ESTIMASI PARAMETER REGRESI TERPOTONG KIRI DENGAN METODE MAKSIMUM LIKELIHOOD

Ratna Pratiwi dan Dwi Ispriyanti
Jurusan Matematika FMIPA Universitas Diponegoro
Jl. Prof. H. Soedarto SH Tembalang, Semarang 50275

Abstract. Left truncated regression model is a regression model which the parameters prevailed for the dependent variable values Y > *a*, with *a* is a left truncation point that is choosen according to the experiment goals, and *a* is also the observation object itself at once. Thus, truncated regression model is more correct if it is used for the experiment that oriented on a specific characteristic of the observation object, that is dependent variable. The distribution that is used for this regression model is normal left truncated distribution. Estimation of regression parameters used Maksimum Likelihood method with Newton's iteratif method.

Keywords: left trucated regression, Maksimum likehood method

1. PENDAHULUAN

Analisis regresi merupakan salah satu teknik statistik yang digunakan untuk menyelidiki hubungan antara dua variabel atau lebih, sekaligus merumuskan model matematisnya. Model regresi linier adalah suatu persamaan yang berhubungan dengan nilai satu variabel dependen (Y) yang didasarkan pada satu atau beberapa variabel independen (X) yang diketahui.

Dalam suatu penelitian, seringkali dijumpai bahwa variabel dependen Y perlu dibatasi untuk tujuan tertentu. Adanya pembatasan atau pemotongan terhadap suatu nilai tertentu terhadap variabel dependen Y**,** sebut saja *a,* mengakibatkan distribusi data tersebut berubah. Distribusi yang akan digunakan adalah distribusi normal terpotong yang selanjutnya menjadi dasar dalam pembentukan model regresi terpotong (*truncated regression*). Sebelum pembentukan model regresi terpotong, terlebih dahulu harus diketahui mean terpotong (*truncated mean*) dari distribusi normal terpotong. Hal ini akan mempermudah langkah selanjutnya yaitu melakukan estimasi terhadap parameter-parameternya.

Dalam penulisan ini, akan dibahas mengenai estimasi model regresi terpotong kiri (*left truncated regression*), yaitu model yang parameter-parameternya berlaku untuk nilai-nilai Y > *a*, dengan *a* suatu konstanta yang merupakan titik potong kiri bagi variabel dependen Y**.** Parameter regresi tersebut dapat diestimasi menggunakan metode Maksimum Likelihood. Menurut [5], efek pemotongan terjadi ketika data sampel yang diambil berasal dari suatu bagian populasi yang besar. Sehingga untuk membentuk model regesi terpotong sebaiknya menggunakan data yang berjumlah besar.

2. METODE MAKSIMUM LIKELIHOOD

Metode Maksimum Likelihood merupakan salah satu metode yang paling baik untuk memperoleh estimator. Diberikan X variabel random distribusi probabilta $f(x|\theta)$, dimana parameter tunggal θ tidak diketahui. Jika diberikan X_1, X_2, …, X_n adalah sampel random dari populasi dengan densitas $f(x_i|\theta_1, \theta_2,...,\theta_k)$. Maka fungsi Likelihood didefinisikan:

$$L(\theta_1, \theta_2,...,\theta_k|X) = \prod_{i=1}^{n} f(x_i|\theta_1,\theta_2,\cdots,\theta_k)$$

Diberikan Y_1, Y_2, …, Y_n adalah sampel random dari distribusi normal dengan fungsi densitas

$$f(y_i \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2}$$

maka fungsi Likelihoodnya adalah

$$L(\mu,\sigma^2 \mid \mathrm{Y}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(y_i-\mu)^2\right)$$
$$= \sigma^{-n}(2\pi)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right).$$

Pandang model regresi linier berganda dalam bentuk matriks

$$\mathrm{Y} = \mathrm{X}\beta + \varepsilon$$

$\varepsilon \sim N(0, \sigma^2)$, X fixed dan β konstan, maka $E(\mathrm{Y}) = \mathrm{X}\beta$, sehingga $\mathrm{Y} \sim N(\mathrm{X}\beta, \sigma^2)$.
Dengan substistusi $\mu = E(\mathrm{Y}) = \mathrm{X}\beta$, maka fungsi Likelihood menjadi

$$L(\beta,\sigma^2 \mid \mathrm{Y}) = \sigma^{-n}(2\pi)^{-n/2}. \exp\left(-\frac{1}{2\sigma^2}(\mathrm{Y}-\mathrm{X}\beta)'(\mathrm{Y}-\mathrm{X}\beta)\right) \quad (2.1).$$

