cross section ke-i persamaan di atas dapat ditulis **$y_i = X_i\beta + (\mu_i 1 + e_i)$**.
Varians komponen dari unsur gangguan **($\mu_i 1 + e_i$)** untuk unit cross section ke-I adalah :

$$\underset{TxT}{\Omega} = \begin{bmatrix} \sigma_\mu^2 + \sigma_e^2 & \sigma_\mu^2 & \ldots & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\mu^2 + \sigma_e^2 & \cdots & \sigma_\mu^2 \\ \cdots & \cdots & \ddots & \cdots \\ \sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 + \sigma_e^2 \end{bmatrix} \tag{13}$$

Varians komponen Ω identik untuk setiap unit cross section. Sehingga varians komponen untuk seluruh observasi dapat ditulisakan :

$$\underset{NTxNT}{W} = \begin{bmatrix} \Omega & 0 & \cdots & 0 \\ 0 & \Omega & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & \Omega \end{bmatrix} \tag{14}$$

Jika nilai Ω diketahui maka persamaan… dapat diduga menggunakan Generalized Least Square (GLS) dengan $\hat{\beta} = (X`W^{-1}X)^{-1}(X`W^{-1}y)$. Jika Ω tidak diketahui maka Ω perlu diduga dengan menduga $\hat{\sigma}_\mu^2$ dan $\hat{\sigma}_e^2$, sehingga persamaan di atas diduga dengan $\hat{\beta} = (X`\hat{W}^{-1}X)^{-1}(X`\hat{W}^{-1}y)$, Dimana $\hat{\sigma}_e^2 = \dfrac{\hat{e}`\hat{e}}{NT - N - k}$ dengan $\hat{e} = y - X\hat{\beta}$ adalah residu dari Least Square Dummy Variables (LSDV). Sedangkan $\hat{\sigma}_\mu^2 = \dfrac{\hat{\sigma}_1^2 - \hat{\sigma}_e^2}{T}$

Uji Spesifikasi random Effect Model

Hipotesis nol menyatakan bahwa semua varians error cross section adalah no. artinya efek random dari unit cross section tidak bearti. Jika H_0 diterima maka pooled model lebih baik dair Random effect model. Sedangkan jika H0 ditolak maka Random Effect model lebih sesuai.
Hipotesis statistik yang digunakan adalah :

H_0 : $\sigma_\mu^2 = 0$

H_1 : $\sigma_\mu^2 \neq 0$

Statistik uji :

$$LM = \frac{NT}{2(T-1)} \left[\frac{T^2 \bar{e}`\bar{e}}{e`e} - 1 \right]^2 \sim \chi^2_{(1)} \tag{15}$$

Dimana $\bar{e}$ adalah vektor (Nx1) dari rata-rata residu pooled model dan e`e adalah jumlah kuadrat residu dari pooled model.
Kriteria pengujian

Tolak hipotesi nol (H_0) jika χ^2_{hitung} lebih besar dari χ^2_{tabel}

Uji Hausman

Uji Hausman digunakan untuk memilih model terbaik apakah *Fixed Effect Model* (FEM) atau *Random Effect Model* (REM). Jika H_0 diterima maka Random Effect Model (REM) lebih efisien, sedangkan jika H_0 ditolak maka *Fixed Effect Model* lebih sesuai daripada Random Effect Model.

Hipotesis statisik :

H_0 : $\rho_{\mu_i X_i} = 0$ Effect Cross section tidak berhubungan dengan regresi lain (REM)

H_0 : $\rho_{\mu_i X_i} \neq 0$ Effect Cross section berhubungan dengan regresi lain (FEM)

Statistik uji :

Statistik uji yang digunakan untuk menguji hipotesis di atas adalah statisti uji *m* dengan rumus sebagai berikut :

$$m = (\mathbf{b}_{FEM} - \mathbf{b}_{REM})`\hat{\boldsymbol{\Sigma}}(\mathbf{b}_{FEM} - \mathbf{b}_{REM}) \sim \chi^2_k \qquad (16)$$

