DATA PANEL

****

ANGELINA IKA RAHUTAMI

Program Studi Manajemen

Fakultas Ekonomi

Universitas Katolik Soegijapranata

2011

**Bab IV. Data Panel**

* Panel data adalah data yang memiliki dimensi waktu dan ruang.
* Dalam panel data, data *cross-section* yang sama diobservasi menurut waktu.
* *Balanced panel* 🡪jika setiap *cross-section* unit memiliki jumlah observasi *time-series* yang sama.
* *Unbalanced panel* 🡪 jika jumlah observasi berbeda untuk setiap *cross-section* unit

**ADVANTAGE OF PANEL DATA (Baltagi, 1995)**

1. Panel data estimation dapat mencakup masalah heterogenitas
2. Panel data memberikan informasi lebih banyak, lebih bervariasi, mempersedikit kolinieritas antar variabel, dan lebih efisien
3. Panel data lebih baik digunakan untuk melihat perubahan yang bersifat dinamik.
4. Panel data dapat mendeteksi dan mengukur efek lebih baik.
5. Panel data memungkinkan kita untuk meneliti model yang lebih kompleks 🡪 behavioral models
6. Panel data dapat meminimalkan bias
7. Menghindari masalah multikolinieritas

Kesulitan utama

* Faktor pengganggu akan berpotensi mengandung gangguan yang disebabkan penggunaan observasi runtun waktu, observasi lintas sektoral, serta gabungan keduanya
	+ Penggunaan observasi lintas sektoral mempunyai potensi tidak konsistennya parameter regresi.
	+ Penggunaan observasi runtun waktu mempunyai potensi autokorelasi antar observasi.

**ESTIMASI DATA PANEL**

Terdapat 3 cara estimasi (Pindyck dan Rubinfeld, 1998:

1. Regresi Penggabungan semua data (Pooled OLS)
2. Covariance Model (Fixed Effects 🡪 Least Squares Dummy Variable (LSDV) Model
3. Error Component Model (Random Effects)

REGRESI PENGGABUNGAN (POOLED OLS)

Model Dasar



Untuk i=1,2,….,N dan t=1,2,….,T

N= jumlah unit lintas sektoral

T= jumlah periode waktu

dimana i adalah indeks unit *cross-section* dan t adalah indeks waktu

* Jika μi=0 🡪 berarti tidak ada *individual specific effects* 🡪 pooled OLS akan menghasilkan estimator yang *unbiased*, *consistent,* dan *efficient*
* Jika μi≠0 🡪 berarti ada *individual specific effects* 🡪 pooled OLS akan menghasilkan estimator yang *unbiased*, *consistent,* tetapi *inefficient*
* Inefisiensi juga disebabkan oleh estimasi pooled OLS mengabaikan adanya positive serial correlation pada error

 

akibatnya estimator akan *inefficient*, dan standar error akan *biased* dan *inconsistent*

Jika  estimasi dengan pooled OLS 🡪 *biased* dan *inconsistent*

**Eview pooled**

* Jika data belum tersedia, maka langkah pertama adalah membuat *workfile* untuk data panel.
* FILE
* NEW
* WORKFILE.... dan tuliskan frekuensi data panel.



* Setelah muncul tampilan WORKFILE,
* klik OBJECT
* NEW OBJECT
* POOL.....dan tuliskan identifier di edit window.
* GE
* GM identifier perusahaan
* US
* WEST
* NAME
* POOL01
* OK
* Ingat untuk selalu bekerja di workfile pool selama melakukan estimasi dengan data panel

* Jika anda akan memasukkan data panel,
	+ VIEW
	+ SPREADSHEET(stacked data)......
	+ X2? (pada tampilan series list tuliskan nama variabel dengan disertai (?) untuk menunjukkan identifier)
	+ OK
	+ Kemudian akan muncul tampilan spreadsheet yang siap diedit



Jika data harus di-import dari file lain, dari workfile POOL

* PROCS
* IMPORT POOL DATA.....pilih file data pool yang akan diimport.
* Pada kotak dialog import, isi semua kolom sesuai dengan spesifikasi data yang ada di file asalnya



Dengan asumsi data panel yang akan diolah sudah tersedia, maka klik

* FILE
* OPEN
* WORKFILE (pilih nama file yang mengandung data panel)
* OBJECT
* NEW OBJECT
* POOL

