

LECTURE NOTES #7

MULTIKOLINEARITAS DAN AUTOKORELASI

I. Pendahuluan

Asumsi 4 agar estimator OLS bersifat BLUE adalah tidak adanya kolinearitas sempurna diantara variabel bebas. Istilah ini dikenalkan oleh Ragnar Frisch (1934) yang berarti hubungan linier yang sempurna diantara variabel bebas.

Adanya hubungan diantara variabel bebas adalah hal yang tak terelakkan dan memang diperlukan agar regresi yang diperoleh dapat bersifat valid. Namun demikian hubungan yang bersifat linier hendaknya dihindarkan karena akan membawa konsekuensi gagal estimasi (multikolinearitas sempurna) atau kesulitan dalam inferensi (multikolinearitas tidak sempurna).

Dalam bagian ini akan dibahas berbagai aspek mengenai multikolinearitas, mulai dari pengertian/konsep, implikasi bagi model regresi, teknik deteksi dan prosedur koreksi.

Disamping itu diuraikan pula salah satu pelanggaran asumsi OLS yang lain, yakni adanya korelasi diantara residual. Sebagaimana diketahui agar OLS bersifat BLUE, residual regresi diasumsikan bersifat independen dari satu observasi ke observasi lainnya. Pelanggaran terhadap sifat ini dikenal dengan nama autokorelasi (atau *serial correlation*).

Seperti juga multikolinearitas pembahasan aspek autokorelasi, akan mengikuti sistematika yang dimulai dari pengertian/konsep, implikasi bagi model regresi, teknik deteksi dan prosedur koreksi.

II. Konsep Multikolinearitas

Suatu hubungan antara k variabel bebas disebut linier sempurna jika ia memenuhi

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k = 0 \quad \dots\dots\dots 1)$$

Dengan sedikit manipulasi matematis, dapat ditunjukkan untuk setiap observasi ke i, maka berlaku

$$X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \dots - \frac{\lambda_k}{\lambda_2} X_{ki} \quad \dots\dots\dots 2)$$

Dilain pihak hubungan antara k variabel disebut linier tidak sempurna jika memenuhi

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k + v_i = 0 \quad \dots\dots\dots 3)$$

Dan dengan demikian

$$X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \dots - \frac{\lambda_k}{\lambda_2} X_{ki} - \frac{v_i}{\lambda_2} \dots\dots\dots 4)$$

Ingat kembali bahwa prosedur OLS pada intinya adalah suatu usaha untuk menyelesaikan suatu sistem persamaan simultan berjumlah i dengan k+1 variabel yang dicari (β_k), atau

$$\sum_i^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki}) = 0$$

$$\sum_i^n x_{1i} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki}) = 0 \dots\dots\dots 5)$$

.....

$$\sum_i^n x_{ki} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki}) = 0$$

Dari prinsip matematika matriks, kita mengetahui bahwa jika system persamaan 5 mengalami kondisi ketidak bebasan secara linier maka system persamaan tersebut tidak akan memiliki solusi. Inilah yang terjadi jika variabel bebas yang ada pada model mengalami multikolinearitas sempurna. Sedangkan jika multikolinearitas adalah tidak sempurna, maka kesulitan yang terjadi adalah kesulitan dalam inferensi karena standar error yang bersifat variabel.

Terdapat beberapa penyebab multikolinearitas diantaranya (Montgomery and Peck, 1982):

- a. Cara pengambilan data dan kecilnya ukuran sample.
- b. Pembatas pada model atau populasi yang disampel. Misalnya kita meregresi konsumsi listrik terhadap pendapatan dan ukuran rumah. Disini populasi darimana sample diperoleh memiliki karakteristik kolinearitas, dimana individu yang memiliki pendapatan tinggi umumnya memiliki rumah berukuran besar.
- c. Spesifikasi model. Penambahan polynomial (x^2 , x^3 , dst) berpotensi menimbulkan masalah multikolinearitas terutama jika range nilai x yang dimiliki adalah kecil.

- d. Model yang *overdetermined*. Hal ini terjadi jika model dimaksud memiliki lebih banyak variabel dibandingkan jumlah sample (umumnya terjadi pada penelitian medis).
- e. Common trend. Terutama jika kita menggunakan data *time series*, banyak variabel seperti GDP, konsumsi agregat, PMA, dsb bergerak searah berdasarkan waktu.

III. Implikasi Multi Kolinearitas

Multikolinearitas tidak mengubah sifat parameter OLS sebagai *Best Linear Unbiased Estimator* (BLUE). Parameter yang diperoleh adalah valid untuk mencerminkan kondisi populasi dan ia adalah yang terbaik (dalam artian memiliki varians yang minimum) diantara estimator linier.

Namun demikian keberadaan multikolinearitas bukannya tidak berdampak negatif. Dapat ditunjukkan bahwa keberadaan kolinearitas akan menyebabkan varians parameter yang diestimasi akan menjadi lebih besar dari yang seharusnya, dengan demikian tingkat presisi dari estimasi akan menurun. Konsekuensi lanjutnya adalah rendahnya kemampuan menolak hipotesis null (power of test).