Dengan $\hat{\Sigma} = \text{var}(b_{FEM}) - \text{var}(b_{REM})$

Kriteria pengujian

Tolak hipotesis nol jika m > χ^2_k

Tabel 1 Fixed Effects Model dan Random Effects Model

	Fixed Effect Model	Random Effects Model
Model	$Y_{it} = (\alpha + \mu_i) + \sum_{k=1}^{K} \beta_k X_{kit} + \varepsilon_{it}$	$Y_{it} = \alpha + \sum_{k=1}^{K} \beta_k X_{kit} + (\mu_i + \varepsilon_{it})$
Intersep	Berbeda untuk tiap unit cross section	Konstan
Varians Error	Konstan	Berbeda untuk tiap unit cross section
Slopes	Konstan	Konstan
Metode	LSDV	GLS-FGLS
Hipotesis	Uji F	Uji Lagrange Multiplier (LM)

Contoh Aplikasi

Analisis regresi data panel diterapkan dalam pemodelan penyerapan tenanga kerja di Jawa Barat. Data panel terdiri dari penggabungan data cross setion dan time series yang diambil pada periode waktu dengan objek yang sama. Data cross section yang digunakan adalah 22 wilayah di Jawa Barat sedangkan data time series digunakan sampel pada tahun 2001 sampai 2004 untuk masing-masing wilayah. Sehingga keselurhan data yang dugunakan adalah sebanyak 88 data cross section-time series.

Variabel dependen dalam penelitian ini adalah jumlah penerapan tenaga kerja industri kecil di wilayah Jawa Barat. Sedangkan variabel bebasnya adalah nilai produksi, modal dan upah.

Untuk mengetahui faktor-faktor yang mempengaruhi penyerapan tenanga kerja, digunakan fungsi permintaan tenanga kerja dengan pendekatan maksimasi keuntungan (Gujarati, 1978 : 350). Melalui pendekatan fungsi Cobb Duoglas, permintaan terhadap tenaga kerja merupakan fungsi dari nilai produksi (PQ) dan Upah (W) atau dalam notasi

L=*f*(PQ,W)

Selanjutnya menurut Arsyad (1987 :287) permintaan akan input tenanga kerja juga dipengaruhi oleh input nilai model kerja RK. Sehinga fungsi permintaan tenanga kerja menjadi :

L=*f*(PQ, W, RK)

Secara lebih lengkap fungsi penyerapan tenanga kerja dituliskan sebagai berikut :

$$L_{it} = \alpha (PQ)_{it}^{\beta_1} (RK)_{it}^{\beta_2} (W)_{it}^{\beta_3} e^{\varepsilon_{it}}$$

$$\ln L_{it} = \ln \alpha + \beta_1 \ln (PQ)_{it} + \beta_2 \ln (RK)_{it} + \beta_3 \ln (W)_{it} + \varepsilon_{it}$$

Misalkan :

$$\ln L_{it} = Y_{it} \quad \ln (PQ)_{it} = X_{1it} \quad \ln (RK)_{it} = X_{2it} \quad \ln (W)_{it} = X_{3it} \quad \ln \alpha_{1i} = \beta_{0i}$$

Sehingga model regresi sampelnya dapat ditulis :

$$Y_{it} = \beta_{0i} + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + e_{it}$$

HASIL ANALISIS

Setelah dilakukan pengolahan dengan software SAS 9.1 diperoleh hasil sebagai berikut :

1. *Fixed Effect Model* (FEM)

Memandang bahwa ketidaklengkapan informasi dari setiap wilayah sebagai bagian dari intersep, melalui metode Least Square Dummy Variables (LSVD) diperoleh model taksiran *Fixed Effect Model* sebagai berikut :

$$\begin{aligned} Y_{it} = {} & 1.208 + 0.267 X_{1it} + 0.338 X_{2it} + 0.116 X_{3it} + 0.538 D_2 + 1.077 D_3 + 0.967 D_4 + \\ & 1.392 D_5 + 1.946 D_6 + 1.145 D_7 + 0.960 D_8 \text{ -}0.066 D_9 + 1.262 D_{10} + 1.944 D_{11} + \\ & 0.663 D_{12} \text{-}0.320 D_{13} + 0.768 D_{14} + 1.422 D_{15} \text{-}0.094 D_{16} + 1.467 D_{17} + 0.834 D_{18} + \\ & 0.292 D_{19} \text{-}0.312 D_{20} + 0.258 D_{21} + 0.758 D_{22} \end{aligned}$$

2. Random Effect Model (REM)

Memandang bahwa ketidaklengkapan informasi dari setiap wilayah sebagai bagian dari unsur gangguan, melalui metode Generalized Least Square (GLS) diperoleh nilai taksiran Random Effect Mode sebagai beriktu :

$$Y_{it} = 2.705 + 0.284 X_{1it} + 0.375 X_{2it} + 0.021 X_{3it}$$

3. Uji Hausman

Untuk menentukan apakah model terbaik untuk data adalah *fixed effect modes* atau *Random Effect Model* dapat dilakukan dengan uji Hausman dengan hasil sebagai berikut :