Estimasi data panel: POOLED OLS

* Dari toolbar POOL,
	+ PROCS
	+ ESTIMATE dan muncul tampilan
	+ Isi kotak dialog yang muncul
	+ Y? (DEPENDENT VARIABLE)
	+ X2? X3? (REGRESSORS)
	+ COMMON (INTERCEPT)
	+ Pilih Common untuk Pooled OLS yang didasarkan pada asumsi intersep pada model yang akan diestimasi sama.
	+ NO WEIGHTING
	+ OK



|  |
| --- |
| Dependent Variable: Y? |
| Method: Pooled Least Squares |
| Sample: 1935 1954 |
| Included observations: 20 |
| Total panel observations 80 |
| Variable | Coefficient | Std. Error | t-Statistic | Prob.  |
| C | -63.30414 |  29.61420 | -2.137628 |  0.0357 |
| X2? |  0.110096 |  0.013730 |  8.018809 |  0.0000 |
| X3? |  0.303393 |  0.049296 |  6.154553 |  0.0000 |
| R-squared |  0.756528 |  Mean dependent var |  290.9154 |
| Adjusted R-squared |  0.750204 |  S.D. dependent var |  284.8528 |
| S.E. of regression |  142.3682 |  Sum squared resid |  1560690. |
| Log likelihood | -422.5523 |  F-statistic |  119.6292 |
| Durbin-Watson stat |  0.218717 |  Prob(F-statistic) |  0.000000 |

* Jika dalam data panel terdapat autokolerasi antar observasi (runtun waktu maupun lintas sektor) maka digunakan pendekatan GLS. Dengan GLS, korelasi antar observasi harus dikoreksi dengan koefisien autokorelasinya.
	+ Estimasi GLS:
	+ Dari workfile POOL
	+ PROCS
	+ ESTIMATE
	+ Y? (DEPENDENT VARIABLE)
	+ X2? X3? (REGRESSORS)
	+ COMMON (INTERCEPT)
	+ CROSS SECTION WEIGHTS
	+ OK

Hasil estimasi data panel dengan Pooled GLS:

|  |
| --- |
| Dependent Variable: Y? |
| Method: GLS (Cross Section Weights) |
| Sample: 1935 1954 |
| Included observations: 20 |
| Balanced sample |
| Total panel observations 80 |
| Variable | Coefficient | Std. Error | t-Statistic | Prob.  |
| C | -52.07358 |  7.833245 | -6.647765 |  0.0000 |
| X2? |  0.100023 |  0.008168 |  12.24582 |  0.0000 |
| X3? |  0.330065 |  0.035398 |  9.324311 |  0.0000 |
| Weighted Statistics |  |  |  |  |
| R-squared |  0.869942 |  Mean dependent var |  373.4338 |
| Adjusted R-squared |  0.866563 |  S.D. dependent var |  383.8047 |
| S.E. of regression |  140.1999 |  Sum squared resid |  1513513. |
| Log likelihood | -420.3084 |  F-statistic |  257.5208 |
| Durbin-Watson stat |  0.340369 |  Prob(F-statistic) |  0.000000 |
| Unweighted Statistics |  |  |  |  |
| R-squared |  0.754727 |  Mean dependent var |  290.9154 |
| Adjusted R-squared |  0.748356 |  S.D. dependent var |  284.8528 |
| S.E. of regression |  142.8938 |  Sum squared resid |  1572236. |
| Durbin-Watson stat |  0.211474 |  |  |  |

**MODEL COVARIANCE (LSDV Models)**

Pada model ditambahkan variabel dami untuk agar intersep dimungkinkan berubah



Dalam model terdapat (N-1)+(T-1) variabel dami. Jika diestimasi dengan OLS akan diperoleh estimasi yang tidak bias dan konsisten

There are several possibilities

1. The intercept and slope coefficients are constant across time and space
2. The slope coefficients are constant but the intercept varies over individuals
3. The slope coefficients are constant but the intercept varies over individuals and time
4. All coefficients vary over individuals
5. All coefficients vary over individuals and time

Estimasi dengan variabel dami

|  |
| --- |
| **Dependent Variable: Y?** |
| **Method: Pooled Least Squares** |
| **Sample: 1935 1954** |
| **Included observations: 20** |
| **Total panel observations 80** |
| **Variable** | **Coefficient** | **Std. Error** | **t-Statistic** | **Prob.**  |
| **C** | **-245.7924** |  **35.81112** | **-6.863577** |  **0.0000** |
| **X2?** |  **0.107948** |  **0.017509** |  **6.165319** |  **0.0000** |
| **X3?** |  **0.346162** |  **0.026664** |  **12.98212** |  **0.0000** |
| **D2?** |  **161.5722** |  **46.45639** |  **3.477932** |  **0.0008** |
| **D3?** |  **339.6328** |  **23.98633** |  **14.15943** |  **0.0000** |
| **D4?** |  **186.5666** |  **31.50681** |  **5.921468** |  **0.0000** |
| **R-squared** |  **0.934563** |  **Mean dependent var** |  **290.9154** |
| **Adjusted R-squared** |  **0.930141** |  **S.D. dependent var** |  **284.8528** |
| **S.E. of regression** |  **75.28890** |  **Sum squared resid** |  **419462.9** |
| **Log likelihood** | **-382.6675** |  **F-statistic** |  **211.3706** |
| **Durbin-Watson stat** |  **0.807158** |  **Prob(F-statistic)** |  **0.000000** |

