UJI ASUMSI KLASIK

****

ANGELINA IKA RAHUTAMI

Program Studi Manajemen

Fakultas Ekonomi

Universitas Katolik Soegijapranata

2011

**Bab III. Asumsi Klasik**

**Asumsi: Classical Linear Regression Model (CLRM)**

1. Linier Regression Model

2. X’s (regressors) are non stochastic

3. Zero Mean value of disturbances

4. Homoscedasticity = equal variances

5. No Autocorrelation between disturbances

6. Zero Covariances between u and X

7. N > estimated parameters

8. Variability in X’s values

9. Correct Specification

10. No Perfect Multicolinearity

**Asumsi Klasik Multikolinieritas**

Ditemukan oleh Ragnar Frisch tahun 1934. multikolinieritas bila terjadi hubungan linier yang sempurna (*perfect*) atau pasti (*exact*) di antara beberapa atau semua variabel bebas dari suatu model regresi. Akibatnya akan kesulitan untuk dapat melihat pengaruh variabel penjelas terhadap variabel yang dijelaskan (Maddala, 1992: 269-270). Multikolinieritas pada hakekatnya adalah fenomena sampel Multikolinieritas adalah persoalan derajat (*degree*) dan bukan persoalan jenis (*kind*) Masalah multikolinieritas hanya berkaitan dengan adanya hubungan linier di antara variabel-variabel bebas

**Penyebab multiko**

1. Metode pengumpulan data yang dipakai (*the data collection method employed*).
2. Kendala dalam model atau populasi yang menjadi sampel (*constraint on the model or in the population being sampled*), sehingga populasi yang dijadikan sampel tersebut kurang realistis.
3. Spesifikasi model (*specification model*).
4. Model yang berlebihan (*an overdetermined model*) 🡺 jumlah variabel penjelas lebih banyak dibandingkan dengan jumlah data (observasi).
5. Penggunaan nilai kelambanan (*lagged value*)

Efek kolinieritas

A high degree of collinearity will produce:

1. no least squares output when collinearity is exact.
2. large standard errors and wide confidence intervals.
3. insignificant t-values even with high R**2** and a significant F-value.
4. estimates sensitive to deletion or addition of a few observations or “insignificant” variables.
5. good “within-sample”(same proportions) but poor “out-of-sample”(different proportions) prediction

Identifying Collinearity

Evidence of high collinearity include**:**

1. a high pairwise correlation between two explanatory variables.
2. a high R-squared when regressing one explanatory variable at a time on each of the remaining explanatory variables.
3. a statistically significant F-value when the t-values are statistically insignificant.
4. an R-squared that doesn’t fall by much when dropping any of the explanatory variables.

Mitigating Collinearity

1. collect more data with better information.
2. impose economic restrictions as appropriate.
3. impose statistical restrictions when justified.
4. if all else fails at least point out that the poor
5. model performance might be due to the collinearity problem (or it might not).

menggunakan korelasi parsial 🡺 nilai R21 lebih tinggi dari nilai R2 pada regresi antar variabel bebas, maka dalam model empirik tidak terdapat adanya multikolinieritas, dan sebaliknya

|  |
| --- |
| Dependent Variable: LOG(Y) |
| Method: Least Squares |
| Sample: 1960 1982 |
| Included observations: 23 |
| Variable | Coef | Se | t-Statistic | Prob.  |
| C | 2.514 | 0.182 | 13.804 | 0.000 |
| LOG(X3) | -0.790 | 0.117 | -6.729 | 0.000 |
| LOG(X4) | 0.473 | 0.082 | 5.741 | 0.000 |
| LOG(X5) | 0.440 | 0.074 | 5.957 | 0.000 |
| R-squared | 0.966 |  |  |
| Adjusted R-squared | 0.960 |  |  |