Hal ini dapat dilihat melalui suatu ilustrasi, misalnya kita mengestimasi suatu model regresi linier dengan 1 variabel tergantung dan 2 variabel bebas tanpa intersep sbb:

$$y = \beta_1 x_1 + \beta_2 x_2 + u \quad \dots\dots\dots 6)$$

Varians β_1 dan β_2 serta kovarians $\beta_1\beta_2$ dapat diperoleh sbb (lihat appendiks LNo4):

$$\begin{aligned} \text{var}(\hat{\beta}_1) &= \frac{\sigma^2}{\sum_{i=1}^n x_{1i}^2 (1 - r_{12}^2)} \\ \text{var}(\hat{\beta}_2) &= \frac{\sigma^2}{\sum_{i=1}^n x_{2i}^2 (1 - r_{12}^2)} \quad \dots\dots\dots 7) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) &= \frac{-r_{12}\sigma^2}{(1 - r_{12}^2) \sqrt{\sum_{i=1}^n x_{1i}^2 \sum_{i=1}^n x_{2i}^2}} \end{aligned}$$

Dimana r_{12} adalah koefisien korelasi antara x_1 dan x_2 . Dapat dilihat disini dengan semakin besarnya koefisien tersebut maka varians β_1 dan β_2 akan semakin besar.

Selanjutnya kita mengetahui bahwa standar error parameter (yang merupakan akar dari varians) diperlukan untuk menghitung signifikansi. Dengan demikian meningkatnya varians akibat terjadinya kolinearitas akan menyebabkan nilai t statistik menjadi kecil. Akibatnya akan semakin rendah kemampuan pula model untuk menolak hipotesis null (derajat signifikansi koefisien adalah rendah). Secara praktis model dikatakan bias kepada hipotesis null (dalam jargon statistik memiliki power yang rendah).

IV. Teknik Deteksi

Gujarati (2003) menyatakan bahwa multikolinearitas adalah fenomena sampling. Ia terjadi pada sample dan bukan pada populasi. Hal ini tentu saja jika kita telah menspesifikasikan variabel yang masuk kedalam model dengan benar (misalnya tidak ada variabel yang merupakan multiplikasi dari variabel lainnya). Dengan kata lain jika dimungkinkan untuk bekerja pada populasi maka multikolinearitas tidak akan pernah menjadi suatu masalah.

Dalam istilah lain yang terkait, Kmenta (1986) menyatakan permasalahan multikolinearitas adalah persoalan derajat bukan apakah ada atau tidak ada suatu kolinearitas pada data yang dimiliki. Beberapa metoda yang dapat digunakan untuk mengukur derajat kolinearitas adalah

1. R^2 yang tinggi tetapi sedikit variabel yang signifikan. Meskipun kolinearitas menyebabkan standar error dari parameter menjadi lebih besar tetapi hal ini tidak terjadi pada model secara keseluruhan. Residual model adalah tidak bias dan dengan demikian R^2 yang dimiliki adalah valid. Dengan demikian jika kita memiliki model dengan R^2 yang tinggi (misalnya >0.7) tetapi sedikit variabel yang signifikan, kita dapat menduga bahwa model yang dimiliki mengalami multikolinearitas.
2. Koefisien korelasi yang tinggi antara regressor. Cara langsung mendeteksi adanya multikolinearitas adalah dengan menghitung koefisien korelasi diantara variabel bebas. Koefisien korelasi yang dihitung dapat bersifat *pairwise correlation* (zero order correlation): yang menunjukkan korelasi antara variabel x_i atau bersifat parsial (Farrar-Glauber, 1967): menghitung korelasi antara dua koefisien korelasi yang terpisah ($r_{12,34}$, hitung korelasi variabel x_1 dengan x_2 (r_{12}) dan x_3 dan x_4 (r_{34}) kemudian hitung korelasi antara r_{12} dengan r_{34}).
3. *Overall significance dari Auxiliary Regression*. Kita membuat regresi auxiliary antara variabel-variabel yang dicurigai mengalami multikolinearitas dan menghitung *overall significance* (F Test). Suatu regresi auxiliary yang signifikan mendukung dugaan atas adanya multikolinearitas.