Intersep untuk setiap perusahaan yang berbeda adalah

|  |  |
| --- | --- |
| GE | -245,7924 |
| GM | (-245,7924) + 161,5722 = -84,220 |
| US | (-245,7924) + 339,6328 = 93,8404 |
| WEST | (-245,7924) + 186,5666 = -59,2258 |

Perbedaan tersebut disebabkan oleh perbedaan dalam tiap perusahaan, misal ukuran perusahaan, umur perusahaan, gaya manajerial, dll

Membandingkan antara pooled OLS dan regresi dengan variabel dami

Membandingkan antara pooled OLS dan regresi dengan variabel dami

* Model mana yang lebih baik?
* Restricted F test
	+ Estimasi OLS 🡪 persamaan restricted
	+ Estimasi LSDV 🡪 persamaan unrestricted



* Bandingkan dengan F tabel df(3;74) 🡪 model LSDV lebih baik.

Fixed Effect Panel Model

* Intersep regresi berbeda antar individu, yang berarti tiap individu memiliki karakteristik yang berbeda
* Also known as the least squares dummy variable model.

Why use Fixed Effects

* Fixed Effects are generally used when there is a correlation between the individual intercept and the independent variables.
* Generally used when n is relatively small and t is relatively large.

Fixed Effects

Model dasar





Jumlahkan secara series (t) dan dibagi T, dihasilkan rata-rata grup estimator



dimana







Kurangkan kedua persamaan, didapat



Jika 

estimasi dengan fixed effect akan menghasilkan estimator unbiased, consistent, tetapi inefficient relative terhadap estimasi dengan random effects

Estimasi data panel: Fixed Effect

* Dari toolbar POOL,
	+ PROCS
	+ ESTIMATE dan muncul tampilan kotak dialog
	+ Y? (DEPENDENT VARIABLE)
	+ X2? X3? (REGRESSORS)
	+ FIXED EFFECT (INTERCEPT)
	+ NO WEIGHTING
	+ OK

Hasil estimasi data panel dengan Fixed Effect:

|  |
| --- |
| Dependent Variable: Y? |
| Method: Pooled Least Squares |
| Sample: 1935 1954 |
| Included observations: 20 |
| Total panel observations 80 |
| Variable | Coefficient | Std. Error | t-Statistic | Prob.  |
| X2? |  0.107948 |  0.017509 |  6.165319 |  0.0000 |
| X3? |  0.346162 |  0.026664 |  12.98212 |  0.0000 |
| Fixed Effects |  |  |  |  |
| GE--C | -245.7924 |  |  |  |
| GM--C | -84.22017 |  |  |  |
| US--C |  93.84046 |  |  |  |
| WEST--C | -59.22581 |  |  |  |
| R-squared |  0.934563 |  Mean dependent var |  290.9154 |
| Adjusted R-squared |  0.930141 |  S.D. dependent var |  284.8528 |
| S.E. of regression |  75.28890 |  Sum squared resid |  419462.9 |
| Log likelihood | -382.6675 |  F-statistic |  1056.853 |
| Durbin-Watson stat |  0.807158 |  Prob(F-statistic) |  0.000000 |

**Fixed effects dan Specific firm**

|  |
| --- |
| Dependent Variable: Y? |
| Method: Pooled Least Squares |
| Sample: 1935 1954 |
| Included observations: 20 |
| Total panel observations 80 |
| Variable | Coefficient | Std. Error | t-Statistic | Prob.  |
| GE--X2GE |  0.026551 |  0.037881 |  0.700903 |  0.4855 |
| US--X2US |  0.119210 |  0.019083 |  6.246796 |  0.0000 |
| GM--X2GM |  0.171430 |  0.052390 |  3.272221 |  0.0016 |
| WEST--X2WEST |  0.053055 |  0.104485 |  0.507779 |  0.6131 |
| GE--X3GE |  0.151694 |  0.062553 |  2.425046 |  0.0177 |
| US--X3US |  0.371525 |  0.027401 |  13.55858 |  0.0000 |
| GM--X3GM |  0.408709 |  0.102966 |  3.969367 |  0.0002 |
| WEST--X3WEST |  0.091694 |  0.373394 |  0.245568 |  0.8067 |
| Fixed Effects |  |  |  |  |
| GE--C | -9.956306 |  |  |  |
| US--C | -149.4667 |  |  |  |
| GM--C | -50.07804 |  |  |  |
| WEST--C | -0.580403 |  |  |  |
| R-squared |  0.951157 |  Mean dependent var |  290.9154 |
| Adjusted R-squared |  0.943256 |  S.D. dependent var |  284.8528 |
| S.E. of regression |  67.85489 |  Sum squared resid |  313091.5 |
| Log likelihood | -345.5921 |  F-statistic |  189.1733 |
| Durbin-Watson stat |  0.974483 |  Prob(F-statistic) |  0.000000 |

