BAB III

REGRESI TERSENSOR (TOBIT)

3.1 Model Regresi Tersensor (Tobit)

Model regresi yang didasarkan pada variabel terikat tersensor disebut model regresi tersensor (tobit). Untuk variabel terikat yang mengelompok akibat adanya batas bawah (tersensor kiri), persamaan model regresi tersebut adalah:

$$Y_i = \begin{cases} Y_i^* = \boldsymbol{X}_i\boldsymbol{\beta} + \varepsilon_i, & \text{jika } Y_i^* > c \\ c_Y, & \text{jika } Y_i^* \leq c \end{cases}$$

dimana:

Y_i : variabel terikat yang diamati

$\boldsymbol{X}_i = [1 \quad X_{1i} \quad X_{2i} \quad \cdots \quad X_{ki}]$: vektor variabel bebas

$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$: vektor koefisien regresi

c : titik sensor

i : $1,2,\cdots,n$

ε_i: error, dan diasumsikan berdistribusi normal dengan rata-rata 0 dan varians σ^2.

Banyak penulis yang mengasumsikan $c = c_Y = 0$ atau $c = c_Y$. Sayangnya penyederhanaan ini dapat membingungkan, karena banyak kasus dimana $c \neq c_Y \neq 0$ (Long, 1997:197).

Karena ε_i diasumsikan berdistribusi normal, maka dapat diasumsikan $Y_i^* \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \sigma^2)$. Peluang suatu observasi tersensor apabila diketahui atau diberikan nilai $\boldsymbol{X}_i$ adalah daerah pada kurva normal yang kurang dari atau sama dengan c:

$$\begin{aligned} P(Y_i \text{ tersensor bila diberikan } \boldsymbol{X}_i) &= P(\text{tersensor}|\boldsymbol{X}_i) \\ &= P(Y_i^* \leq c|\boldsymbol{X}_i) \\ &= P(\boldsymbol{X}_i\boldsymbol{\beta} + \varepsilon_i \leq c|\boldsymbol{X}_i) \\ &= P(\varepsilon_i \leq c - \boldsymbol{X}_i\boldsymbol{\beta}|\boldsymbol{X}_i) \end{aligned}$$

Karena $\varepsilon \sim N(0, \sigma^2)$, maka $\frac{\varepsilon}{\sigma}$ berdistribusi N(0,1), sehingga:

$$P(\text{tersensor}|\boldsymbol{X}_i) = P\left(\frac{\varepsilon_i}{\sigma} \leq \frac{c - \boldsymbol{X}_i\boldsymbol{\beta}}{\sigma}\middle| \boldsymbol{X}_i\right) = \Phi\left(\frac{c - \boldsymbol{X}_i\boldsymbol{\beta}}{\sigma}\right) \tag{3.1}$$

dan

$$P(\text{tidak tersensor}|\boldsymbol{X}_i) = 1 - \Phi\left(\frac{c - \boldsymbol{X}_i\boldsymbol{\beta}}{\sigma}\right) = \Phi\left(\frac{\boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{c}}{\sigma}\right) \tag{3.2}$$

Menurut Joreskog (2002) dalam Larissa, I.D. dan Ispriyanti, Dwi, (2008:136), fungsi kepadatan peluang dari Y_i adalah:

$$f(y_i) = \left[\frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)\right]^{1-k} \left[\Phi\left(\frac{c - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)\right]^{\mathrm{k}} \tag{3.3}$$

dengan

$$k = \begin{cases} 1, \text{jika } y_i = c_y \\ 0, \text{jika lainnya} \end{cases} \tag{3.4}$$

Nilai ekspektasi dari Y_i bila diberikan $\boldsymbol{X}_i$ adalah jumlah komponen dari observasi yang tersensor dan yang tidak tersensor, dengan mensubstitusikan persamaan (3.1), (3.2) dan (2.4), diperoleh nilai ekspektasi dari Y_i bila diberikan $\boldsymbol{X}_i$ sebagai berikut.