V. Prosedur Koreksi

Terdapat pandangan yang cukup berbeda diantara *econometricians* mengenai tingkat kepentingan dari masalah multikolinearitas. Seperti yang

diuraikan diatas bahwa kolinearitas adalah masalah sample, dan seperti pada penelitian sosial umumnya tidak banyak yang dapat dilakukan untuk “menurangi” data yang ada ditangan. Ini adalah masalah defisiensi data (atau dalam istilah Goldberger,1991: *micronumerosity*). Jika dimungkinkan kita dapat mengurangi dampak kolinearitas dengan menambah data, atau jika tidak memungkinkan maka diterima apa adanya. Dilain pihak terdapat mereka yang menyatakan bahwa terdapat suatu kemungkinan memperbaiki dengan data yang ada. Beberapa hal yang disarankan untuk dilakukan diantaranya:

1. Penggunaan Informasi Apriori. Informasi apriori adalah informasi yang bersifat non sample. Ia tidak berasal dari data melainkan dari teori, penelitian lainnya, atau judgement peneliti. Misalnya kita meregresi model berikut

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \quad \dots\dots\dots 8)$$

dan menduga bahwa x_1 dan x_2 mengalami kolinearitas. Jika kita memiliki informasi apriori yang menyatakan bahwa katakanlah $\beta_2=0.1 \beta_1$ maka kita dapat mereestimasi kembali persamaan 8 menjadi

$$y = \beta_0 + \beta_1 x^* + u; \quad \dots\dots\dots 9)$$

dimana

$$x^* = x_1 + 0.1x_2 \quad \dots\dots\dots 10)$$

2. Penggunaan data panel. Data semacam ini memiliki beberapa karakter yang berguna bagi penelitian dan robust terhadap beberapa pelanggaran asumsi (termasuk multikolinearitas), lihat Wooldrige (2005) bab 13 untuk suatu diskusi.
3. Penggantian/mengeluarkan variabel. Hal ini dilakukan jika tidak menyebabkan *specification error* (variabel yang dihilangkan tidak berasal dari teori) dan bersifat substitusi terhadap variabel lainnya. Misalnya kita melakukan regresi konsumsi gas terhadap pendapatan, ukuran rumah dan jumlah kendaraan. Ukuran rumah dan jumlah kendaraan adalah konsep kekayaan sehingga kita dapat mengeluarkan salah satunya.
4. Tranformasi variabel. Beberapa bentuk transformasi yang umum digunakan adalah first difference, rasio transformation (seperti pada WLS) dan bentuk log.

Contoh 1.

Dengan menggunakan Longley Data (1967), kita akan melakukan regresi y (jumlah orang yang bekerja) terhadap x_1 s/d x_5 (gnp deflator, gnp nominal, jumlah pengangguran, jumlah personel angkatan bersenjata, non

institutionalized population) . Hasil yang diperoleh dapat dirangkum pada tabel 1, sbb:

Dependent Variable: Y
 Method: Least Squares
 Date: 06/13/08 Time: 09:08
 Sample: 1947 1962
 Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	77270.12	22506.71	3.433204	0.0075
X1	1.506187	8.491493	0.177376	0.8631
X2	-0.035819	0.033491	-1.069516	0.3127
X3	-2.020230	0.488400	-4.136427	0.0025
X4	-1.033227	0.214274	-4.821985	0.0009
X5	-0.051104	0.226073	-0.226051	0.8262
TIME	1829.151	455.4785	4.015890	0.0030

R-squared	0.995479	Mean dependent var	65317.00
Adjusted R-squared	0.992465	S.D. dependent var	3511.968
S.E. of regression	304.8541	Akaike info criterion	14.57718
Sum squared resid	836424.1	Schwarz criterion	14.91519
Log likelihood	-109.6174	F-statistic	330.2853
Durbin-Watson stat	2.559488	Prob(F-statistic)	0.000000

Tabel 1. Print Out Regresi Contoh 1

Seperti yang dapat dilihat pada hasil regresi $R^2=0.99$ suatu hasil yang sangat tinggi tetapi 3 dari 5 variabel x adalah tidak signifikan. Dengan demikian kita menduga pada data terdapat masalah multikolinearitas. Hal ini ditunjukkan dengan melakukan perhitungan pairwise correlation sbb:

	TIME	X1	X2	X3	X4	X5
TIME	1.000000	0.991149	0.995273	0.668257	0.417245	0.993953
X1	0.991149	1.000000	0.991589	0.620633	0.464744	0.979163
X2	0.995273	0.991589	1.000000	0.604261	0.446437	0.991090
X3	0.668257	0.620633	0.604261	1.000000	-0.177421	0.686552
X4	0.417245	0.464744	0.446437	-0.177421	1.000000	0.364416
X5	0.993953	0.979163	0.991090	0.686552	0.364416	1.000000

Tabel 2. Pairwise Correlation antara variabel bebas

Koefisien korelasi antara variabel X_1X_2 , X_1X_5 dan X_2X_5 adalah sangat tinggi (>0.90) yang menjustifikasi dugaan kita atas multikolinearitas yang diperoleh diatas.

Koreksi dilakukan dengan melakukan respesifikasi model, yakni

- Bentuk variabel baru yakni gnp riil karena akan lebih sesuai dengan teori. GNP riil (RGNP) diperoleh dengan membagi GNP nominal dengan GNP deflator ($=x_2/x_1$).
- Variabel X_5 karena pertumbuhan alami akan berkorelasi dengan variabel waktu, sehingga perlu dipilih salah satu. Dalam kasus ini kita memilih untuk mengeluarkan time.
- Tidak ada alasan untuk memasukkan variabel jumlah pengangguran sebagai penjas dari jumlah orang bekerja, karena itu ia dikeluarkan.

