KOLINEARITAS GANDA (*MULTICOLLINEARITY*)

Oleh Bambang Juanda

Model: $Y_i = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon_i$

1. Hubungan Linear Sempurna (eksak), Jika

$\sum_{i=1}^{k} C_i X_i = 0$ C_i konstanta yg tdk semuanya 0.

- Mudah diketahui krn tdk ada dugaan parameter koef dgn OLS, juga ragamnya.
- Kolinearitas ganda hanya untuk hubungan linear.
  bukan, Y_i (biaya)=f(output)
  $= \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_p X^p$
- Karena asumsi X nonstokastik *(fixed),* multikolinearitas hanya fenomena sample saja.

2. Hubungan Linear Tidak Sempurna, jika :

$$\sum_{i=1}^{k} C_i X_i + v_i = 0, \quad v_i = \text{sisaan}$$

Konsekuensinya (dan Juga mendeteksinya) :

- Masih bersifat Tak Bias, tapi tidak berarti ragamnya harus kecil

$$E\left(\hat{\beta}\right) = \beta_i \Rightarrow \textbf{Tidak Bias, Tapi}\ \sigma^2_{\hat{\beta}_i} \Rightarrow \textbf{Besar}$$

- Tidak bisa (sulit) memisahkan pengaruh masing-masing peubah bebas,karena $\breve{\ }\ \beta_i$ besar

- Koefisien sulit diinterpretasi (Asumsi Ceteris Paribus?)
- R^2 tinggi, tapi tidak ada (sedikit) koefisien yang nyata, bahkan tandanya bisa terbalik
- Koefisien korelasi sederhana atau $R_j^2\, utk\ X_j = f\left(X \text{ lainnya}\right)$ tinggi.

Note : $r_{x_i x_j} \uparrow$ merupakan syarat cukup bukan syarat perlu Multikolinearitas

- $$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\sum x_j^2} \frac{1}{\left(1 - R_j^2\right)}$$
- VIF : Variance Inflation Factor = $\frac{1}{1 - R_j^2}$ Kenaikan $Var\left(\hat{\beta}_j\right)$ karena korelasi antara peubah penjelas.
- $K = \left(\frac{\lambda_{max}}{\lambda_{min}}\right)^{1/2}$ λ = akar ciri matriks (X\`X)

Aturan praktis : kolinearitas jika K ≥ 30; atau K ≥ $(VIF_{jmax})^{1/2}$……………(Berk 1977)

Mengatasi Kolinearitas Ganda :

1. Manfaatkan Informasi sebelumnya (*a Prior information*)

   Mis: tingkat perubahan konsumsi (Y) terhadap perubahan kekayaan (x_3) sepersepuluh dari tingkat perubahannya terhadap perubahan pendapatan (x_2) → $\beta_3 = 0{,}1\ \beta_2$

   $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$; + $X_i = X_{2i} + 0.1\ X_{3i}$

2. Mengeluarkan peubah dengan kolinearitas tinggi (Kesalahan Spesifikasi)

3. **Transformasi data dengan perbedaan pertama (*first differnce form*) utk data *time series*.**
4. **Menggunakan PCA , *Ridge Regression* → berbias tapi Var(β_i) kecil**
5. **Menggabungkan data '*cross section'* dan '*time series*"**

   Mis :

$$\ln Q_t = \beta_1 + \beta_2 \ln P_t + \beta_3 \ln Y_t + \varepsilon_t$$

   Karena r_{PY} maka **, Misal : β_3 dari SUSENAS dimana harga tidak begitu bervariasi**

$$Q_t^* = \beta_1 + \beta_2 \ln P_t$$

6. **Penambahan data baru →** $\sum_i x_j^2 \uparrow \quad \rightarrow \sigma^2_{\beta_j} \downarrow$
7. **Cek kembali asumsi waktu membuat model (Mis :CRS,IRS,..komponen error) seringkali jika tidak nyata dianggap masalah kolinearitas ganda ?! Dan umumnya di pecahkan dengan mencari prosedur pendugaan**

HETEROSCEDASTICITY (Heteroskedastisitas)

-Sering terjadi dlm data “cross section”
misal: Hubungan pendapatan & pengeluaran RT, Perusahaan

-Biasanya tdk terjadi dlm data “ Time Series”

-Jika Var (ε_i) , E (ε_i^2)=σ_i^2, penduga koefisien OLS tetap tak bias dan konsisten, tapi tidak efisien, bahkan Var($\hat{\beta}$)$_{OLS}$ berbias