$$E[Y_i|\boldsymbol{X}_i] = [P(\text{tidak tersensor}|\boldsymbol{X}_i) \times E(Y_i|Y_i > c, \boldsymbol{X}_i)] + P(\text{tersensor}|\boldsymbol{X}_i) \times c_Y$$

$$= \left[\Phi\left(\frac{\boldsymbol{X}_i\boldsymbol{\beta} - c}{\sigma}\right) \times E(Y_i|Y_i > c, \boldsymbol{X}_i)\right] + \Phi\left(\frac{c - \boldsymbol{X}_i\boldsymbol{\beta}}{\sigma}\right) c_Y$$

$$= \Phi\left(\frac{\boldsymbol{X}_i\boldsymbol{\beta} - c}{\sigma}\right)\left[\boldsymbol{X}_i\boldsymbol{\beta} + \sigma\lambda\left(\frac{c - \boldsymbol{X}_i\boldsymbol{\beta}}{\sigma}\right)\right] + \Phi\left(\frac{c - \boldsymbol{X}_i\boldsymbol{\beta}}{\sigma}\right) c_Y \qquad (3.5)$$

3.2 Penaksiran Parameter

Parameter-parameter dalam model regresi tobit tidak disarankan untuk ditaksir dengan menggunakan metode kuadrat terkecil, karena menyebabkan perhitungan parameter akan cenderung mendekati titik sensor, hubungan variabel menjadi tidak signifikan, atau ketika hubungan tersebut signifikan maka nilainya akan bias serta tidak konsisten karena hasil penelitian yang baru tidak sesuai dengan hasil sebelumnya.

Metode yang disarankan untuk menaksir parameter-parameter regresi tobit adalah metode kemungkinan maksimum. Penaksiran parameter dengan metode kemungkinan maksimum membagi observasi menjadi dua himpunan. Himpunan pertama berisi observasi yang tidak tersensor, dan himpunan kedua berisi observasi yang tersensor. Untuk observasi yang tersensor, nilai spesifik dari Y_i^* tidak diketahui, tetapi peluang tersensor dapat dihitung dan nilai ini digunakan dalam fungsi kemungkinan.

Untuk observasi yang tidak tersensor:

$$Y_i = \boldsymbol{X}_i\boldsymbol{\beta} + \varepsilon_i \quad \text{untuk } Y_i^* > c$$

dimana $\varepsilon \sim N(0, \sigma^2)$. Karena ε_i diasumsikan berdistribusi normal, maka dapat diasumsikan $Y_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \sigma^2)$, sehingga fungsi kemungkinan untuk observasi yang tidak tersensor adalah sebagai berikut.

$$L_{\text{tidak tersensor}}(\boldsymbol{\beta}, \sigma | y_i, \boldsymbol{x}_i) = \prod_{y_i > c} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{y_i - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)^2\right]$$

$$= \prod_{y_i > c} \frac{1}{\sigma} \phi\left(\frac{y_i - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

$$\ln L_{\text{tidak tersensor}}(\boldsymbol{\beta}, \sigma | y_i, \boldsymbol{x}_i) = \sum_{y_i > c} \ln \frac{1}{\sigma} \phi\left(\frac{y_i - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

Sementara untuk observasi yang tersensor, $\boldsymbol{X}_i$ diketahui dan $Y_i = c_Y$, sehingga berdasarkan persamaan (3.3) dan (3.4) diperoleh fungsi kemungkinan:

$$L_{\text{tersensor}}(\boldsymbol{\beta}, \sigma | \boldsymbol{x}_i) = \prod_{y_i = c_y} \Phi\left(\frac{c - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

$$\ln L_{\text{tersensor}}(\boldsymbol{\beta}, \sigma | \boldsymbol{x}_i) = \sum_{y_i = c_y} \ln \Phi\left(\frac{c - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

Apabila observasi yang tidak tersensor dan yang tersensor digabungkan, maka diperoleh:

$$\ln L(\boldsymbol{\beta}, \sigma | y_i, \boldsymbol{x}_i) = \sum_{y_i > c} \ln \frac{1}{\sigma} \phi\left(\frac{y_i - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right) + \sum_{y_i = c_y} \ln \Phi\left(\frac{c - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

$$= \sum_{y_i > c} \ln \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{y_i - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)^2\right)\right] + \sum_{y_i = c_y} \ln \Phi\left(\frac{c - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

$$= -\frac{1}{2}\sum_{y_i>c}\left[\ln 2\pi + \ln\sigma^2 + \left(\frac{y_i - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)^2\right] + \sum_{y_i=c_y} \ln\Phi\left(\frac{c - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