Dengan melakukan penyesuaian tersebut dan melakukan regresi kembali diperoleh hasil sbb:

Dependent Variable: Y
 Method: Least Squares
 Date: 06/13/08 Time: 09:16
 Sample: 1947 1962
 Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	65720.37	10624.81	6.185558	0.0000
RGNP	9.736496	1.791552	5.434671	0.0002
X4	-0.687966	0.322238	-2.134965	0.0541
X5	-0.299537	0.141761	-2.112965	0.0562
R-squared	0.981404	Mean dependent var		65317.00
Adjusted R-squared	0.976755	S.D. dependent var		3511.968
S.E. of regression	535.4492	Akaike info criterion		15.61641
Sum squared resid	3440470.	Schwarz criterion		15.80955
Log likelihood	-120.9313	F-statistic		211.0972
Durbin-Watson stat	1.654069	Prob(F-statistic)		0.000000

Tabel 3. Respesifikasi dan reestimasi model

Dapat dilihat meskipun R^2 sudah menurun tetapi sekarang seluruh variabel sudah signifikan (tidak ada sinyal multikolinearitas).

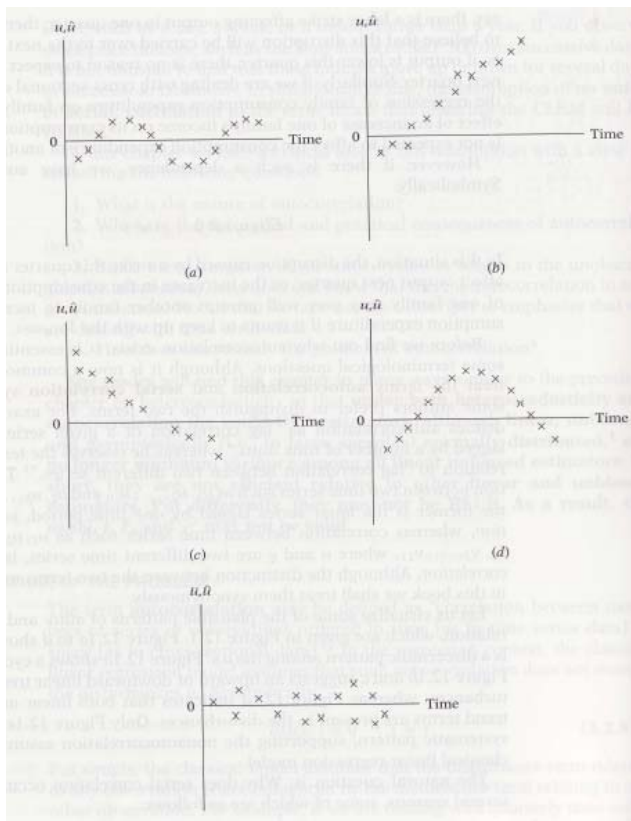
VI. Konsep Auto Korelasi

Autokorelasi menunjukkan sifat residual regresi yang tidak bebas dari satu observasi ke observasi lainnya, atau secara formal

$$E(u_i u_j) \neq 0; i \neq j \quad \dots\dots\dots 11)$$

Fenomena ini umum ditemukan pada regresi dengan data yang bersifat time series tetapi kadang juga ditemukan pada data *cross section*.

Keberadaan autokorelasi dapat dilihat secara kasual (melalui grafik). Grafik 1 menunjukkan berbagai pola residual yang diplot terhadap waktu.



Grafik 1. Pola-pola plot residual terhadap waktu: Autokorelasi vs Non Autokorelasi.

Pola 1a s/d d menunjukkan kondisi autokorelasi, sebagai contoh pada pola 1b ia bersifat positif monotonic (meningkat sejalan dengan berjalannya waktu). Sedangkan pola e menunjukkan kondisi dimana tidak ada autokorelasi, disini residual tersebar disekitar nol pada berbagai titik waktu.

Terdapat beberapa penyebab mengapa fenomena ini timbul, diantaranya

1. Inertia. Salah satu karakteristik umum dari data bersifat time series adalah adanya inertia (sluggishness). Penyesuaian akibat suatu

goncangan terhadap variabel makro ekonomi adalah bersifat bertahap, dan berlangsung sepanjang waktu tertentu. Hal ini juga terjadi pada sekelompok variabel. Dengan demikian kita dapat mengobservasi adanya pergerakan bersama diantara, misalnya: GDP, pengangguran dan tingkat harga yang sebenarnya disebabkan adanya guncangan pada variabel-variabel tersebut dan mereka saat ini berada dalam penyesuaian menuju ekuilibrium. Dalam kondisi ini tentu saja model regresi yang menggunakan variabel-variabel dimaksud akan mengalami autokorelasi.