|  |
| --- |
| Dependent Variable: LOG(X3) |
| Method: Least Squares |
| Sample: 1960 1982 |
| Included observations: 23 |
| Variable | Coef | se | t-Statistic | Prob.  |
| C | 1.295 | 0.191 | 6.779 | 0.000 |
| LOG(X4) | 0.368 | 0.134 | 2.752 | 0.012 |
| LOG(X5) | 0.193 | 0.134 | 1.446 | 0.164 |
| R-squared | 0.906 |  |  |
| Adjusted R-squared | 0.897 |  |  |

|  |
| --- |
| Dependent Variable: LOG(X4) |
| Method: Least Squares |
| Sample: 1960 1982 |
| Included observations: 23 |
| Variable | Coef | se | t-Statistic | Prob.  |
| C | -1.038 | 0.436 | -2.382 | 0.027 |
| LOG(X3) | 0.746 | 0.271 | 2.752 | 0.012 |
| LOG(X5) | 0.548 | 0.158 | 3.462 | 0.003 |
| R-squared | 0.935 |  |  |
| Adjusted R-squared | 0.929 |  |  |

|  |
| --- |
| Dependent Variable: LOG(X5) |
| Method: Least Squares |
| Sample: 1960 1982 |
| Included observations: 23 |
| Variable | Cof | se | t-Statistic | Prob.  |
| C | -0.163 | 0.551 | -0.296 | 0.770 |
| LOG(X3) | 0.489 | 0.338 | 1.446 | 0.164 |
| LOG(X4) | 0.684 | 0.198 | 3.462 | 0.003 |
| R-squared | 0.919 |  |  |
| Adjusted R-squared | 0.911 |  |  |

**Pendekatan Koutsoyiannis**

Metode yang dikembangkan oleh Koutsoyiannis (1977) menggunakan coba-coba dalam memasukkan variabel

|  |
| --- |
| Dependent Variable: LOG(Y) |
| Variable | coef | se | t | Prob.  |
| C |  2.513867 |  0.182108 |  13.80426 |  0.0000 |
| LOG(X3) | -0.789744 |  0.117361 | -6.729179 |  0.0000 |
| LOG(X4) |  0.473400 |  0.082462 |  5.740837 |  0.0000 |
| LOG(X5) |  0.439705 |  0.073808 |  5.957421 |  0.0000 |
| R-squared |  0.965683 |  |  |
| Adjusted R-squared |  0.960265 |  |  |

|  |
| --- |
| Dependent Variable: LOG(Y) |
| Variable | coef | se | t | Prob.  |
| C |  1.046069 |  0.422623 |  2.475183 |  0.0219 |
| LOG(X3) |  0.680506 |  0.109687 |  6.204055 |  0.0000 |
| R-squared |  0.647001 |  |  |
| Adjusted R-squared |  0.630192 |  |  |

|  |
| --- |
| Dependent Variable: LOG(Y) |
| Variable | Coef | Se | T | Prob.  |
| C |  1.640004 |  0.183436 |  8.940460 |  0.0000 |
| LOG(X4) |  0.456435 |  0.041225 |  11.07183 |  0.0000 |
| R-squared |  0.853746 |  |  |
| Adjusted R-squared |  0.846781 |  |  |

|  |
| --- |
| Dependent Variable: LOG(Y) |
| Variable | Coef | Se | T | Prob.  |
| C |  1.640004 |  0.183436 |  8.940460 |  0.0000 |
| LOG(X4) |  0.456435 |  0.041225 |  11.07183 |  0.0000 |
| R-squared |  0.853746 |  |  |
| Adjusted R-squared |  0.846781 |  |  |

R2 pada regresi dengan tiga variabel bebas memberikan nilai R2 yang lebih tinggi dibanding R2 pada regresi dengan masing-masing variabel bebas

Perbaikan multikolinearitas

1. informasi apriori
2. pooling data
3. drop variabel – uji dengan wald test
4. transformasi variabel
5. tambah observasi

**Wald Test**

1. wald test dilakukan untuk mengeluarkan variabel dari model
2. Regresi model A dengan memasukkan semua variabel bebas
3. Lakukan Wald test terhadap variabel yang paling tidak signifikan pada regresi awal, misal X2.
	* Dari Equation:
	* klik VIEW
	* COEFFICIENT TESTS
	* WALD – Coefficient Restrictions,
	* C(2)=0 (Pada kotak dialog tuliskan koefisien yang akan direstriksi, yaitu c(2)=0)
	* OK
4. Probabilitas F tidak signifikan, berarti variabel X6 bisa dikeluarkan dari model.