Misalkan $\boldsymbol{\gamma} = \frac{\boldsymbol{\beta}}{\sigma}$ dan $\theta = \frac{1}{\sigma}$, maka diperoleh:

$$\ln L(\boldsymbol{\gamma},\theta|y_i,\boldsymbol{x}_i) = -\frac{1}{2}\sum_{y_i>c}[\ln 2\pi - \ln\theta^2 + (\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma})^2] + \sum_{y_i=c_y}\ln\Phi(\theta c - \boldsymbol{x}_i\boldsymbol{\gamma})$$

$\frac{\partial L(\boldsymbol{\gamma},\theta|y_i,\boldsymbol{x}_i)}{\partial\theta} = 0$ maka:

$$\frac{\partial}{\partial\theta}\left[-\frac{1}{2}\sum_{y_i>c}(\ln 2\pi - \ln\theta^2 + (\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma})^2) + \sum_{y_i=c_y}\ln\Phi(\theta c - \boldsymbol{x}_i\boldsymbol{\gamma})\right] = 0$$

$$-\frac{1}{2}(-2)\sum_{y_i>c}\frac{1}{\theta} - \frac{1}{2}(2)\sum_{y_i>c} y_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma}) + \sum_{y_i=c_y}\frac{c\,\phi(\theta c - \boldsymbol{x}_i\boldsymbol{\gamma})}{\Phi(\theta c - \boldsymbol{x}_i\boldsymbol{\gamma})} = 0$$

$$\sum_{y_i>c}\frac{1}{\theta} - \sum_{y_i>c} y_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma}) + \sum_{y_i=c_y}\frac{c\,\phi(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)}{1 - \Phi(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)} = 0$$

$$\sum_{y_i>c}\frac{1}{\theta} - \sum_{y_i>c} y_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma}) + \sum_{y_i=c_y} c\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) = 0$$

$$\sum_{y_i>c}\left[\frac{1}{\theta} - y_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma})\right] + c\sum_{y_i=c_y}\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) = 0 \quad (3.6)$$

$\frac{\partial \ln L(\boldsymbol{\gamma},\theta|y_i,\boldsymbol{x}_i)}{\partial\boldsymbol{\gamma}} = 0$ maka:

$$\frac{\partial}{\partial\boldsymbol{\gamma}}\left[-\frac{1}{2}\sum_{y_i>c}(\ln 2\pi - \ln\theta^2 + (\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma})^2) + \sum_{y_i=c_y}\ln\Phi(\theta c - \boldsymbol{x}_i\boldsymbol{\gamma})\right] = 0$$

$$-\frac{1}{2}(-2)\sum_{y_i>c}\boldsymbol{x}_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma}) - \sum_{y_i=c_y}\frac{\boldsymbol{x}_i\,\phi(\theta c - \boldsymbol{x}_i\boldsymbol{\gamma})}{\Phi(\theta c - \boldsymbol{x}_i\boldsymbol{\gamma})} = 0$$

$$\sum_{y_i>c} \boldsymbol{x}_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma}) - \sum_{y_i=c_y} \boldsymbol{x}_i\,\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) = 0 \tag{3.7}$$

Untuk mencari parameter-parameter $\boldsymbol{\gamma}$ dan θ yang memaksimumkan fungsi kemungkinan, persamaan (3.6) dan (3.7) diselesaikan dengan bantuan metode Newton Raphson.

3.3 Metode Newton Raphson Untuk Regresi Tersensor

Persamaan (3.6) dan (3.7) diselesaikan dengan menggunakan metode Newton Raphson. Misalkan:

$$T_m = \begin{bmatrix} \sum_{y_i>c}\left[\frac{1}{\theta} - y_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma})\right] + c\sum_{y_i=c_y}\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) \\ \sum_{y_i>c}\boldsymbol{x}_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma}) - \sum_{y_i=c_y}\boldsymbol{x}_i\,\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) \end{bmatrix}$$

