BAB III
METODOLOGI PENELITIAN

Bab ini akan berisi mengenai metode penelitian dan teknik analisis yang digunakan pada penelitian ini. Metode analisi yang digunakan adalah regresi Heckit.

3.1 Prosedur Penelitian

Berikut merupakan penjelasan ringkas mengenai prosedur penelitian pada penelitian ini adalah:

1. Melakukan studi literatur mengenai konsep dasar regresi Heckit.
2. Melakukan survei konsumsi susu di Kota Bandung dalam periode April 2019 sampai dengan Juli 2019.
3. Memodelkan variable Z_i dengan variabel bebas (X) sebagai model awal.
4. Menduga parameter dengan model probit untuk memperoleh model persamaan seleksi.
5. Hasil dari tahap persamaan seleksi menghasilkan variabel baru yang dikenal sebagai *Inverse Mills Ratio* (λ), dimana nilai dari λ adalah
$$\lambda(\, w_i{}'\hat{\gamma}) = \frac{\phi(\, w_i{}'\hat{\gamma})}{\Phi(\, w_i{}'\hat{\gamma})}$$
6. Variabel baru yang dihasilkan akan diregresikan pada tahap persamaan hasil yaitu $Y_i = X_i\beta + \varepsilon_i$.
7. Menduga parameter dengan OLS untuk model persamaan hasil.
8. Dilakukan pengujian asumsi normalitas, homokedastisitas, multikolinieritas dan autokorelasi.
9. Dilakukan pengujian terhadap penduga parameter yang telah didapat menggunakan uji F dengan rumus $F = \frac{JKR/(p)}{JKG/(n-p-1)}$ untuk pengujian secara bersama-sama, sedangkan pada pengujian secara parsial digunakan uji t dengan rumus $t = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)}; dengan \quad Se(\hat{\beta}_j) = \sqrt{\mathrm{var}(C_{jj})}$.
10. Penentuan model terbaik berdasarkan nilai AIC, MSE, dan SIC.

3.2 Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data primer yang diperoleh dari hasil survei konsumsi rumah tangga untuk konsumsi susu di Kota Bandung dan data yang digunakan ini terdapat variabel yang diduga dapat mempengaruhi pengeluaran konsumsi rumah tangga untuk konsumsi susu di Kota Bandung antara lain pengeluaran konsumsi susu selama satu bulan, jumlah anggota rumah tangga, status bekerja kepala rumah tangga, pendidikan terakhir kepala rumah tangga, pendapatan rumah tangga selama satu bulan, dan usia kepala rumah tangga.

Penelitian pada regresi Heckit mengamati dua variabel tak bebas dan beberapa variabel bebas. Di antaranya adalah persamaan seleksi dan persamaan hasil.

1. Variabel Tak Bebas

Variabel tak bebas pada model Heckit pada persamaan seleksi adalah probabilitas mengonsumsi susu yang didefinisikan sebagai variabel dikotomi seperti berikut:

$$Z_i = \begin{cases} 0, & \text{tidak mengonsumsi susu} \\ 1, & \text{mengonsumsi susu} \end{cases}$$

Kemudian untuk variabel tak bebas pada persamaan hasil adalah jumlah konsumsi susu rumah tangga untuk konsumsi susu selama satu bulan.

2. Variabel Bebas

Variabel bebas yang digunakan pada persamaan seleksi dan persamaan hasil terdiri dari lima variabel yaitu pendidikan terakhir kepala rumah tangga, status bekerja kepala rumah tangga, jumlah anggota rumah tangga, pendapatan rumah tangga tiap bulan, dan usia kepala rumah tangga yang didefinisikan sebagai berikut:

Z_i : Status mengonsumsi susu

Y_i : Pengeluaran konsumsi susu rumah tangga

X_1 : Pendidikan terakhir kepala rumah tangga.

X_2 : Status bekerja kepala rumah tangga

X_3 : Jumlah anggota rumah tangga

X_4 :Pendapatan Rumah tangga, dalam hal ini diukur berdasarkan pendekatan pengeluaran pengeluaran rumah tangga selama satu bulan (rupiah).