2. *Specification bias*. Yakni kesalahan dalam menspesifikasi model. Terdapat dua tipe kesalahan, yakni (1) mengeluarkan variabel yang seharusnya ada pada model dan (2) bentuk fungsional yang tidak benar. Pada kasus pertama misalnya kita memodelkan fungsi permintaan daging sapi terhadap pendapatan, jumlah keluarga dan harga daging ayam. Jika karena suatu hal kita tidak memasukkan harga daging ayam, padahal pada populasi variabel ini adalah penting maka kita akan memiliki residual regresi sebagai fungsi dari harga daging ayam. Sedangkan kasus bentuk fungsional terjadi kita memodelkan regresi sebagai fungsi linier yang seharusnya mungkin kuadratik.
3. Fenomena Cobweb. Sering terjadi pada pemodelan dimana terdapat suatu fenomena lagged *response*. Hal ini sering terjadi misalnya pada estimasi fungsi supply produk pertanian. Petani akan mendasarkan keputusan jumlah produksi berdasarkan harga satu periode yang lalu. Dengan demikian residual pada model ini tidak akan bersifat independen, terjadinya overproduksi disuatu periode (sehingga harga turun) akan mendorong petani untuk memproduksi lebih sedikit diperoleh selanjutnya. Dampak variabel lag tidak hanya yang bersifat *successive*, satu periode langsung dibelakangnya, namun bisa terjadi pada beberapa periode. Perilaku persediaan umumnya mengikuti pola seperti ini, persediaan pada waktu t tergantung tidak hanya pada $t-1$, tetapi juga beberapa periode sebelumnya $t-2$, $t-3$, dst.
4. “Manipulasi” Data. Karena satu hal dan lainnya seorang peneliti kadang harus “menukangi” data. Salah satu hal yang sering terjadi adalah akibat perbedaan frekuensi. Misalnya kita akan meregresi nilai tukar terhadap gdp dan suku bunga dengan frekuensi data bulanan. Data GDP adalah bersifat kuartalan dengan demikian dilakukan suatu teknik interpolasi yang menungknkan data gdp kuartalan dipecah menjadi data bulanan. Penggunaan metoda interpolasi ini menimbulkan dampak smoothing yang berpotensi autokorelasi.

VII. Implikasi Autokorelasi

Jika kita memiliki model regresi yang mengalami autokorelasi, maka estimator OLS yang diperoleh adalah tetap tidak bias, konsisten dan secara asimtotik akan terdistribusi dengan normal (Gujarati, 2003). Namun demikian ia menjadi tidak BLUE karena varians residual regresi adalah tidak minimum pada estimator kelas linier.

Untuk melihat sifat ini kita dapat kembali pada model regresi dua variabel (dengan intersep) sbb:

$$y = \beta_0 + \beta_1 x + u \quad \dots\dots\dots 12)$$

dimana β_1 dapat diestimasi sebagai

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{Cov(x, y)}{Var(x)} \quad \dots\dots\dots 13)$$

dengan varians,

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2}{s_x^2} \quad \dots\dots\dots 14)$$

Seperti juga heterokedastisitas, terdapat berbagai bentuk autokorelasi. Untuk ilustrasi ini kita menggunakan bentuk paling sederhana yang disebut autoregressive orde 1 (AR1), dimana

$$u_t = \rho u_{t-1} + \varepsilon_t; -1 < \rho < 1 \quad \dots\dots\dots 15)$$

Koefisien ρ disebut dengan koefisien autokorelasi (autokovarian antara u_t dengan u_{t-1}). Semakin ρ mendekati -1 atau 1 maka, sifat autokorelasi semakin substansial.

Dengan mengasumsikan bahwa regressor juga mengikuti pola autoregressive dengan koefisien autokorelasi r , maka dapat ditunjukkan bahwa varians pada persamaan 14 tidak lagi valid. Formulasi yang tidak bias untuk varians parameter β_1 diberikan oleh

$$\begin{aligned} Var(\hat{\beta}_1)_{AR} &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \frac{1 + rp}{1 - rp} \\ &= Var(\hat{\beta}_1)_{OLS} \frac{1 + rp}{1 - rp} \quad \dots\dots\dots 15) \end{aligned}$$