Untuk memperoleh estimasi dari β dan σ^2, maka persamaan (2.1) ekivalen dengan

$$\log L(\beta,\sigma^2 \mid \mathrm{Y}) = -n\log\sigma - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}(\mathrm{Y}-\mathrm{X}\beta)'(\mathrm{Y}-\mathrm{X}\beta) \quad (2.2)$$

selanjutnya

$$\frac{\partial}{\partial\beta}\log L(\beta,\sigma^2 \mid \mathrm{Y}) = 0 \Rightarrow -\frac{1}{2\sigma^2}(-2\mathrm{X'Y} + 2\hat{\beta}\mathrm{X'X}) = 0 \quad (2.3)$$

dan

$$\frac{\partial}{\partial\sigma}\log L(\beta,\sigma^2 \mid \mathrm{Y}) = 0 \Rightarrow -\frac{n}{\sigma} + \frac{1}{\sigma^3}(\mathrm{Y}-\mathrm{X}\beta)'(\mathrm{Y}-\mathrm{X}\beta) = 0 \quad (2.4)$$

Dari penyelesaian persamaan (2.3) dan (2.4) diperoleh

$$\hat{\beta} = (\mathrm{X'X})^{-1}\mathrm{X'Y},$$
$$\hat{\sigma}^2 = \frac{1}{n}(\mathrm{Y}-\mathrm{X}\beta)'(\mathrm{Y}-\mathrm{X}\beta).$$

3. DISTRIBUSI NORMAL TERPOTONG KIRI

Distribusi normal terpotong adalah distribusi normal dengan nilai variabel random X terbatas pada interval $[a, b]$ atau $a \le X \le b$. Titik a adalah titik terpotong di sebelah kiri (disebut juga titik terpotong kiri) dan titik b adalah titik terpotong kanan (disebut juga titik terpotong kanan). Menurut [3], didefinisikan fungsi densitas normal terpotong sebagai berikut.

$$f(x \mid a < x < b) = \frac{f(x)}{P(a < x < b)}, \quad (3.1)$$

dimana $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ adalah fungsi densitas normal.

Diberikan transformasi $z = \frac{x-\mu}{\sigma}$ sedemikian sehingga

$$f(x \mid a < x < b) = \frac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \quad (3.2)$$

dimana $\phi(.)$ dan $\Phi(.)$ adalah fungsi densitas dan fungsi distribusi dari distribusi normal baku.

Teorema 3.1.
Jika variabel random X mempunyai fungsi densitas $f(x)$ dan a konstanta, maka fungsi densitas terpotong pada $x > a$ adalah

$$f(x \mid x > a) = \frac{f(x)}{P(x > a)} \quad (3.3)$$

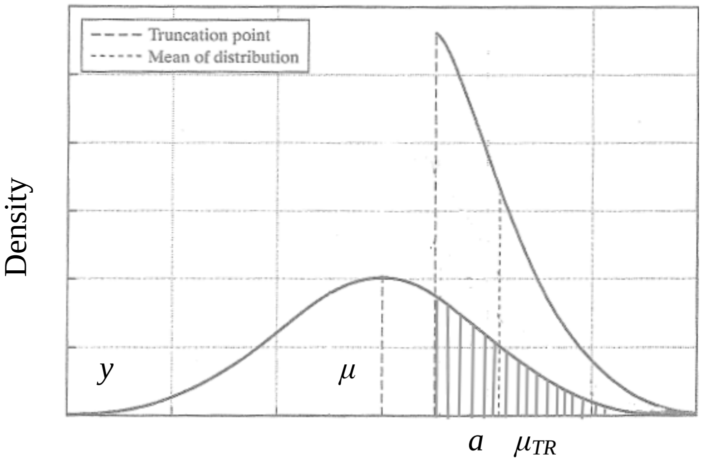
Bukti.
Karena pemotongan hanya dilakukan pada titik $x > a$, maka titik b dianggap bernilai positif tak berhingga, sehingga dari persamaan (3.1) dan (3.2) diperoleh

$$f(x \mid x > a) = \frac{f(x)}{\Phi(\infty) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

$$= \frac{f(x)}{1 - P(x \le a)}$$

$$= \frac{f(x)}{P(x > a)} \qquad \blacksquare$$

Distribusi normal yang terpotong tunggal di sebelah kiri oleh suatu konstanta *a* sebagai titik pemotongan disebut juga distribusi normal terpotong kiri (Gambar 1).