Dlm model linear sederhana:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(\beta x_i + \varepsilon_i)}{\sum x_i^2}$$

$$E(\overset{\Lambda}{\beta}) = \beta + \frac{E(\sum x_i \varepsilon_i)}{\sum x_i^{\ 2}} \ ; \qquad \mathrm{Var}(\overset{\Lambda}{\beta})_{OLS} = \frac{\sigma^2}{\sum x_i^{\ 2}}$$

$$c_i = \frac{x_i}{\sum x_i^{\ 2}} \rightarrow Var(\overset{\Lambda}{\beta}) = \sum c_i^{\ 2} \sigma_i^{\ 2} = \frac{\sum x_i^{\ 2} \sigma_i^{\ 2}}{(\sum x_i^{\ 2})^2} \overset{\Lambda}{=} \sum c_i^{\ 2} \overset{\Lambda}{e}_i^{\ 2} (White, 1980)$$

Tetapi tetap inefisien dibandingkan penduga tak bias berikut ini :

Mis. $\sigma_i^2 = \sigma^2 k_i$, k_i: konstanta yg tdk harus sama

$$Var(\hat{\beta}) = \frac{\sigma^2 \sum x_i^2 k_i}{\left(\sum x_i^2\right)^2} = \frac{\sigma^2}{\left(\sum x_i^2\right)} \frac{\sum x_i^2 k_i}{\left(\sum x_i^2\right)}$$

$Jika \dfrac{\sum x_i^2 k_i}{\left(\sum x_i^2\right)} > 1$ Maka $Var\left(\hat{\beta}\right)_{OLS}$ underestimate dan t_{hit} overestimate

Mendeteksi Heteroskedastisitas

- Metode grafik $(\hat{Y}, e_i^{\,2})$ atau $(\mathrm{x}, \mathrm{e}^2)$ diplotkan.
- Uji Heteroskedastisitas :

  Ho : $\sigma_1^{\,2} = \sigma_2^{\,2} = ... = \sigma_N^{\,2}$

  H1 : tergantung pendugaan yang dianggap akan menghasilkan koreksi heteroskedastisitas yang paling diinginkan.

1. UJI PARK (Econometrica, vol 34, No. 4, 1966)

   menganjurkan fungsi :

$$\sigma_i^{\,2} = \sigma^2 x_i^{\,\beta} e^{vi} \rightarrow \beta \text{ signifikan ? masalah pola vi?}$$

$$Plot\, \sigma_{\mathrm{i}}^{\,2} \text{ dengan } \mathrm{e}_{\mathrm{i}}^{\,2}$$

2. UJI GLEJSER (seperti uji Park)

Fungsi Linear $|e_i|$ terhadap : $x_i, \sqrt{x_i}, \dfrac{1}{x_i}, \dfrac{1}{\sqrt{x_i}}, \ldots$

3. UJI Korelasi Pangkat SPEARMAN (urutan e_i dan X_i)

$$r_s = 1 - 6\left[\frac{\sum d_i^{\,2}}{n(n^2-1)}\right] \rightarrow t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^{\,2}}} \approx t_{(n-2)}$$

4. UJI GOLDFELD-QUANDT (JASS, Vol 60, 1965) $\rightarrow H_1 : \sigma_i^2 = cx_i^2$

Cara : 1. Urutkan data peubah bebas x.

2. Keluarkan d pengamatan di tengah (jika tidak ada `natural break`) misal : d= $\frac{1}{5}N$
3. Hitung 2 model regresi terpisah tersebut, dengan

   dbe=(N-d-2k)/2

4. Hitung JKS_1 dan JKS_2
5. Dengan asumsi masing-masing ε_i ~Normal, $\frac{JKS_2}{JKS_1} \approx F_{(dbe_1, dbe)}$

Note : - Jika k>2, urutkan pengamatan berdasarkan salah satu peubah bebasnya

- Supaya kuasa uji tinggi (salah jenis II lebih keci), harus dengan restriksi dimana kedua model regresi tersebut mempunyai parameter koefisien yang sama

5. UJI BREUSCH-PAGAN (Econometrica, Vol 47, 1979)

Misal Model : $Y_i = \alpha + \beta x_i + \varepsilon_i$

dengan asumsi umum: $\sigma_i^2 = f(\gamma + \delta z_i)$

z dapat merupakan peubah bebas x atau suatu kelompok peubah bebas selain x.