**Uji Autokorelasi**

1. Terdapat korelasi antara ui dan uj
2. biasanya muncul dalam data *time-series* 🡺 observasi diurutkan secara kronologis
3. For **efficiency** (accurate estimation/prediction) all systematic information needs to be incorporated into the regression model.
4. **Autocorrelation** is a systematic pattern in the errors that can be either attracting (**positive**)or repelling (**negative**) autocorrelation.

*E*(*u*i *u*j) = 0 *i ≠ j*

*E(ui uj) ≠ 0 i ≠ j*

**Autokorelasi dan serial korelasi**

Gerhard Tihtner tahun 1965 (lihat Gujarati, 2003: 442) mendefinisikan otokorelasi sebagai korelasi kelambanan (*lag correlation*) suatu deretan tertentu dengan dirinya sendiri, tertinggal oleh sejumlah unit waktu (*u*1, *u*2, ... , *u*10 dan *u*2, *u*3, ... , *u*11,). Sedangkan serial korelasi adalah korelasi kelambanan (*lag correlation*) antara dua seri atau rangkaian yang berbeda (*u*1, *u*2, ... , *u*10 dan *v*2, *v*3, ... , *v*11, di mana *u* dan *v* merupakan dua deretan waktu yang berbeda). Meskipun demikian saat ini dalam praktek yang lazim, istilah otokorelasi dan serial korelasi adalah dianggap sinonim.



Penyebab autokorelasi

1. Kesalahan model (linier – non linier)
2. Penggunaan Lag (inertia) 🡺 data observasi pada periode sebelumnya dan periode sekarang, kemungkinan besar akan saling ketergantungan (*interdependence*)
3. fenomena cobweb 🡺 Munculnya fenomena sarang laba-laba terutama terjadi pada penawaran komoditi sektor pertanian 🡺 Misalnya, panen komoditi permulaan tahun dipengaruhi oleh harga yang terjadi pada tahun sebelumnya 🡺 ui tidak lagi bersifat acak (*random*), tetapi mengikuti suatu pola yaitu sarang laba-laba.
4. Tidak memasukkan variabel yang penting
5. Manipulasi data

Akibat autokorelasi

1. Penaksir-penaksir OLS masih linier dan tidak bias.
2. Tetapi penaksir-penaksir OLS tidak efisien dibandingkan dengan prosedur dalam otokorelasi. Dengan kata lain, penaksir-penaksir OLS tidak lagi BLUE.
3. Penaksir-penaksir OLS akan mempunyai varian yang bias.
4. Oleh karena itu, uji *t* dan uji *F*, secara umum tidak dapat dipercaya.
5. Biasanya formulasi untuk menghitung varian kesalahan (*error variance*) yaitu (RSS/*degree of freedom*) adalah suatu penaksir yang bias untuk *σ*2 yang sebenarnya dan dalam kasus ini akan mengestimasi terlalu rendah (*underestimate*).
6. Sebagai konsekuensinya, nilai *R*2 yang dihasilkan akan lebih tinggi daripada yang seharusnya, sehingga nilai *R*2 tersebut tidak dapat dipercaya.
7. Perhitungan nilai varian dan kesalahan baku yang akan digunakan untuk peramalan tidak akan efisien