Maka turunan dari T_m adalah:

$$\frac{\partial^2 \ln L(\boldsymbol{\gamma},\theta|y_i,\boldsymbol{x}_i)}{\partial\theta\partial\theta} = \frac{\partial}{\partial\theta}\left[\sum_{y_i>c}\left[\frac{1}{\theta} - y_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma})\right] + c\sum_{y_i=c_y}\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\right]$$

$$= -\sum_{y_i>c}\left(\frac{1}{\theta^2} + {y_i}^2\right) + \frac{\partial}{\partial\theta}\left[c\sum_{y_i=c_y}\frac{\phi(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)}{1 - \Phi(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)}\right]$$

$$= -\sum_{y_i>c}\left(\frac{1}{\theta^2} + {y_i}^2\right) + \frac{\partial}{\partial\theta}\left[c\sum_{y_i=c_y}\frac{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)^2\right]}{1 - \Phi(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)}\right]$$

$$= -\sum_{y_i>c}\left(\frac{1}{\theta^2} + {y_i}^2\right) + c^2\sum_{y_i=c_y}\left[(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) + \lambda^2(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\right]$$

$$\frac{\partial^2 \ln L(\boldsymbol{\gamma},\theta|y_i,\boldsymbol{x}_i)}{\partial\theta\partial\boldsymbol{\gamma}} = \frac{\partial}{\partial\theta}\left[\sum_{y_i>c} \boldsymbol{x}_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma}) - \sum_{y_i=c_y} \boldsymbol{x}_i\,\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\right]$$

$$= \sum_{y_i>c} \boldsymbol{x}_i y_i + c\sum_{y_i=c_y} \boldsymbol{x}_i[(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) + \lambda^2(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)]$$

$$\frac{\partial^2 \ln L(\boldsymbol{\gamma},\theta|y_i,\boldsymbol{x}_i)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}} = \frac{\partial}{\partial\boldsymbol{\gamma}}\left[\sum_{y_i>c} \boldsymbol{x}_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma}) - \sum_{y_i=c_y} \boldsymbol{x}_i\,\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\right]$$

$$= -\sum_{y_i>c} \boldsymbol{x}_i\boldsymbol{x}_i{}' - \frac{\partial}{\partial\boldsymbol{\gamma}}\left[\sum_{y_i=c_y} \boldsymbol{x}_i\frac{\phi(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)}{1 - \Phi(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)}\right]$$

$$= -\sum_{y_i>c} \boldsymbol{x}_i\boldsymbol{x}_i{}' - \frac{\partial}{\partial\boldsymbol{\gamma}}\left[\sum_{y_i=c_y} \boldsymbol{x}_i\frac{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)^2\right]}{1 - \Phi(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)}\right]$$

$$= -\sum_{y_i>c} \boldsymbol{x}_i\boldsymbol{x}_i{}' + \sum_{y_i=c_y} \{\boldsymbol{x}_i\boldsymbol{x}_i{}'[(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) - \lambda^2(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)]\}$$

$$\frac{\partial^2 \ln L(\boldsymbol{\gamma},\theta|y_i,\boldsymbol{x}_i)}{\partial\boldsymbol{\gamma}\partial\theta} = \frac{\partial}{\partial\boldsymbol{\gamma}}\left[\sum_{y_i>c}\left[\frac{1}{\theta} - y_i(\theta y_i - \boldsymbol{x}_i\boldsymbol{\gamma})\right] + c\sum_{y_i=c_y} \lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\right]$$

$$= \sum_{y_i>c} y_i\,\boldsymbol{x}_i + c\sum_{y_i=c_y} \{\boldsymbol{x}_i[(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) - \lambda^2(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)]\}$$

Misalkan matriks turunan dari T_m adalah H_m, maka:

$$H_m = \begin{bmatrix} \frac{\partial^2 \ln L(\boldsymbol{\gamma},\theta|y_i,\boldsymbol{x}_i)}{\partial\theta\partial\theta} & \frac{\partial^2 \ln L(\boldsymbol{\gamma},\theta|y_i,\boldsymbol{x}_i)}{\partial\theta\partial\boldsymbol{\gamma}} \\ \frac{\partial^2 \ln L(\boldsymbol{\gamma},\theta|y_i,\boldsymbol{x}_i)}{\partial\boldsymbol{\gamma}\partial\theta} & \frac{\partial^2 \ln L(\boldsymbol{\gamma},\theta|y_i,\boldsymbol{x}_i)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}} \end{bmatrix}$$