X_5 : Usia kepala rumah tangga

3.3 Analisis Data

3.3.1 Regresi Heckit

Salah satu metode yang digunakan dalam kasus data tersensor seperti peneltian ini adalah metode regresi Heckit. Model regresi Heckit ini pertama kali dikenalkan oleh Heckman (1979). Menurut Heckman (1979), model ini muncul dalam praktek karena beberapa alasan antara lain peluang adanya seleksi diri untuk keluar atau masuk dalam bagian sampel oleh unit atau individu yang sedang diamati dalam penelitian dan keputusan pemilihan sampel oleh peneliti berlaku dalam banyak cara yang sama seperti seleksi diri (Heckman J. J., 1979). Model Heckit digunakan pada kasus estimator yang bias dan tidak konsisten yang ditangani dengan proses seleksi sampel. Seleksi sampel ini dilakukan terhadap variabel Z_i^*, dimana nilai dari Z_i^* tersebut dilambangkan menjadi nilai 0 untuk nilai dari Z_i^* yang sama dengan nol dan nilai 1 untuk nilai dari Z_i^* yang lebih dari nol yang kemudian disimbolkan menjadi variabel Z_i. (Siegelmen, Lee ; Zeng, Langche;, 1999). Pada model Heckit terdapat dua persamaan, yaitu persamaan seleksi dan persamaan hasil.

Persamaan seleksi didefinisikan :

$$Z_i^* = W_i^T\gamma + u_i \, , \, i = 1, \ldots, N \qquad (20)$$

di mana,

Z_i^* adalah variabel laten

γ adalah vektor dari parameter, dengan $\gamma = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_p \end{bmatrix}$

${W_i}^T$ adalah matriks tranpose variabel bebas,

dengan ${W_i}^T = \begin{bmatrix} w_{11} & w_{21} & \cdots & w_{n1} \\ \cdots & \cdots & \cdots & \cdots \\ w_{1p} & w_{2p} & \cdots & w_{np} \end{bmatrix}$

u_i adalah galat ~ N(0,1)

Karena Z_i^* merupakan variabel laten, maka tidak bisa diamati secara langsung. Meskipun demikian dapat didefinisikan dengan suatu variabel dikotomi yang disimbolkan sebagai Z_i seperti berikut:

$Z_i = 1,$ jika $Z_i^* > 0$

$Z_i = 0,$ jika $Z_i^* = 0$

Untuk persamaan hasil adalah model linier yang sering ditemui pada analisis regresi berganda, yaitu:

$$Y_i = {X_i}^T\beta + \varepsilon_i\ , i = 1, \ldots, n; N > n \qquad (3.1)$$

di mana Y_i adalah variabel yang teramati,

β adalah vektor dari parameter, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$

${X_i}^T$ adalah matrik s transpose dari variabel bebas,

$${X_i}^T = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix}$$

ε_i adalah galat ~N(0,σ²).

3.3.2 Pendugaan Parameter Model Heckit

Untuk menduga parameter model Heckit digunakan prosedur *Heckman Two-Step Estimator*.Adapun langkah-langkah dalam mengestimasi parameter model Heckit menggunakan prosedur dua langkah Heckman adalah sebagai berikut:

Misal akan mengestimasi suatu model regresi *Y* terhadap serangkaian variabel penjelas *X* sebagai berikut:

$$Z = W^T\gamma + u > 0$$

$$Y = X\beta + \varepsilon \,; E(\varepsilon|X) = 0$$

Tahap 1

Sampel dibedakan antara kelompok yang teramati dengan yang tidak teramati sebagai berikut:

$$Z = 0 \,, jika\; W_i{}^T\gamma + u_i = 0 \text{ dan } Z = 1, jika\; W_i{}^T\gamma + u_i > 0$$

Menduga nilai γ dengan menggunakan model Probit dengan probabilitas dari data yang teramati sebagai fungsi dari peubah penjelas *W* yang kemudian akan diperoleh nilai pendugaan dari γ, kemudian menghitung nilai duga dari Z dengan persamaan $\hat{Z}_i = W_i{}^T\hat{\gamma}$ dan menghasilkan fungsi kepadatan peluang $\left(\phi(W_i{}^T\hat{\gamma})\right)$ dan fungsi kepadatan peluang kumulatif $(\Phi(W_i{}^T\hat{\gamma}))$. λ yang dihasilkan merupakan hasil bagi antara fungsi kepadatan peluang dan fungsi kepadatan peluang kumulatif seperti pada persamaan (219).