Gambar 1. Distribusi Normal Terpotong Kiri

Teorema 3.2.
Jika $X \sim N(\mu, \sigma^2)$ dan *a* konstanta, maka mean dan varian dari distribusi normal terpotong kiri adalah

$$E(X \mid X > a) = \mu + \sigma\lambda(\alpha), \qquad (3.4)$$
$$Var(X \mid X > a) = \sigma^2[1 - \delta(\alpha)],$$

dengan

$$\alpha = \frac{x - \mu}{\sigma},$$
$$\lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)},$$
$$\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha] = \lambda'(\alpha).$$

Mean dari distribusi terpotong kiri disebut mean terpotong kiri, sedangkan variannya disebut varian terpotong kiri. Harga mean terpotong selalu lebih besar dari harga mean yang tidak terpotong, namun sebaliknya harga varian terpotong selalu lebih kecil dari harga varian yang tidak terpotong.

Fungsi $\lambda(\alpha) = \frac{\phi(\alpha)}{1 - \Phi(\alpha)}$ disebut fungsi Hazard dari distribusi normal. Dalam literatur *micro-econometrics*, fungsi ini biasanya disebut dengan *inverse Mills ratio*.

4. MODEL REGRESI TERPOTONG KIRI

Diberikan model regresi berganda dalam bentuk skalar

$$Y_i = X_i\beta + \varepsilon_i,$$

dengan

$$X_i = \begin{bmatrix} 1 & X_{i1} & X_{i2} & \cdots & X_{ip} \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \qquad i = 1, 2, \ldots, n \text{ observasi.}$$

Diasumsikan $\varepsilon_i \sim N(0, \sigma^2)$ dan $\mathbf{X_i}$ fixed maka $E(Y_i) = X_i\beta$, sehingga $Y_i \sim N(X_i\beta, \sigma^2)$.

Dalam regresi terpotong kiri, dikehendaki suatu distribusi dengan Y_i lebih besar dari pemotongan suatu nilai *a*, maka dari persamaan (3.4) diperoleh mean dan varian terpotong kiri

$$E(Y_i \mid Y_i > a) = X_i\beta + \sigma\lambda\left(\frac{a - X_i\beta}{\sigma}\right),$$

$$Var(Y_i \mid Y_i > a) = \sigma^2\left[1 - \delta\left(\frac{a - X_i\beta}{\sigma}\right)\right],$$

dengan

$$\lambda\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)=\frac{\phi\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)}{1-\Phi\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)}$$

$$\delta\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)=\lambda\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right).\left[\lambda\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)-\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)\right]$$

Dengan demikian diperoleh model regresi terpotong kiri sebagai berikut.

$$Y_i \mid Y_i > a = E(Y_i \mid Y_i > a) + \varepsilon_i = \mathrm{X_i}\beta + \sigma\lambda\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right) + \varepsilon_i, \qquad (4.1)$$

dengan

$$\lambda\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)=\frac{\phi\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)}{1-\Phi\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)},$$

i = 1, 2, …, n observasi.

Menurut [5], adanya penambahan suku $\sigma\lambda\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)$ menyebabkan model regresi terpotong berbentuk nonlinier dalam β dan $\mathrm{X_i}$. Akibatnya, prosedur estimasi parameter $\hat{\beta}$ dan $\hat{\sigma}$ sedikit lebih kompleks dibandingkan model regresi klasik

5. ESTIMASI PARAMETER REGRESI TERPOTONG KIRI

Menurut [7], regresi terpotong kiri merupakan keluarga dari *Limited Dependent Variables* (LDV). Dalam LDV ada beberapa metode untuk melakukan estimasi parameter, yaitu metode Heckman, Amemiya dan Maksimum Likelihood. Untuk regresi terpotong kiri ini dipilih metode Maksimum Likelihood.

Dari persamaan (3.3), fungsi densitas normal terpotong kiri pada nilai a adalah

$$f(Y_i, Y_i > a \mid \beta,\sigma)=\frac{\frac{1}{\sigma}\phi\left(\frac{Y_i-\mathrm{X_i}\beta}{\sigma}\right)}{1-\Phi\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)},$$

dengan demikian fungsi Likelihoodnya

$$L(\beta,\sigma \mid Y_i, Y_i > a)=\frac{\prod_{i=1}^{n}\left[\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{Y_i-\mathrm{X_i}\beta}{\sigma}\right)^2}\right]}{\prod_{i=1}^{n}\left[1-\Phi\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)\right]} \qquad (5.1)$$

untuk memperoleh estimasi dari β dan σ, maka persamaan (5.1) ekivalen dengan

$$\log L(\beta,\sigma \mid Y_i, Y_i > a)= -\frac{1}{2}\sum_{i=1}^{n}\left[\log(2\pi)+\log\sigma^2+\left(\frac{Y_i-\mathrm{X_i}\beta}{\sigma}\right)^2+ 2\log\left(1-\Phi\left(\frac{a-\mathrm{X_i}\beta}{\sigma}\right)\right)\right].$$