Hasil perhitungan denga SAS 9.0 memberikan nilai *m* dan *p-value* sebagai berikut :

Tabel 2. Uji Hausman

Hausman Test for Random Effects		
DF	m Value	Pr > m
3	7.4	0.0601

Hasil uji Hausman menunjukan bahwa hipotesis nol diterima dimana terihat nilai p-value nya lebih besar diabandingkan dengan 0,05 sehingga dapat disimpulkan bahwa model yang lebih tepat untuk data adalah *Random Effect Model*.

4. Uji Spesifikasi Model

Untuk menguji apakah Random Effect Model lebih baik dibandingkan dengan Pooled Model, maka dilakukan pengujian dengan menggunakna Lagrange Multiplier (LM). Dari haisl perhitungan diperoleh nilai LM = 79.94522. Sedangkan nilai $\chi^2_{(1;0,05)}$ = 3.841459. Nilai LM lebih besar dibandingkan dengan nilai Chis-Square pada derajat bebas 1 dan tingkat signifikansi 5%. Artinya model yang lebih tepat adalah model REM.

5. Interpretasi Hasil

Dari hasil pengujian Hausman diperoleh kesimpulan bahwa Random Effect Model lebih baik dibandingkan dengan *Fixed Effect Model* . Uji spesifikasi model juga menunjukkan model terbaik adalah REM.

Setelah diketahui model terbaik, selanjutnya dilakukan pengujian signfikansi model baik secara parsial ataupun overall. Hasil pengujian ditujukkan pada tabel-tabel di bawah ini :

Tabel 4. Uji Overall

Fit Statistics			
SSE	3.7602	DFE	84
MSE	0.0448	Root MSE	0.2116
R-Square	0.6028		

Tabel 5. Dugaan Parameter

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.705	1.7568	1.54	0.1274
Produksi (x_1)	1	0.284	0.0788	3.6	0.0005
Modal (x_2)	1	0.375	0.0638	5.87	<.0001
Upah (x_3)	1	0.021	0.1614	0.13	0.8956

Hasil pengujian menunjukkan bahwa faktor produksi dan modal berpengaruh signifikan terhadap penyerapan tenanga kerja di Jawa Barat. Sedangkan upah dinyatakan tidak berpengaruh signifikan pada taraf signifikansi 5%.

KESIMPULAN

1. Untuk analisis regresi dengan data panel terdapat dua pendekatan yaitu *Fixed Effect Model* dan Random Effect Model. Untuk mengetahui apakah *Fixed Effect Model* yang lebih baik dibandingkan dengan Random Effect Model dapat dilakukan dengan uji Hausman.
2. Untuk pemodelan penyerapan tenanga kerja di Jawa Barat, model yang dinyatakan lebih tepat adalah model REM. Model ini mampu menjelaskan variansi data mencapai 60,28%. Penyerapan tenaga kerja secara signifikan dipengaurhi oleh produksi dan modal. Sedangkan upah dinyatakan tidak berpengaruh.

DAFTAR PUSTAKA

Myoung Park, Hun Linear Regression Models for Panel Data Using SAS, STATA, LIMDEP, and SPSS, 2005 The Trustees of Indiana University

Wooldridge, Jeffrey M. Econometric Analysis of Cross Section and Panel Data, The MIT Press, 2008, Cambridge, Massachusetts, London, England

LAMPIRAN

Program SAS

```
Data Panels;
Input Kab Thn y x1 x2 x3 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13 d14 d15 d16 d17 d18 d19 d20 d21 d22;
cards;
1	2001	9.4926	11.3125	9.1653	12.9408	1	0	0	0	0…….0
………………………………………………………………………………………………………………………………
22	2004	8.7058	9.9735	9.3580	13.3723	0	0	0	0	0……..0
;
;
Proc Reg Data=Panels;
Model y=x1 x2 x3 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13 d14 d15 d16 d17 d18 d19 d20 d21 d22;
Run;
Proc Sort Data=Panels;
by Kab Thn;
run;
Proc Tscsreg Data=Panels;
model y=x1 x2 x3/FULLER;
Id Kab Thn;
Run;
```