atau $H_m = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$, dengan:

$$h_{11} = -\sum_{y_i > c}\left(\frac{1}{\theta^2} + {y_i}^2\right) + c^2 \sum_{y_i = c_y}[(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) + \lambda^2(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)]$$

$$h_{12} = \sum_{y_i > c} \boldsymbol{x}_i y_i + c \sum_{y_i = c_y} \boldsymbol{x}_i[(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) + \lambda^2(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)]$$

$$h_{21} = \sum_{y_i > c} y_i\, \boldsymbol{x}_i + c \sum_{y_i = c_y} \{\boldsymbol{x}_i[(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) - \lambda^2(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)]\}$$

$$h_{22} = -\sum_{y_i > c} \boldsymbol{x}_i{\boldsymbol{x}_i}' + \sum_{y_i = c_y} \{\boldsymbol{x}_i{\boldsymbol{x}_i}'[(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)\lambda(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c) - \lambda^2(\boldsymbol{x}_i\boldsymbol{\gamma} - \theta c)]\}$$

Berdasarkan persamaan (2.11), iterasi metode Newton Raphson untuk menaksir parameter dengan metode kemungkinan maksimum adalah:

$$\begin{bmatrix}\hat{\theta}_{m+1}\\ \hat{\boldsymbol{\gamma}}_{m+1}\end{bmatrix} = \begin{bmatrix}\hat{\theta}_{m}\\ \hat{\boldsymbol{\gamma}}_{m}\end{bmatrix} - H_m^{-1}T_m$$

Jika nilai taksiran parameter $\hat{\theta}$ dan $\hat{\boldsymbol{\gamma}}$ telah diperoleh, maka dengan menggunakan persamaan $\hat{\boldsymbol{\beta}} = \frac{\hat{\gamma}}{\hat{\theta}}$ dan $\hat{\sigma} = \frac{1}{\hat{\theta}}$ nilai taksiran parameter $\hat{\boldsymbol{\beta}}$ dan $\hat{\sigma}$ dapat diperoleh.

3.4 Pelanggaran Asumsi

Penaksir kemungkinan maksimum untuk model regresi tersensor mengasumsikan bahwa error berdistribusi normal dan homoskedastis (variansi tetap/konstan). Dalam model regresi linear, jika asumsi ini dilanggar maka penaksir masih konsisten tetapi tidak efisien, namun tidak demikian dalam kasus regresi tersensor. Hasil penelitian Maddala dan Nelson (1975), dan Arabmazar dan Schmidt (1981) menunjukan bahwa apabila asumsi normalitas dan homoskedastisitas dari error dilanggar, maka pelanggaran ini akan menyebabkan penaksir kemungkinan maksimum menjadi tidak konsisten (Long, 1997:206).

3.4.1 Uji Normalitas Jarque-Bera

Untuk mengetahui apakah asumsi normalitas dari error dipenuhi, maka perlu dilakukan uji normalitas. Tetapi nilai error tidak diketahui sebelum dilakukan penaksiran parameter, sementara berdasarkan persamaan model regresi tersensor (tobit):

$$Y_i = \begin{cases} Y_i^* = \boldsymbol{X}_i\boldsymbol{\beta} + \varepsilon_i, & \text{jika } Y_i^* > c \\ c_Y, & \text{jika } Y_i^* \leq c \end{cases}$$

sehingga uji normalitas dilakukan terhadap data Y_i^* .

Jarque-Bera adalah statistik uji untuk mengetahui apakah data berdistribusi normal atau tidak. Uji ini mengukur perbedaan kemiringan (*skewness*) dan kurtosis data. Statistik Jarque-Bera mengikuti distribusi Chi-Kuadrat dengan derajat kebebasan dua.

Langkah-langkah dalam melakukan uji Jarque-Bera:

1. Perumusan hipotesis

   H_0 : Data berdistribusi normal

   H_1 : Data berdistribusi tidak normal

2. Besaran yang diperlukan

$$\frac{N-k}{6}\left(S^2 + \frac{(K-3)^2}{4}\right)$$

   dimana S adalah koefisien kemiringan, K adalah koefisien kurtosis, dan k adalah banyak koefisien yang digunakan dalam persamaan.