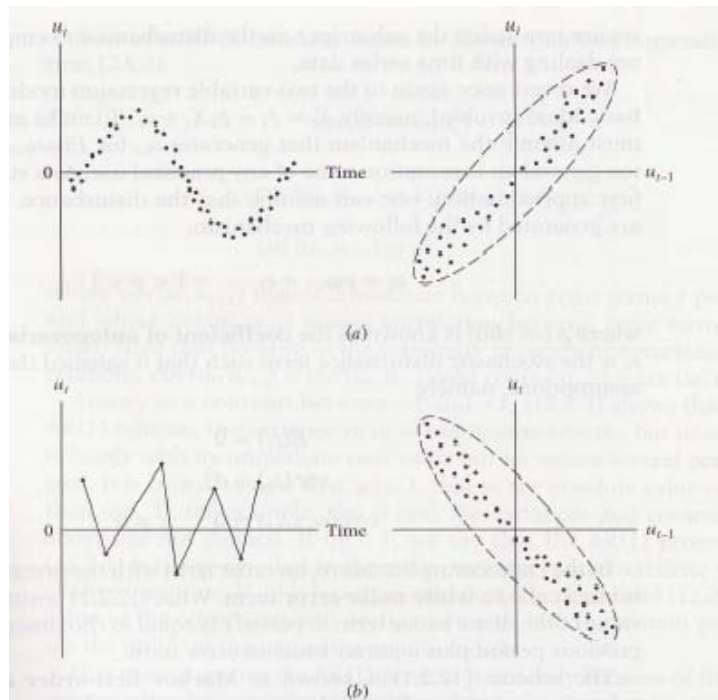
Perhatikan bahwa dari persamaan 13, dapat ditarik kesimpulan estimator OLS adalah tidak bias karena koefisien autokorelasi r dan ρ tidak ada dalam perhitungan. Namun demikian dalam perhitungan varians mereka ada. Semakin substansial masalah autokorelasi, ditunjukkan dengan semakin besarnya angka r dan ρ maka varians dari parameter yang diestimasi akan semakin besar. Sebagai konsekuensinya koefisien regresi menjadi kurang presisi dan terkait dengan kerangka inferensi (pengujian hipotesis) model akan bias kearah null hipotesis.

Hal ini tidak valid karena terdapat metoda estimasi pada kelas linier lainnya yang dapat memberikan varians parameter yang lebih kecil. Dengan kata lain estimator OLS tidak lagi paling efisien.

VIII. Teknik Deteksi

Mengingat dampak autokorelasi adalah negatif terhadap inferensi maka perlu dilakukan suatu tindakan. Namun demikian sebelum melakukan tindakan perlu diketahui terlebih dahulu apakah model yang dimiliki mengalami autokorelasi. Beberapa metoda yang dapat digunakan diantaranya:

1. Observasi Grafik Residual terhadap Waktu. Teknik ini bersifat kasual namun cukup efektif sebagai evaluasi awal. Kita dapat menggunakan plot regresi sebagaimana diberikan pada grafik 1. Disini jika kita menemui residual mengikuti pola 1a, 1b, 1c dan 1d maka dapat diduga bahwa model mengalami autokorelasi.



Grafik 2. Pola-pola Autokorelasi: (a) Positif dan (b) Negatif

Cara lain yang lazim digunakan adalah plot u_t terhadap u_{t-1} . Disini jika pola yang ditemui adalah menyerupai 2a, maka dikatakan terdapat autokorelasi positif dan jika menyerupai 2b, dikatakan terdapat autokorelasi negatif. Kita jug dapat menghitung koefisien korelasinya dan melihat apakah nilainya cukup besar (>0.7).

2. Statistik Durbin Watson/DW (1951). Penggunaan statistik ini dilakukan jika mengasumsikan bahwa pola autokorelasi adalah AR(1) sebagaimana diberikan pada persamaan 15. Disini asumsinya adalah

$$\begin{aligned}
 H_0 : \rho &= 0 \text{ (tidak ada autokorelasi)} && \dots\dots\dots 16) \\
 H_1 : \rho &\neq 0
 \end{aligned}$$

Adapun statistik DW dapat dihitung dengan formula sbb:

$$DW = \frac{\sum_{t=2}^N \hat{u}_t - \hat{u}_{t-1}^2}{\sum_{t=2}^N \hat{u}_t^2} \dots\dots\dots 17)$$

dengan rejection rule

$$\begin{aligned}
 4 - d_l < DW < 4; & \text{ Negative Autocorrelation} \\
 4 - d_u < DW < 4 - d_l; & \text{ Indeterminate} \\
 2 < DW < 4 - d_u; & \text{ No Autocorrelation} \dots\dots\dots 18) \\
 d_l < DW < d_u; & \text{ Indeterminate} \\
 0 < DW < d_l; & \text{ Positive Autocorrelation}
 \end{aligned}$$

Dimana d_l dan d_u adalah batas bawah dan batas atas nilai kritis yang dapat dicari dari Tabel Durbin Watson berdasarkan k (jumlah variabel bebas) dan n (jumlah sample) yang relevan. Statistik DW adalah suatu prosedur rutin yang umum ditemukan pada banyak software statistik, sehingga yang dilakukan adalah melihat apakah nilai dimaksud terletak diantara $2 < DW < 4 - d_u$ untuk menentukan ada tidaknya autokorelasi.

3. Pengujian Autokorelasi Linier Berorde Tinggi (Breusch-Godfrey Test). Adakalanya kita menduga bahwa autokorelasi yang terjadi

adalah pada orde yang tinggi (bukan AR(1)). Untuk itu pengujian dilakukan dengan tahap-tahap berikut:

- a. Asumsikan bahwa autokorelasi bersifat AR(k), atau

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_q u_{t-q} + \varepsilon_t; \quad \dots\dots\dots 19)$$

dimana setiap koefisien autokorelasi ρ_l ; $l=1, \dots, q$ diasumsikan memiliki nilai absolut kurang dari 1.

