

PEMODELAN REGRESI HECKIT UNTUK KONSUMSI SUSU DI PROVINSI JAWA TENGAH

Dwi Asti Rakhmawati, Dwi Ispriyanti², Agus Rusgiyono³

¹Mahasiswa Departemen Statistika FSM Universitas Diponegoro

^{2,3}Dosen Departemen Statistika FSM Universitas Diponegoro

ABSTRACT

In multiple regression if the response variable is dummy variable then it can not be used because it will produce biased and inconsistent estimator. The appropriate method for binary response variables is Heckit Regression. Estimation of Heckit Regression parameter using Two Step Method of Procurement is the selection equation and the result equation. In the selection equation will get new variable that is Invers Mills Ratio. While in Equation Result of new variable of Inverse Mills Ratio is added as independent variable along with other independent variable. Heckit Regression method is applied to household milk consume data obtained from 2015 SUSENAS results as many as 201 households. The response variable used is household expenditure for milk consumption. The independent variables used are the working status of the head of the household, the last education of the head of the household, the number of household members, the number of toddler age in the family and the income of the household.

Keywords: Multiple Regression, OLS, Heckit Regression, Two Step Procedure, Milk consumption expenditure.

1. PENDAHULUAN

Dalam penelitian bidang sosial-ekonomi, terdapat suatu variabel yang tidak dapat diukur secara langsung. Variabel semacam itu biasa disebut sebagai variabel laten. Salah satu kasus variabel laten di bidang sosial-ekonomi yaitu pada data tersensor. Disebut data tersensor karena dalam suatu observasi dimana informasi yang tersedia hanya untuk beberapa observasi karena terjadi perbedaan variasi responden yang memilih nol dan selain nol atau (Ekananda, 2015).

Dalam penelitian ini mengkaji gambaran pengeluaran konsumsi susu dan menganalisis berapa pengeluaran konsumsi susu rumah tangga tiap bulannya. Pada kasus ini variabel terikat (respon) yang diambil adalah pengeluaran konsumsi susu dan variabel bebasnya berupa status bekerja kepala rumah tangga, pendidikan terakhir kepala rumah tangga, pendapatan per bulan, banyaknya anggota keluarga usia balita, dan jumlah anggota rumah tangga. Pada kasus ini variabel respon bernilai 1 apabila rumah tangga mengonsumsi susu dan bernilai 0 jika rumah tangga tidak mengonsumsi susu, maka regresi berganda tidak dapat digunakan karena akan memberikan estimator yang bias dan tidak konsisten. Hal tersebut terjadi dalam penelitian sebelumnya yang dilakukan oleh Firdaus dan Suryaningsih (2010) menggunakan analisis regresi berganda dengan OLS sebagai metode pendugaan untuk mengetahui fungsi konsumsi telur pada rumah tangga di Pulau Jawa. Bias disini adalah nilai harapan penduga memberikan hasil yang berbeda dengan parameter yang diduga. Dan tidak konsisten karena model sampelnya tidak mendekati model populasi secara keseluruhan sehingga jumlah galatnya menjadi lebih besar sehingga memberikan gambaran yang kurang sesuai.

Metode yang dapat menangani kasus pengeluaran konsumsi susu sehingga dapat diperoleh hasil estimasi tak bias dan konsisten yang dapat memberikan gambaran keadaan yang sesuai pada kasus tersebut. Salah satu metode yang dapat digunakan untuk menyelesaikan kasus tersebut adalah Regresi Heckit. Regresi Heckit merupakan suatu

metode yang sampelnya terseleksi dimana dari keseluruhan data hanya data yang terobservasi saja yang digunakan. Metode ini cocok untuk permasalahan pengeluaran konsumsi susu dalam rumah tangga karena metode ini mampu melakukan seleksi data yang diperlukan sehingga data yang digunakan dalam analisis lebih diprioritaskan yang terobservasi.