> gunakan $\hat{\varepsilon}_i$ untuk menghitung $\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{N}$

> Lakukan Regresi:

$$\frac{\hat{\varepsilon}_i^2}{\hat{\sigma}_i^2} = \gamma + \delta z_i + v_i$$

> Jika ε_i ~Normal, $\frac{JKR}{2} \approx \chi^2_{(p)}$ merupakan statistik uji yang cocok.

Jika nyata (heteroskedastisitas), koreksinya menggunakan peubah z

6. WHITE TEST (Econometrica, Vol. 48, 1980)

- Tidak perlu asumsi kenormalan seperti B-test.
- Dengan asumsi umum : $\hat{\varepsilon}_i^{\,2} = \gamma + \delta z_i + v_i,$

  dengan R^2 sebagai ukuran `goodness of fit`.

  Jika homoskedastisitas, $NR^2 \approx \chi_{(p)}^{\;2}$

CARA MENGATASI (MENGOREKSI) HETEROSKEDASTISITAS

a) Jika σ^2 diketahui $\rightarrow$ *Weighted Least Squares* (MKT tertimbang); kasus khusus dari GLS, yg dpt diturunkan dari fungsi kemungkinan maximum.

Note : $JKS = \sum \frac{1}{\sigma_i^{2}} (Y_i - \hat{\alpha} - \hat{\beta}\, x_i)^2 = \sum (\frac{Y_i - \beta x_i}{\sigma_i})^2$

simpangan (pengamatan) ekstrim dpt timbangan kecil

$$\hat{\beta} = \frac{\sum x_i y_i / \sigma_i^{2}}{\sum x_i^{2} / \sigma_i^{2}} = \frac{\sum x_i^* y_i^*}{\sum x_i^*} \qquad \text{dimana } x_i^* = \frac{x_i}{\sigma_i} \text{ dan } y_i^* = \frac{y_i}{\sigma_i}$$

Cara Transformasi Model: $Y_i = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i \quad | \times \frac{1}{\sigma_\mathrm{i}} |$

$$\frac{Y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{x_{2i}}{\sigma_i} + \ldots + \beta_k \frac{x_{ki}}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}$$

$$Y_i^* = \beta_1 x_1^* + \beta_2 x_{2i}^* + \ldots + \beta_k x_{ki}^* + \varepsilon_i^*$$

$$\rightarrow Var(\varepsilon_i^*) = \frac{1}{\sigma_i^{2}} Var(\varepsilon_i) = 1$$

b. Jika σ_i^2 tidak diketahui → sering menggunakan asumsi tentang σ_i^2

Misal Asumsi : **$Var(\varepsilon_i)=C\ X_{2i}^2$** → lakukan seperti di atas dengan transformasi: **x $(X_{2i})^{-1}$**

$$\frac{Y_i}{x_{2i}} = \beta_1 \frac{1}{x_{2i}} + \beta_2 + \beta_3 \frac{x_{3i}}{x_{2i}} + \ldots + \beta_k \frac{x_{ki}}{x_{2i}} + \frac{\varepsilon_i}{x_{2i}}$$

$$\rightarrow Var(\varepsilon_i *) = c$$

c. Dapat dengan Transformasi Log (memperkecil skala) → kadangkala dapat masalah baru, seperti *Spurious correlation,* kolinearitas.

Teladan : Dengan OLS : $\hat{Y}_i = 0.89 + 0.237x_i$; $R^2 = 0.93$; $F = 252.7$

(4.4) (15.9) ← statistik t

Dengan WLS :

$$\frac{Y_i}{x_i} = \beta * + \alpha * \frac{1}{x_i} + \varepsilon_i *$$

$$\frac{\hat{Y}_i}{x_i} = \underset{(21.3)}{0.249} + \underset{(7.7)}{0.7529}\frac{1}{x_i} \; ; \; R^2 = 0.76 \; ; F = 58.7$$

$R^2_{WLS} < R^2_{OLS}$ → jangan dianggap sebagai indikasi bahwa koreksi heteroskedastisitas kurang baik karena prosedur WLS melibatkan transformasi peubah tak bebas (Y*)

Dengan indikator :

$$\hat{\varepsilon}_i = Y_i - \underbrace{(0.7529 + 0.249X_i)}_{\hat{Y}_i} \rightarrow R^2 = 1 - \frac{JKS}{JKT} \rightarrow \text{tidak harus 0 - 1}$$

$$\rightarrow r_{Y_i \hat{Y}_i}$$