Untuk mempermudah pemodelan maka dilakukan reparameterisasi Olsen (1978):

$$\gamma=\frac{\beta}{\sigma} \quad \text{dan} \quad \theta=\frac{1}{\sigma} \qquad (5.2)$$

maka fungsi log Likelihood diatas menjadi:

$$\log L(\theta,\gamma \mid Y_i, Y_i > a)= -\frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left[\log(2\pi)-\log\theta^2+(\theta Y_i-\mathrm{X_i}\gamma)^2+ 2\log(1-\Phi(\theta a-\mathrm{X_i}\gamma))\right].$$

Nilai θ dan **γ** akan diestimasi kembali dengan metode Maksimum Likelihood

$$\frac{\partial}{\partial\theta}\log L(\theta,\gamma \mid Y_i, Y_i > a)=0 \Rightarrow \sum_{i=1}^{n}\left[\frac{1}{\theta}-Y_i(\theta Y_i-\mathrm{X_i}\gamma)+a\lambda(\alpha_i)\right]=0 \qquad (5.3)$$

$$\frac{\partial}{\partial\gamma}\log L(\theta,\gamma \mid Y_i, Y_i > a)=0 \Rightarrow \sum_{i=1}^{n}\left[(\theta Y_i-\mathrm{X_i}\gamma)\mathrm{X_i'}-\lambda(\alpha_i)\mathrm{X_i'}\right]=0 \qquad (5.4)$$

dimana

$$\lambda(\alpha_i) = \frac{\phi(\alpha_i)}{1-\Phi(\alpha_i)} \text{ dengan } \alpha_i = \theta a - X_i\gamma$$

Persamaan (5.3) dan (5.4) belum dapat memberikan suatu penyelesaian karena bentuknya yang tidak "closed form" yaitu masing-masing persamaan masih mempunyai parameter lain yang belum diketahui. Oleh karena itu, diperlukan suatu metode numerik iterasi yang dalam hal ini dipilih metode Newton. Metode Newton adalah metode iterasi yang digunakan untuk menyelesaikan suatu persamaan nonlinier seperti persamaan likelihood. Dasar dari metode Newton adalah pendekatan deret Taylor *f(x)* di sekitar titik $x = x^0$, yaitu:

$$f(x) \cong f(x^0) + f'(x^0)(x-x^0) + \frac{1}{2} f''(x^0)(x-x^0)^2 + \cdots$$

Dalam penggunaannya diperlukan derivatif parsial pertama dan kedua dari fungsi Likelihood.

Diberikan g sebagai vektor derivatif parsial pertama:

$$g = \begin{bmatrix} \dfrac{\partial \log L(\theta,\gamma \mid Y_i, Y_i > a)}{\partial \theta} \\[2ex] \dfrac{\partial \log L(\theta,\gamma \mid Y_i, Y_i > a)}{\partial \gamma} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \left[\dfrac{1}{\theta} - Y_i(\theta Y_i - X_i\gamma) + a\lambda(\alpha_i) \right] \\[2ex] \sum_{i=1}^{n} \left[(\theta Y_i - X_i\gamma)X_i^{'} - \lambda(\alpha_i)X_i^{'} \right] \end{bmatrix}$$

Sedangkan **G** sebagai matriks Hessian atau matriks derivatif parsial kedua

$$G = \begin{bmatrix} \dfrac{\partial^2 \log L(\theta,\gamma \mid Y_i, Y_i > a)}{\partial\theta\partial\theta} & \dfrac{\partial^2 \log L(\theta,\gamma \mid Y_i, Y_i > a)}{\partial\theta\partial\gamma'} \\[2ex] \dfrac{\partial^2 \log L(\theta,\gamma \mid Y_i, Y_i > a))}{\partial\gamma\partial\theta} & \dfrac{\partial^2 \log L(\theta,\gamma \mid Y_i, Y_i > a))}{\partial\gamma\partial\gamma'} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \left[-\dfrac{1}{\theta^2} - Y_i^2 + a^2\delta(\alpha_i) \right] & \sum_{i=1}^{n} \left[Y_i X_i - \delta(\alpha_i) a X_i \right] \\[2ex] \sum_{i=1}^{n} \left[Y_i X_i^{'} - a\delta(\alpha_i) X_i^{'} \right] & \sum_{i=1}^{n} \left[-X_i X_i^{'} + \delta(\alpha_i) X_i X_i^{'} \right] \end{bmatrix}.$$