Model Regresi Heckit ini terdiri dari dua langkah prosedur pendugaan. Pada langkah pertama digunakan model probit untuk menduga probabilitas dari seorang konsumen memutuskan untuk mengonsumsi susu. Hal ini diperlukan karena kemungkinan ada faktor tak terukur lain yang mempengaruhi keinginan untuk mengonsumsi susu tersebut. Pada langkah kedua digunakan penduga OLS untuk menduga model tobit dengan seleksi sampel dan menambah sebuah variabel baru (*invers mills ratio*) yang diturunkan dari pendugaan probit. Dari hasil tersebut dapat diketahui besar konsumsi susu pada individu yang secara aktual mengonsumsi susu (Gujarati dan Porter, 2012).

2. TINJAUAN PUSTAKA

2.1. Konsumsi Susu

Konsumsi susu yang dimaksud dalam penelitian ini yaitu besarnya uang atau rupiah yang dibelanjakan oleh suatu rumah tangga untuk mengonsumsi susu dalam jangka waktu satu bulan (BPS, 2015).

2.2. Faktor-faktor yang Mempengaruhi Konsumsi Susu

Menurut BPS (2015), faktor faktor yang mempengaruhi pengaruh pengeluaran konsumsi susu dengan menggunakan data hasil susenas tersebut antara lain :

- a. Tingkat Pendidikan Kepala Rumah Tangga
- b. Status Pekerjaan Kepala Rumah Tangga
- c. Jumlah Anggota Rumah Tangga
- d. Jumlah Anggota Rumah Tangga Usia Balita
- e. Pendapatan Rumah Tangga per Bulan

2.3. Analisis Regresi Linear Berganda

Model regresi linier berganda (*Multiple Linear Regression Model*) merupakan regresi yang digunakan untuk menguji dua atau lebih variabel bebas terhadap satu variabel tak bebas (Montgomery dan Runger, 2011). Bentuk umum persamaan regresi linier berganda adalah:

$$Y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \varepsilon_i \text{ dengan } i = 1, 2, \dots, n \quad (1)$$

dengan variabel y_i dinamakan variabel respon, variabel x_{ij} disebut variabel prediktor, β_0 adalah konstanta (*intersept*), dan β_j adalah koefisien regresi atau nilai parameter dengan $i=1, 2, \dots, n$ dan $j=1, 2, \dots, p$.

Dalam analisis ini, harus memenuhi beberapa asumsi seperti normalitas, multikolinearitas, heteroskedastisitas, dan autokorelasi.

2.4. Model Probit

Menurut Bliss (1935), model Probit disebut juga model Normit yang menggunakan bilangan biner sebagai variabel tak bebasnya. Pada model Probit responnya bersifat biner yaitu hanya bernilai 0 dan 1.

Persamaan model Probit antara Z_i dengan variabel bebasnya W_i sebagai berikut.

$$Z_i = W_i^T \gamma + u_i, i = 1, 2, \dots, n \quad (2)$$

dengan variabel Z_i dinamakan variabel respon biner, variabel W_i adalah matriks tranpose dari variabel bebas, γ adalah parameter model regresi, dan u_i adalah galat yang berdistribusi Normal $N(0, \sigma^2)$.

2.5. Model Heckit

Model Heckit digunakan pada kasus estimator yang bias dan tidak konsisten yang ditangani dengan proses seleksi sampel. Seleksi sampel ini dilakukan terhadap variabel P_i , dimana nilai dari P_i tersebut dilambangkan menjadi nilai 0 untuk nilai dari P_i yang sama dengan nol dan nilai 1 untuk nilai dari P_i yang lebih dari nol yang kemudian disimbolkan menjadi variabel Z_i (Siegelman dan Zeng, 1999).