3. Statistik Uji

$$JB = \frac{N-k}{6}\left(S^2 + \frac{(K-3)^2}{4}\right)$$

4. Kriteria Pengujian

   Dengan mengambil taraf nyata α, maka dari Tabel Distribusi Chi-Kuadrat diperoleh nilai $\chi^2_{(1-\alpha;2)}$

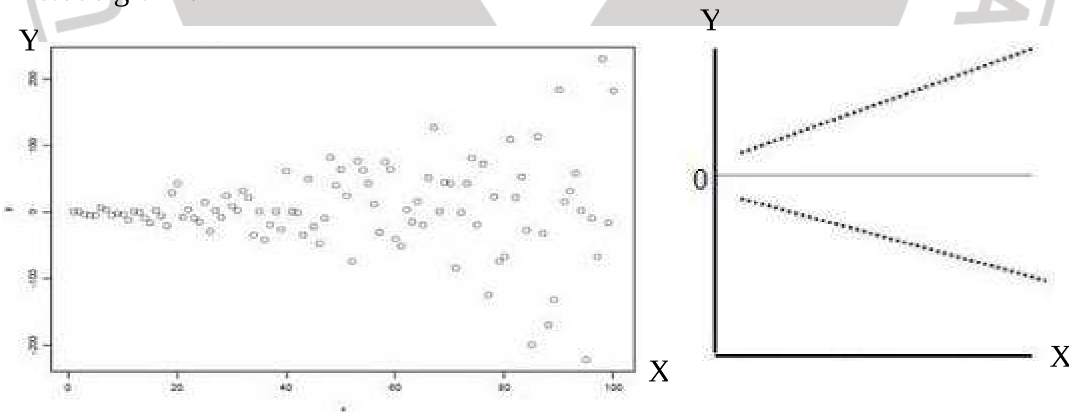
   H_0 ditolak jika $JB > \chi^2_{(1-\alpha;2)}$ atau jika Probabilitas (JB) < α

5. Kesimpulan

   Penafsiran dari H_0 ditolak atau diterima.

3.4.2 Uji Heteroskedastisitas

Uji heteroskedastisitas dilakukan untuk mengetahui apakah error memiliki variansi yang tetap (homoskedastisitas) atau tidak. Salah satu metode untuk menguji heteroskedastisitas yang sering digunakan dan relatif mudah adalah metode grafik.

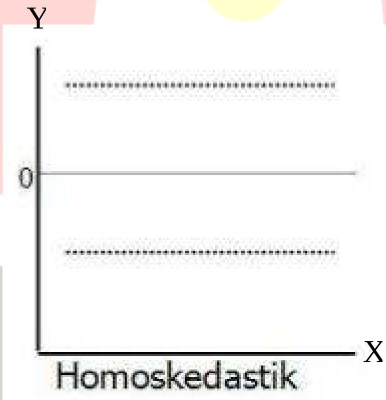


Gambar 3.1

Indikasi Adanya Heteroskedastisitas

Mendeteksi heteroskedastisitas dengan metode grafik adalah dengan melihat ada tidaknya pola tertentu pada grafik, dengan sumbu-X adalah nilai prediksi variabel terikat yang telah distandarkan, dan sumbu-Y adalah nilai error yang telah distandarkan (Ariyoso, 2009).

Jika dari grafik tersebut terlihat adanya pola tertentu, seperti titik-titik yang ada membentuk suatu pola tertentu yang teratur (bergelombang, melebar, kemudian menyempit), maka dapat disimpulkan bahwa terjadi heteroskedastisitas. Tetapi jika tidak ada pola yang jelas, serta titik-titik menyebar di atas dan di bawah angka 0 pada sumbu-Y, maka tidak terjadi heteroskedastisitas (Ariyoso, 2009).



Gambar 3.2

Indikasi Homoskedastisitas

Selain dari variabel error yang telah distandarkan dan nilai-nilai prediksi variabel terikat yang telah distandarkan, uji heteroskedatisitas juga dapat dilakukan dengan menampilkan grafik sebar (*scatter plot*) dari variabel error kuadrat dan variabel bebas. Variabel error dapat diperoleh apabila proses penaksiran parameter telah dilakukan. Oleh karena itu uji heteroskedastisitas dengan metode grafik harus dimulai dengan menjalankan proses penaksiran parameter terlebih dahulu (Winarno, 2009:5.9).