Deteksi autokorelasi

1. Uji Durbin-Watson

 2. Uji Langrage Multiplier

 3. Uji Breusch-Godfrey

 4. Runs Test

**Uji Durbin-Watson**

1. Estimasi model dengan OLS – simpan ei (error)
2. Menghitung nilai statistik d
3. Menentukan nilai kritis berdasarkan ukuran sampel dan jumlah variabel penjelas – Tabel Durbin Watson
4. Membagi distribusi untuk menentukan posisi nilai kritis segingga bisa disimpulkan
5. Lihat tabel DW untuk menentukan nilai kritis
6. Nilai DW terletak antara 0-4
7. bandingkan nilai D-W statistik dengan nilai D-W tabel
8. Tolak H0 yang mengatakan tidak ada otokorelasi positif, bila nilai D-W statistik terletak antara 0 < d < dL.
9. Ragu-ragu (*inconclusive*) tidak ada otokorelasi positif bila dL ≤ d ≤ dU.
10. Tolak H0 yang mengatakan tidak ada otokorelasi negatif, bila nilai D-W statistik terletak antara 4 – dL < d < 4.
11. Ragu-ragu (*inconclusive*) tidak ada otokorelasi negatif bila 4 – dU ≤ d ≤ 4 – dL.
12. Terima H0 yang mengatakan tidak ada otokorelasi positif ataupun otokorelasi negatif, bila nilai D-W statistik terletak antara dU < d < 4 – dL.

f(d)

d

0

dL

dU

(4-dU)

(4-dL)

2

4

Daerah Tidak Menolak H0: Tidak Ada Otokorelasi Positif atau Otokorelasi Negatif

Menolak H0:

Ada Otokorelasi Positif

Daerah

Ragu-Ragu

Daerah

Ragu-Ragu

Menolak H0:

Ada Otokorelasi Negatif

|  |
| --- |
| Dependent Variable: LOG(Y) |
| Method: Least Squares |
| Sample(adjusted): 1961 1982 |
| Included observations: 22 after adjusting endpoints |
| Variable | Coefficient | Std. Error | t-Statistic | Prob.  |
| C |  1.134049 |  0.339709 |  3.338299 |  0.0039 |
| LOG(X3) | -0.360771 |  0.136537 | -2.642298 |  0.0171 |
| LOG(X4) |  0.229115 |  0.087129 |  2.629603 |  0.0176 |
| LOG(X5) |  0.171344 |  0.081166 |  2.111040 |  0.0499 |
| LOG(Y(-1)) |  0.574380 |  0.130804 |  4.391163 |  0.0004 |
| R-squared |  0.981073 |  Mean dependent var |  3.679289 |
| Adjusted R-squared |  0.976620 |  S.D. dependent var |  0.176570 |
| S.E. of regression |  0.026998 |  Akaike info criterion | -4.189368 |
| Sum squared resid |  0.012391 |  Schwarz criterion | -3.941404 |
| Log likelihood |  51.08305 |  F-statistic |  220.3026 |
| Durbin-Watson stat |  1.689083 |  Prob(F-statistic) |  0.000000 |

**Uji Autokorelasi: Breusch-Godfrey LM**

- mengidentifikasi masalah otokorelasi dengan tingkat derajat tinggi

Tahapan Pengujian:

1. Estimasi model
2. Simpan residual, ambil lag residual
3. Regresi residual dengan lag residual dan semua regresor
4. Test statistik BG = (N-p)\*R2 dimana R2 adalah koefisien determinasi regresi ~ chi-square (p)

|  |
| --- |
| Breusch-Godfrey Serial Correlation LM Test: |
| F-statistic |  0.513803 |  Probability |  0.482695 |
| Obs\*R-squared |  0.638305 |  Probability |  0.424325 |
| Test Equation: |
| Dependent Variable: RESID |
| Method: Least Squares |
| Variable | Coefficient | Std. Error | t-Statistic | Prob.  |
| C | -0.042557 |  0.193802 | -0.219592 |  0.8287 |
| LOG(X3) |  0.043102 |  0.133233 |  0.323508 |  0.7500 |
| LOG(X4) |  0.006757 |  0.084068 |  0.080372 |  0.9368 |
| LOG(X5) | -0.032264 |  0.087274 | -0.369691 |  0.7159 |
| RESID(-1) |  0.202659 |  0.282728 |  0.716800 |  0.4827 |
| R-squared |  0.027752 |  Mean dependent var |  1.63E-15 |
| Adjusted R-squared | -0.188303 |  S.D. dependent var |  0.034764 |
| S.E. of regression |  0.037896 |  Akaike info criterion | -3.518308 |
| Sum squared resid |  0.025849 |  Schwarz criterion | -3.271461 |
| Log likelihood |  45.46054 |  F-statistic |  0.128451 |
| Durbin-Watson stat |  1.929031 |  Prob(F-statistic) |  0.970114 |