Pada model Heckit terdapat dua persamaan, yaitu persamaan seleksi dan persamaan hasil. Pada persamaan seleksi P_i variabel laten, maka tidak bisa diamati secara langsung. Meskipun demikian dapat didefinisikan dengan suatu variabel dikotomi yang disimbolkan sebagai Z_i seperti berikut.

$$\begin{aligned} Z_i &= 1, && \text{jika } P_i > 0 \\ Z_i &= 0, && \text{jika } P_i = 0 \end{aligned}$$

Persamaan seleksi didefinisikan :

$$Z_i = W_i^T \gamma + u_i, i = 1, \dots, N \quad (3)$$

dengan variabel Z_i dinamakan variabel respon biner, variabel W_i adalah matriks tranpose dari variabel bebas, γ adalah parameter model regresi, dan u_i adalah galat yang berdistribusi Normal $N(0, \sigma^2)$. Dari persamaan seleksi didapatkan variabel baru yaitu *Invers Mills Ratio* (hasil bagi antara fungsi kepadatan peluang dan fungsi kepadatan peluang kumulatif) yang kemudian ditambahkan sebagai variabel bebas pada persamaan hasil.

Persamaan hasilnya sebagai berikut.

$$Y_i = X_i^T \beta + \beta_\lambda \lambda + \varepsilon_i, i = 1, \dots, n; N > n \quad (4)$$

dengan variabel Y_i dinamakan variabel respon biner, variabel X_i^T adalah matriks tranpose dari variabel bebas, β adalah parameter model regresi, β_λ adalah koefisien dari *inverse mills ratio*, λ adalah *inverse mills ratio* dan ε_i adalah galat yang berdistribusi Normal $N(0, \sigma^2)$.

2.6. Pendugaan Parameter Model Heckit

Tahap 1

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

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

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 $(\phi(W_i^T \hat{\gamma}))$ 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.

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

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

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 (4) menjadi:

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

$$\begin{bmatrix} X_i^T & \hat{\lambda}_i \end{bmatrix} \begin{bmatrix} \beta \\ \beta_\lambda \end{bmatrix} + \varepsilon_i = X_i^T \beta^* + \varepsilon_i$$

Pendugaan parameter β^* menggunakan metode kuadrat terkecil
 $b^* = (X^T X)^{-1} X^T Y$

2.7 Uji Kecocokan Model

Dari model persamaan (4) dilakukan uji kecocokan model yang bertujuan untuk mengetahui pengaruh variabel bebas secara bersama-sama terhadap variabel tak bebas. Berikut langkah-langkah menguji hipotesis :

1. Merumuskan hipotesis

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_p = 0$$

H_1 : paling sedikit terdapat satu $\beta_j \neq 0$ dengan $j = 0, 1, 2, \dots, p$

2. Menentukan taraf signifikansi yang digunakan sebesar α .
3. Menentukan uji statistik nilai F

$$F = \frac{JKR/(p)}{JKG/(n-p-1)} \quad (5)$$

Keterangan :

p = banyaknya parameter (koefisien regresi) model regresi linier

n = banyaknya pengamatan

4. Menentukan kriteria pengujian, yaitu H_0 ditolak jika $F \text{ hitung} \geq F_{\alpha; (p); (n-p-1)}$ atau $p\text{-value} \leq \alpha$ yang artinya terdapat pengaruh yang signifikan secara bersama-sama antara variabel bebas terhadap variabel terikat.

2.8 Uji Koefisien Regresi Secara Individual

Uji ini merupakan pengujian secara individu yang digunakan untuk mengetahui apakah masing-masing variabel independen benar-benar memberikan pengaruh signifikan terhadap variabel dependen. (Douglas C. Montgomery dan George C. Runger, 2011:473)

Berikut langkah-langkah menguji hipotesis :

- 1) Merumuskan hipotesis

$$H_0 : \beta_j = 0, \text{ dengan } j = 1, 2, \dots, p$$

$$H_1 : \beta_j \neq 0$$

- 2) Menentukan taraf signifikansi yang digunakan sebesar α
- 3) Menentukan uji statistik nilai t

$$t = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)}; \text{ dengan } Se(\hat{\beta}_j) = \sqrt{\text{var}(C_{jj})} \quad (6)$$

dimana C_{jj} merupakan elemen diagonal dari matriks $(X^T X)^{-1}$

- 4) Menentukan kriteria pengujian, yaitu H_0 ditolak jika $|t_0| \geq t_{(\alpha/2; n-p)}$ atau $p\text{-value} \leq \alpha$ yang artinya parameter tersebut berbeda dengan nol secara signifikan pada tingkat signifikansi α terhadap model.