3.5 Uji Signifikansi Parameter

Uji signifikansi parameter dilakukan untuk mengetahui apakah variabel bebas secara signifikan berpengaruh terhadap variabel terikat, baik secara simultan maupun secara parsial. Karena uji *t* dan uji *F* hanya dapat digunakan untuk menguji signifikansi parameter pada model regresi linear (konteks linear dalam parameter), maka untuk model regresi secara umum dibutuhkan metode yang dapat digunakan untuk menguji signifikansi parameter model regresi, baik linear maupun nonlinear. Untuk keperluan tersebut dapat digunakan uji Wald dan uji perbandingan kemungkinan (*likelihood ratio test*).

3.5.1 Uji Perbandingan Kemungkinan

Untuk melihat peranan variabel bebas secara bersama-sama terhadap variabel terikat, dilakukan uji perbandingan kemungkinan (*likelihood ratio test*). Uji ini membandingkan dua buah model, yang pertama adalah model penuh (model yang memuat semua β_i), dan yang kedua adalah model alternatif (model yang hanya memuat konstanta saja (β_0)). Uji perbandingan kemungkinan didasarkan pada nilai perbandingan kemungkinan (*likelihood ratio*), yang menunjukkan berapa kali data lebih mungkin berada dalam model yang satu dibandingkan model yang lain.

Langkah-langkah dalam melakukan uji perbandingan kemungkinan:

1. Perumusan hipotesis

   $H_0 : \beta_1 = \cdots = \beta_k = 0$

   H_1 : Paling sedikit terdapat satu $\beta_j \neq 0$

2. Besaran yang diperlukan

$$-2[\ln(L_{alternatif}) - \ln(L_{full})] = -2\ln\left(\frac{L_{alternatif}}{L_{full}}\right)$$

dengan

$L_{alternatif}$: nilai kemungkinan maksimum tanpa memuat variabel bebas

L_{full} : nilai kemungkinan maksimum yang memuat semua variabel bebas

3. Statistik Uji

$$\chi^2{}_{hitung} = -2\ln\left(\frac{L_{alternatif}}{L_{full}}\right)$$

4. Kriteria Pengujian

Dengan mengambil taraf nyata α, maka dari Tabel Distribusi Chi-Kuadrat dengan peluang $= 1 - \alpha$ dan $dk = k$ diperoleh $\chi^2{}_{(1-\alpha;k)}$

H_0 ditolak jika $\chi^2{}_{hitung} > \chi^2{}_{(1-\alpha;k)}$ atau jika probabilitas $\chi^2{}_{hitung} < \alpha$

5. Kesimpulan

Penafsiran dari H_0 ditolak atau diterima.

3.5.2 Uji Wald

Untuk melihat peranan suatu variabel bebas terhadap variabel terikat, dilakukan uji signifikansi parameter parsial, yaitu dengan menggunakan uji Wald. Langkah-langkah dalam melakukan uji Wald:

1. Perumusan hipotesis

H_0 : $\beta_i = 0$ (koefisien β_i tidak signifikan secara statistik)

H_1 : $\beta_i \neq 0$ (koefisien β_i signifikan secara statistik)

dengan $i = 0, 1, 2, \cdots, k$

2. Besaran yang diperlukan

$$\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

dimana $\hat{\beta}_i$ adalah penaksir parameter β_i dan $SE(\hat{\beta}_i)$ adalah *standard error* dari β_i

3. Statistik Uji

$$Z_{hitung} = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

4. Kriteria Pengujian

Dengan mengambil taraf nyata α, maka dari Tabel Distribusi Normal Baku diperoleh $Z_{\alpha/2}$. H_0 ditolak jika $Z_{hitung} < -Z_{\alpha/2}$ atau $Z_{hitung} > Z_{\alpha/2}$ atau jika probabilitas $Z_{hitung} < \alpha$

5. Kesimpulan

Penafsiran dari H_0 ditolak atau diterima.