nilai probabilitas lebih besar dari probabilitas 5%, maka hipotesa yang menyatakan pada model tidak terdapat autokorelasi tidak ditolak. Bararti model empirik lolos dari masalah autokorekasi

**Heteroscedasticity:***Nature*

1. Assumption variance of the true error term to be constant, i.e., Var(ui) = σ2 for all i.
2. When this is violated, we have the problem of heteroscedasticity-- variance of the true error term changes form observation to observation
3. 🡺 Residual tidak mempunyai varian yang sama

 Homoskedastis pattern of errors

 

Homoskedastic case



Heteroskedastic pattern of error

 

Heteroskedastic case

 

**Konsekuensi heteroskedastisitas**

1. These are generally the same as those of pure positive first-order autocorrelation
2. *OLS* estimators of βk remain unbiased and consistent
3. The true variance of βk becomes larger so that *OLS* estimators will be inefficient 🡺 inferensi menjadi keliru karena variannya keliru
4. *σ^2 = RSS/(n-k)* akan menjadi estimator yang bias bagi homoscedastic error variance, *σ2*.
5. Biasanya, jika varians naik akibat kenaikan salah satu variabel independen, maka *σ^2*will be biased *downwar*
6. Akibatnya, *Var(βk)* menghasilkan estimator yang bias
7. Since σ*2* is usually underestimated in the presence of heteroscedasticity, Var(β^*k*) is underestimated.
8. Uji t dan F tidak lagi valid ketika terjadi heteroskedastisitas
9. t ratios will be overestimated thus giving us more confidence than warranted.
10. All F statistics and R*2* will be overestimated

**Uji Heteroskedasitas**

1. uji Park 🡺 
2. Uji White
3. Uji Bruce-Pagan-Godfrey LM

***White’s Test***

1. This is a large sample test
2. Consider the following model, 🡺 Yi = β1 + β2X2i + β3X3i + ui
3. Assumes the following pattern of heteroscedasticity

σ2i = α1 + α2X2 + α3X3 + α4X22 + α5X32 + α6X2X3

1. Because σ2i is not known, White suggests using squared residual from the original regression as a proxy
2. So estimate the original model and retrieve the estimated residuals, u^.
3. Use the squared residuals as a proxy for the heteroscedastic error variance, σ2i, and estimate the following *auxiliary regression*,

u^2 = α1+ α2X2 + α3X3 + α4X22 + α5X32 + α6X2X3

1. Now test the null of no heteroscedasticity, i.e.,

H0: α2 = α3 = α4 = α5 = α6 = 0

1. using nR2~χ2(k-1), where n is the sample size, R2 is the unadjusted R2 from the auxiliary regression in Step 2, and k is the number of parameters in the auxiliary regression

Langkah pengujian dengan Eviews: dari menu *equation*, klik *residual test/white heteroscedasticity*

Ho: tidak terdapat heteroskedastisitas

Ha : terdapat heteroskedastisitas

df = jumlah variabel tak bebas pada regresi White.

🡺 Jika nilai chi-square lebih besar dari nilai kritis chi-square 🡪signifikan, terdapat heteroskedastisitas 🡺 atau lihat probabilitas

**Normalitas**

* Uji normalitas berkaitan dengan kondisi distribusi probabilitas gangguan.
* Asumsi yang dibuat mengenai ut bahwa faktor pengganggu mempunyai nilai rata-rata yang diharapkan adalah sama dengan nol, tidak berkorelasi dan mempunyai varian yang konstan atau Ut ~ N(0, σ2).
* Notasi tersebut menunjukkan bila dua variabel yang didistribusikan secara normal, maka kovarians atau korelasi merupakan dua variabel yang bebas, yaitu ui dan uj tidak hanya tidak berkorelasi tetapi juga didistribusikan secara bebas atau *Normally and Independent Distributed* (ui ~ NID(0, σ2)).