Tahap 2

1. Hasil pada tahap 1 menghasilkan variabel baru yang didapatkan dari model Probit dan dikenal sebagai *Inverse Mills Ratio*, dimana nilai dari λ adalah

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} \frac{\phi(W_i{}^T\hat{\gamma})}{\Phi(W_i{}^T\hat{\gamma})} \qquad (3.2)$$

2. Variabel baru yang dihasilkan akan diestimasi pada tahap berikutnya dengan meregresikan Y_i pada X_i dan juga λ_i. Jadi model pada persamaan (21) menjadi:

$$Y_i = X_i{}^T\beta + \beta_\lambda\hat{\lambda} + \varepsilon_i$$

$$[X_i{}^T \quad \hat{\lambda}_i] \begin{bmatrix} \beta \\ \beta_\lambda \end{bmatrix} + \varepsilon_i = \underbrace{X_i}{}^T \beta^* + \underbrace{\varepsilon_i}$$

Pendugaan parameter β^* menggunakan metode kuadrat terkecil

$$b^* = (X^T X)^{-1} X^T Y$$

Estimator dari $\hat{\sigma}_e^2 = \left(\frac{\hat{\varepsilon}\prime\hat{\varepsilon}}{n}\right) + \bar{\hat{\delta}}\hat{\beta}_\lambda^2$ di mana $\hat{e} = Y - bX$, b_λ adalah estimator β_λ dengan metode kuadrat terkecil, $\hat{\delta}_i$ adalah penduga bagi δ_i, dan $\bar{\hat{\delta}}_i = \sum_i^n \frac{\hat{\delta}_i}{n}$. Sedangkan estimator dari $\hat{\rho}^2$ $yaitu$:

$$\hat{\rho}^2 = \frac{b_\lambda^2}{\hat{\delta}_e^2}.$$

3.3.3 Pemilihan Model Terbaik

Penentuan model terbaik untuk mengetahui pengeluaran konsumsi susu di kota Bandung dapat menggunakan metode antara lain:

1. *Akaike Info Criterion* (AIC)

   AIC didefinisikan sebagai (Gujarati & Porter, 2010) :

   $$AIC = e^{\frac{2k}{n}} \frac{\sum_{i=1}^n \hat{u}_i}{n} \tag{3.3}$$

   di mana k adalah jumlah parameter yang diestimasi dalam model regresi, n adalah jumlah observasi, dan $\hat{u}_i$ adsalah residual. Dalam membandingkan dua atau lebih model yang memiliki nilai AIC paling rendah merupakan model yang terpilih.

2. *Mean Square Error* (MSE)

   Mean Square Error (MSE) digunakan untuk menghitung rata-rata kuadrat dari suatu kesalahan atau merupakan varians dari suatu estimator. Dalam pemilihan model terbaik dari beberapa model yang dihasilkan, MSE juga dapat digunakan dengan memilih nilai MSE yang paling mendekati nol.

   $$MSE = \frac{1}{n}\sum_{i=1}^n \left(\hat{Y}_i - Y_i\right)^2 \tag{3.4}$$

3. *Schwarz Information Criterion* (SIC)

   Menurut metode SIC, model regresi terbaik adalah regresi yang mempunyai nilai SIC terkecil (Widarjono, 2007). Dari persamaan (3.3) dapat ditulis juga sebagai:

   $$\ln AIC = \frac{2k}{n} + \ln\left(\frac{\sum_{i=1}^n \hat{u}_i^2}{n}\right) \tag{3.5}$$

   $$SIC = n^{\frac{k}{n}} \frac{\sum_{i=1}^n \hat{u}_i^2}{n} \tag{3.6}$$

3.3.4 Penentuan Kesesuaian Model Heckit

Menentukan kesesuaian model dalam analisis regresi salah satunya dapat menggunakan koefisien determinasi R^2. Penentuan kesesuaian model did asarkan pada besarnya nilai R^2. Pada kasus regresi Heckit penulisan R^2 dapat dituliskan sebagai berikut (Heckman J. J., 1979).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \widehat{u_i^2}}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \tag{3.7}$$