- b. Regresikan model dan peroleh estimasi residual \hat{u}_t .
 c. Laksanakan auxiliary regression berikut

$$\hat{u}_t = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \rho_1 u_{t-1} + \dots + \rho_q u_{t-q} + v_t \quad \dots\dots\dots 19)$$

- d. Pada hipotesisi null no autocorrelation, yakni

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0 \quad \dots\dots\dots 20)$$

H_1 : Paling tidak satu koefisien tidak sama dengan nol

dapat dihitung statistik uji LM sebagai *rejection rule*. Statistik LM dihitung dengan formula sbb:

$$LM = (n - q)R_u^2 \quad \dots\dots\dots 21)$$

Statistik LM memiliki distribusi χ^2 dengan $df=k$ dan R_u^2 adalah koefisien determinasi yang diperoleh pada regresi persamaan 19. Catatan: meskipun teknik ini umumnya digunakan untuk mendeteksi autokorelasi dengan orde tinggi, tetapi dapat juga digunakan untuk mendeteksi autokorelasi berpola AR(1).

IX. Prosedur Koreksi

Jika pada model regresi yang diperoleh ternyata terdeteksi adanya autokorelasi, maka dilakukan prosedur koreksi. Prosedur koreksi dilakukan berdasarkan kasus yang relevan (bentuk dan asumsi autokorelasi) yang dapat diuraikan sbb:

1. Autokorelasi yang disebabkan oleh fenomena cobweb. Jika kita yakin bahwa autokorelasi disebabkan karena adanya mekanisme cobweb (*lagged response*) maka prosedur koreksi dapat dilakukan dengan menambahkan term lag (y_{t-1}) pada model regresi awal.
2. Autokorelasi berbentuk AR(1) dan ρ diketahui. Jika kita dapat memperoleh estimasi tidak bias atas koefisien autokorelasi, misalnya melalui data, penelitian lainnya, maupun estimasi terhadap persamaan 15 (disebut dengan *Feasible GLS*) maka prosedur koreksi yang dilakukan

adalah suatu varian dari GLS. Kita melakukan transformasi data sbb: dari model awal

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u \quad \dots\dots\dots 22)$$

Kita mentransformasikan setiap variabel (regresand, regressor dan intersep) menjadi *quasi differenced*, sehingga diperoleh

$$\tilde{y}_t = (1 - \rho)\beta_0 + \beta_1 \tilde{x}_{1t} + \beta_2 \tilde{x}_{2t} + \dots + \beta_k \tilde{x}_{kt} + \tilde{u}_t \quad \dots\dots\dots 23)$$

dimana

$$\begin{aligned} \tilde{y}_t &= y_t - \rho y_{t-1} \\ \tilde{x}_{jt} &= x_{jt} - \rho x_{jt-1} \\ \tilde{u}_t &= u_t - \rho u_{t-1} \end{aligned} \quad \dots\dots\dots 24)$$

Transformasi ini dilakukan untuk semua observasi dimana $t \geq 2$. Sedangkan untuk $t=1$ dilakukan bentuk berikut:

$$\tilde{y}_1 = (1 - \rho^2)^{1/2} \beta_0 + \beta_1 \tilde{x}_{11} + \beta_2 \tilde{x}_{21} + \dots + \beta_k \tilde{x}_{k1} + \tilde{u}_1 \quad \dots\dots\dots 25)$$

dimana

$$\begin{aligned} \tilde{y}_1 &= (1 - \rho^2)^{1/2} y_1 \\ \tilde{x}_{j1} &= (1 - \rho^2)^{1/2} x_{j1} \\ \tilde{u}_1 &= (1 - \rho^2)^{1/2} u_1 \end{aligned} \quad \dots\dots\dots 26)$$

Dapat ditunjukkan bahwa GLS estimator (melalui persamaan 23 dan 25) adalah bersifat BLE, karena standard error dari persamaan yang ditransformasi adalah tidak bersifat autokorelasi dan homokedastis.

3. ***Serial Correlation Robust Standar Error.*** Seperti yang telah diuraikan diatas dampak dari adanya autokorelasi adalah standar error parameter menjadi bias. Dengan demikian salah satu cara untuk mengkoreksi kondisi ini adalah dengan membuat formulasi standar error parameter yang tidak bias (disebut dengan serial correlation robust standard error). Newey-West (1987) telah mengkonstruksi standar error semacam ini melalui tahap-tahap sebagai berikut:

- a. Misalnya kita memiliki model regresi linier k variabel seperti persamaan 22. Estimasi persamaan tersebut dengan OLS yang biasa dan peroleh residual: \hat{u}_t serta standar error dari setiap parameter.
- b. Lakukan regresi auxiliary dengan meregresikan x_1 terhadap variabel bebas lainnya (x_2 s/d x_k) dan peroleh residualnya, \hat{r}_t .
- c. Pilih sembarang integer $g > 0$ dan hitung

$$v = \sum_{t=1}^T \hat{a}_t^2 + 2 \sum_{h=1}^g [1 - h/(g+1)] \left(\sum_{t=h+1}^T \hat{a}_t \hat{a}_{t-h} \right) \dots\dots\dots 27)$$

dimana

$$\hat{a}_t = \hat{r}_t \hat{u}_t; \quad t = 1, 2, \dots, T.$$

- d. Hitung *Serial Correlation Robust Standar Error* dengan formulasi

$$se_{NW}(\hat{\beta}_j) = \left[se_{OLS}(\hat{\beta}_j) / \hat{\sigma}^2 \right]^2 \sqrt{\hat{v}} \dots\dots\dots 28)$$

dimana

$se_{OLS}(\hat{\beta}_j)$ = standar error parameter j dari regresi awal (pers 22)

$\hat{\sigma}^2$ = estimator varians model regresi awal (pers 22)

Contoh 2.

Dengan menggunakan data PRMINWGE, kita akan melakukan regresi antara variabel persentase orang bekerja (prepop) terhadap upah minimal (mincov), gnp negara sendiri (prgnp) dan gnp negara benchmark (usgnp). Dalam bentuk log dan memasukkan unsur trend diperoleh hasil sbb:

Dependent Variable: LOG(PREPOP)
 Method: Least Squares
 Date: 06/13/08 Time: 08:35
 Sample: 1 38
 Included observations: 38

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.663445	1.257829	-5.297577	0.0000
LOG(MINCOV)	-0.212261	0.040152	-5.286398	0.0000
LOG(PRGNP)	0.285238	0.080492	3.543677	0.0012
LOG(USGNP)	0.486049	0.221982	2.189582	0.0357
T	-0.026663	0.004627	-5.762910	0.0000
R-squared	0.889238	Mean dependent var	-0.944074	
Adjusted R-squared	0.875812	S.D. dependent var	0.092978	

S.E. of regression	0.032766	Akaike info criterion	-3.876796
Sum squared resid	0.035428	Schwarz criterion	-3.661325
Log likelihood	78.65913	F-statistic	66.23381
Durbin-Watson stat	1.013707	Prob(F-statistic)	0.000000

Tabel 4. Print Output Regresi Contoh 2

Dengan menggunakan $k=4$ dan $n=38$, dari tabel Durbin-Watson diketahui bahwa $d_l=1.072$ dan $d_u=1.515$. Dapat dilihat bahwa nilai $DW(=1.013) < d_l(=1.072)$, dengan demikian pada model terindikasi adanya autokorelasi. Pengujian dengan menggunakan Breusch-Godfrey Test juga mendukung hal ini (lihat tabel 5).

Dependent Variable: U
Method: Least Squares
Date: 06/13/08 Time: 08:35
Sample (adjusted): 2 38
Included observations: 37 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.850777	1.092688	-0.778609	0.4421
LOG(MINCOV)	0.037500	0.035212	1.064976	0.2951
LOG(PRGNP)	-0.078466	0.070524	-1.112615	0.2744
LOG(USGNP)	0.203933	0.195158	1.044964	0.3041
T	-0.003466	0.004074	-0.850916	0.4013
U(-1)	0.480510	0.166444	2.886917	0.0070
R-squared	0.242364	Mean dependent var	-0.001763	
Adjusted R-squared	0.120165	S.D. dependent var	0.029372	
S.E. of regression	0.027551	Akaike info criterion	-4.198185	
Sum squared resid	0.023530	Schwarz criterion	-3.936955	
Log likelihood	83.66643	F-statistic	1.983349	
Durbin-Watson stat	1.693340	Prob(F-statistic)	0.108930	

Tabel 5. Breusch-Godfrey Test

Kita dapat melakukan koreksi terhadap situasi ini dengan cara menghitung serial correlation robust standar error (Newey West) yang tersedia pada Eviews. Hasil yang diperoleh dirangkum pada tabel 6.

Dependent Variable: LOG(PREPOP)
Method: Least Squares
Date: 06/13/08 Time: 09:36

Sample: 1 38

Included observations: 38

Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.663445	1.533200	-4.346103	0.0001
LOG(MINCOV)	-0.212261	0.045752	-4.639437	0.0001
LOG(PRGNP)	0.285238	0.096419	2.958321	0.0057
LOG(USGNP)	0.486049	0.275990	1.761112	0.0875
T	-0.026663	0.005768	-4.622321	0.0001
R-squared	0.889238	Mean dependent var	-0.944074	
Adjusted R-squared	0.875812	S.D. dependent var	0.092978	
S.E. of regression	0.032766	Akaike info criterion	-3.876796	
Sum squared resid	0.035428	Schwarz criterion	-3.661325	
Log likelihood	78.65913	F-statistic	66.23381	
Durbin-Watson stat	1.013707	Prob(F-statistic)	0.000000	

Table 6. Newey-West Serial Correlation Robust Standard Error Adjustment