**LSDV beda intersep dan slope**

|  |
| --- |
| Dependent Variable: Y? |
| Method: Pooled Least Squares |
| Sample: 1935 1954 |
| Included observations: 20 |
| Total panel observations 80 |
| Variable | Coefficient | Std. Error | t-Statistic | Prob.  |
| C | -9.956306 |  76.35180 | -0.130400 |  0.8966 |
| D2? | -139.5104 |  109.2808 | -1.276623 |  0.2061 |
| D3? | -40.12174 |  129.2343 | -0.310457 |  0.7572 |
| D4? |  9.375904 |  93.11719 |  0.100689 |  0.9201 |
| X2? |  0.026551 |  0.037881 |  0.700903 |  0.4858 |
| X3? |  0.151694 |  0.062553 |  2.425046 |  0.0180 |
| D2?\*X2? |  0.092658 |  0.042417 |  2.184483 |  0.0324 |
| D2?\*X3? |  0.219831 |  0.068291 |  3.219009 |  0.0020 |
| D3?\*X2? |  0.144879 |  0.064650 |  2.240965 |  0.0283 |
| D3?\*X3? |  0.257015 |  0.120477 |  2.133302 |  0.0365 |
| D4?\*X2? |  0.026504 |  0.111140 |  0.238475 |  0.8122 |
| D4?\*X3? | -0.060000 |  0.378597 | -0.158480 |  0.8745 |
| R-squared |  0.951157 |  Mean dependent var |  290.9154 |
| Adjusted R-squared |  0.943256 |  S.D. dependent var |  284.8528 |
| S.E. of regression |  67.85489 |  Sum squared resid |  313091.5 |
| Log likelihood | -345.5921 |  F-statistic |  120.3830 |
| Durbin-Watson stat |  0.974483 |  Prob(F-statistic) |  0.000000 |

Yang harus diperhatikan jika menggunakan *fixed effect model* dan LSDV

1. Penggunakan variabel dami akan menimbulkan masalah *degree of freedom*.
2. Kemungkinan terjadi multikolineritas.
3. *Fixed Effect* *Model* tidak bisa digunakan untuk mengetahui dampak variabel yang *time invariant*, misal jenis kelamin, ras, dll
4. Hati-hati dengan *error term*. Asumsi Klasik *error term* harus dimodifikasi.

**Random Effects Model**

* Pada dasarnya penyertaan variabel boneka diharapkan dapat mewakili tidak lengkapnya informasi dalam pembuatan model, sehingga wajar jika kekurangan informasi tersebut dianggap tercermin dalam *error term.*
* Model data panel yang di dalamnya melibatkan korelasi antar *error term* karena berubahnya waktu maupun karena berbedanya observasi dapat diatasi dengan pendekatan Model *Error-Components*
* Random efek model mengasumsikan intersep tiap individu adalah random dari populasi yang lebih besar dengan constant mean value
* Diasumsikan bahwa *error component* individual tidak berkorelasi satu sama lain serta tidak terdapat autokorelasi baik karena data *time-series,* maupun karena data *cross-section.*
* Also (less frequently) know as the Error Components model.

Random Effects

Model Dasar



Asumsi

* + E[εit]=E[μi]=0
	+ Var[εit]=σε2
	+ Var[μit]=σμ2
	+ Cov[εitμj]=0 ∀i,t,j
	+ Cov[μiμj]=0 ∀i≠j
	+ Cov[Xitμi]=0
	+ E[ωit2]=σε2+σμ2
	+ E[ωit ωis]= σμ2 ∀t≠s

Implikasi semua asumsi tersebut adalah



Estimasi data panel: Random Effect

* Dari toolbar POOL,
	+ PROCS
	+ ESTIMATE dan muncul tampilan kotak dialog
	+ Y? (DEPENDENT VARIABLE)
	+ X2? X3? (REGRESSORS)
	+ RANDOM EFFECT (INTERCEPT)
	+ NO WEIGHTING
	+ OK