3.6 Kebaikan Model

Model yang sudah dianalisis harus diperiksa apakah kualitasnya sudah baik. Pada umumnya dalam analisis regresi digunakan koefisien determinasi R^2 untuk memperoleh kebaikan model. Rumus R^2 untuk regresi tobit dihitung dengan cara yang sama seperti apabila menggunakan metode kuadrat terkecil (Bierens, 2004).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \hat{U}_i^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

dimana $\widehat{U}_i$ adalah residual yang dihitung dengan menggunakan rumus:

$$\widehat{U}_i = Y_i - E\left(Y_i|\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}}, \hat{\sigma}\right)$$

$$= Y_i - \Phi\left(\frac{\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} - c}{\hat{\sigma}}\right)\left[\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \hat{\sigma}\lambda\left(\frac{\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} - c}{\hat{\sigma}}\right)\right] + \Phi\left(\frac{c - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}}{\hat{\sigma}}\right)c_Y$$

Nilai R^2 selalu berada diantara 0 dan 1. Nilai R^2 menunjukan seberapa besar garis regresi dapat menjelaskan hubungan antara variabel terikat dengan variabel bebas. Jika nilai $R^2 = 0$, artinya variasi dari variabel terikat tidak dapat diterangkan oleh variabel bebas sama sekali. Sementara bila nilai $R^2 = 1$, artinya variasi dari variabel terikat, 100% dapat diterangkan oleh variabel bebas. Kebaikan model juga dapat dilakukan dengan melihat grafik data sesungguhnya dan data menurut perhitungan regresi, serta selisih atau perbedaan di antara keduanya.

3.7 Interpretasi Model

Dengan menggunakan program komputer yang modern, penaksir kemungkinan maksimum untuk model regresi tersensor (tobit) dapat dengan mudah diperoleh, seperti memperoleh penaksir kuadrat terkecil untuk model regresi linear. Terlebih lagi hasil output yang dihasilkan hampir sama, baik untuk regresi tersensor maupun regresi linear. Kesulitan yang dihadapi adalah menginterpretasikan hasilnya.

Dari model regresi:

$$Y_i = \begin{cases} Y_i^* = \boldsymbol{X}_i\boldsymbol{\beta} + \varepsilon_i, & \text{jika } Y_i^* > c \\ c_Y, & \text{jika } Y_i^* \leq c \end{cases}$$

dengan $\boldsymbol{X}_i = [1 \quad X_{i1} \quad X_{i2} \quad \cdots \quad X_{ik}]$ dan $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$

terlihat bahwa β_j mengukur pengaruh parsial dari X_j pada $E[Y_i^*|\boldsymbol{X}]$. Karena error diasumsikan berdistribusi normal dengan rata-rata 0, maka diperoleh nilai harapan dari Y_i^* jika diketahui $\boldsymbol{X}_i$:

$$\begin{aligned} E[Y_i^*|\boldsymbol{X}_i] &= E[\boldsymbol{X}_i\boldsymbol{\beta} + \varepsilon_i] \\ &= \boldsymbol{X}_i\boldsymbol{\beta}E[\varepsilon_i] \\ &= \boldsymbol{X}_i\boldsymbol{\beta} \end{aligned}$$

dan pengaruh parsial dari X_j dapat diperoleh dengan menggunakan kalkulus, yaitu:

$$\frac{\partial E[Y_i^*|\boldsymbol{X}]}{\partial X_j} = \beta_j$$

Bila variabel bebas X_j kontinu, maka dapat dinyatakan bahwa untuk kenaikan satu unit X_j, akan memberikan perubahan nilai harapan sebesar β_j unit pada Y^*, jika variabel lainnya konstan. Sementara bila variabel bebas X_j merupakan variabel dikotomi (variabel boneka), maka dapat dinyatakan bahwa kemunculan/memiliki karakteristik X_j memberikan perubahan nilai sebesar β_j unit pada Y^*, jika variabel lainnya konstan.

Telah disebutkan sebelumnya bahwa regresi tersensor (tobit) adalah perluasan dari regresi probit. Oleh karena itu bila variabel yang akan dijelaskan adalah variabel Y, maka seperti di dalam regresi probit, nilai $P(Y_i = c_Y|\boldsymbol{X}_i)$ dan

$P(Y_i > c|\boldsymbol{X}_i)$ dapat ditaksir (pada persamaan 3.1 dan 3.2). Sementara persamaan 3.5 menjelaskan nilai harapan dari Y sebagai fungsi dari $\boldsymbol{X}$.