Dengan demikian estimasi parameter dengan metode Maksimum Likelihood dengan bantuan metode Newton dapat dirumuskan dalam bentuk iterasi sebagai berikut.

$$\begin{bmatrix} \hat{\theta}_{t+1} \\ \hat{\gamma}_{t+1} \end{bmatrix} = \begin{bmatrix} \hat{\theta}_t \\ \hat{\gamma}_t \end{bmatrix} - G_t^{-1} g_t$$

dimana

$$g_t = \begin{bmatrix} \sum_{i=1}^{n} \left[\dfrac{1}{\hat{\theta}_t} - Y_i(\theta Y_i - X_i\hat{\gamma}_t) + a\lambda(\alpha_i) \right] \\[2ex] \sum_{i=1}^{n} \left[(\hat{\theta}_t Y_i - X_i\hat{\gamma}_t)X_i^{'} - \lambda(\alpha_i)X_i^{'} \right] \end{bmatrix},$$

dan

$$G_t = \begin{bmatrix} \sum_{i=1}^{n} \left[-\dfrac{1}{\hat{\theta}_t^2} - Y_i^2 + a^2\delta(\alpha_i) \right] & \sum_{i=1}^{n} \left[Y_i X_i - \delta(\alpha_i) a X_i \right] \\[2ex] \sum_{i=1}^{n} \left[Y_i X_i^{'} - a\delta(\alpha_i) X_i^{'} \right] & \sum_{i=1}^{n} \left[-X_i X_i^{'} + \delta(\alpha_i) X_i X_i^{'} \right] \end{bmatrix}.$$

dengan *t* adalah ukuran iterasi.

Jika nilai estimasi parameter $\hat{\theta}$ dan $\hat{\gamma}$ telah diperoleh, maka melalui persamaan (5.2) nilai estimasi parameter $\hat{\beta}$ dan $\hat{\sigma}$ juga dapat diperoleh.

Sehingga berdasarkan persamaan (4.1) diperoleh pula estimasi model regresi terpotong kiri sebagai berikut.

$$\hat{Y} \mid Y > a = \mathrm{X}\hat{\beta} + \hat{\sigma}\lambda\left(\frac{a - \mathrm{X}\hat{\beta}}{\hat{\sigma}}\right),$$

dengan

$$\lambda\left(\frac{a - \mathrm{X}\hat{\beta}}{\hat{\sigma}}\right) = \frac{\phi\left(\frac{a - \mathrm{X}\hat{\beta}}{\hat{\sigma}}\right)}{1 - \Phi\left(\frac{a - \mathrm{X}\hat{\beta}}{\hat{\sigma}}\right)} \quad .$$

6. PENUTUP

Pemilihan nilai a sebagai titik potong kiri (*left truncated point*) pada model regresi terpotong kiri harus berdasarkan tujuan peneliti yang memiliki keahlian pada bidang ilmu atau permasalahan yang hendak diobservasi/dianalisis.

Regresi terpotong kiri (*left truncated regression*) merupakan keluarga *Limited Dependent Variables* (LDV) dan untuk estimasi parameternya dipilih metode Maksimum Likelihood. Karena bentuk persamaan Likelihood tidak "*closed form*", maka diperlukan metode iteratif yaitu metode Newton.

DAFTAR PUSTAKA

[1] Becker, W. E. & W. B. Walstad, (1987), *Economics Modelling in Economics Education Research,* Boston, Nijhoff Publishing.

[2] Draper, N. & H. Smith, (1992), *Analisis Regresi Terapan. Edisi Kedua. Terjemahan Bambang Sumantri*, Jakarta, PT Gramedia Pustaka Utama.

[3] Ender, Phil, (2004), *Educaton 213C-Applied Categorical & Non Normal Data Analysis. Los Angeles*, UCLA Departement of Education. htpp:// www.gseis.ucla.edu/courses/ed231c/231c.html

[4] Greene, W. H., (1993), *Econometric Analysis. Second Edition. New York*, Macmillan Publishing Company.

[5] Quantitative Micro Software, LLC., (2004). *EViews 4 User's Guide. Campus Drive*, Irvine CA, USA.

[6] Suppakornkosai, Sumanee, htpp:// faculty.smu.edu/tfomby/eco6375/presentations/Sumanee/Truncation.ppt

[7] Zivot, Eric, (2004). Economics 582 - Econometris Theory II. Departement of Economics. htpp://faculty. washington. edu/~ezivot/econ582/582 notes.htm.