**Uji Normalitas: Jarque-Bera Test**

* Menguji asumsi normalitas residual
* Justifikasi penggunaan distribusi t, F dan lainnya
* Bila tidak terpenuhi, maka tidak bisa dilakukan inferensi (kecuali dengan distribusi asimptotik)
* Prosedur Test: membandingkan moment distribusi residual dari model dengan moment distribusi normal

 Terdapat beberapa test yang digunakan, diantaranya adalah: (1) chi-square goodness of fit test, dan (2) jarque-bera test. Keduanya menguji residual μ dan dengan distribusi probalibitas chi-square



 S : skewness, K : kurtosis, k : jumlah koefisien yang diestimasi.

 Uji Normalitas

* Langkah pengujian dengan Eviews: dari menu *equation*, klik *view/ residual test/histogram-normality test*.
* Inferensi dilakukan dengan membandingkan nilai JB hitung (yang ditunjukkan oleh nilai χ2(k). Bila nilai JB hitung > nilai χ2-tabel, maka histogram akan berbentuk seperti lonceng dan hipotesis bahwa residual ut terdistribusi normal ditolak.

 Contoh normalitas test

 



**Uji Spesifikasi Model (Ramsey’s Reset Test)**

* Regression Specification Error Test and was proposed by Ramsey (1969)
* Uji ini sebenarnya dilakukan untuk menguji ada tidaknya kesalahan spesifikasi dalam regresi
* Uji Linieritas model🡺 Menggunakan uji F
* Langkah pengujian:
	+ Lakukan estimasi OLS terhadap model awal, kemudian hitung nilai fittednya (LYF)
	+ Estimasi model awal dengan ditambah variabel fitted yang tidak linier (misal YF2).
	+ Hitung nilai F dengan rumus F restriksi. Jika nilai F statistik signifikan, maka model awal terjadi mis-spesifikasi.
* Langkah dengan Eviews:
	+ dari equation, klik VIEW
	+ STABILITY TESTS
	+ RAMSEY RESET TEST
	+ 1 (pada kotak dialog isikan pada number of fitted jika fitted pangkat 2)
	+ OK

 **Contoh Ramsey’s Reset Test**

|  |
| --- |
| Ramsey RESET Test: |
| F-statistic |  57.59673 |  Probability |  0.000000 |
| Log likelihood ratio |  47.30640 |  Probability |  0.000000 |
| Test Equation: |
| Dependent Variable: LOG(INVR) |
| Included observations: 104 |
| Variable | Coefficient | Std. Error | t-Statistic | Prob.  |
| C | -92.83684 |  12.13690 | -7.649138 |  0.0000 |
| LOG(RK) |  2.047655 |  0.256772 |  7.974587 |  0.0000 |
| LOG(YDR) |  15.74801 |  1.954297 |  8.058147 |  0.0000 |
| FITTED^2 | -0.807499 |  0.106400 | -7.589251 |  0.0000 |
| R-squared |  0.927131 |  Mean dependent var |  10.00437 |
| Adjusted R-squared |  0.924945 |  S.D. dependent var |  0.470160 |
| S.E. of regression |  0.128806 |  Akaike info criterion | -1.223318 |
| Sum squared resid |  1.659096 |  Schwarz criterion | -1.121611 |
| Log likelihood |  67.61255 |  F-statistic |  424.1091 |
| Durbin-Watson stat |  0.873171 |  Prob(F-statistic) |  0.000000 |

Ho: ~εN(0,σ2I)

Ha: ~εN(μ,σ2I) ≠μ0

Dari hasil di atas terlihat bahwa, nilai F signifikan. Berarti terjadi kesalahan spesifikasi pada model awal.