Keterangan :

p = banyaknya parameter (koefisien) model regresi linier

n = banyaknya pengamatan

$p\text{-value}$ = probabilitas kesalahan dalam menolak H_0 yang seharusnya diperoleh berdasarkan data pengamatan.

2.9 Pemilihan Model Terbaik

1. Akaike Info Criterion (AIC)

AIC didefinisikan sebagai (Gujarati dan Porter, 2012) :

$$AIC = e^{\frac{2k}{n} \sum_{i=1}^n \hat{u}_i} \quad (7)$$

di mana k adalah jumlah parameter yang diestimasi dalam model regresi, n adalah jumlah observasi, dan \hat{u}_i adalah 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.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (8)$$

3. METODE PENELITIAN

3.1. Sumber Data dan Variabel Penelitian

Data yang akan digunakan dalam penelitian ini adalah data hasil Survei Sosial Ekonomi Nasional (Susenas) Provinsi Jawa tengah tahun 2015 yang didapatkan dari Badan Pusat Statistik (BPS) Provinsi Jawa tengah. Unit pengamatan pada penelitian ini adalah rumah tangga yang terdiri dari 201 sampel rumah tangga. Variabel responnya adalah pengeluaran konsumsi susu tiap bulan dan variabel bebasnya adalah status bekerja kepala rumah tangga, pendidikan terakhir kepala rumah tangga, jumlah anggota rumah tangga, banyaknya balita dalam rumah tangga, dan pendapatan rumah tangga tiap bulan.

3.2. Tahapan Analisis Data

Tahapan analisis pola hubungan pengeluaran konsumsi susu rumah tangga Provinsi Jawa Tengah dengan faktor-faktor yang diduga mempengaruhi sebagai berikut.

1. Memodelkan variabel Z_i dengan variabel bebas sebagai model awal.
2. Menduga parameter dengan model probit untuk model persamaan seleksi.
3. Hasil dari persamaan seleksi menghasilkan variabel baru yang dikenal sebagai *Inverse Mills Ratio* (λ), dimana nilai dari λ adalah

$$\lambda(w_i' \hat{\gamma}) = \frac{\phi(w_i' \hat{\gamma})}{\Phi(w_i' \hat{\gamma})}$$

4. Variabel baru (λ) yang dihasilkan akan diregresikan pada tahap persamaan hasil dengan variabel bebas lainnya.
5. Menduga parameter dengan OLS untuk model persamaan hasil.
6. Dilakukan pengujian asumsi normalitas, homoskedastisitas, multikolinieritas dan autokorelasi.
7. Dilakukan pengujian terhadap penduga parameter yang telah didapat menggunakan uji F dengan rumus (7) untuk pengujian secara bersama-sama, sedangkan pada pengujian secara parsial digunakan uji t dengan rumus (8).

4. HASIL DAN PEMBAHASAN

4.1 Model Heckit Tahap Pertama

Model yang diperoleh adalah sebagai berikut.

$$\hat{Z} = -1,118502 + 1,663954X_1 + 0,4230281X_2 + 0,1171111X_3 - 0,1042591X_4 - 0,00000824X_5$$

$$\hat{Y} = -1240320 + 762721,9X_1 - 29356,64X_2 - 2264,741X_3 + 29122,18X_4 + 1241011X_5 - 0,0385442IMR$$

Model regresi Heckit tahap pertama tersebut setelah diuji asumsi hasilnya adalah bahwa model tersebut memenuhi keempat asumsi yaitu residual berdistribusi normal, homoskedastisitas, non-multikolinieritas, dan non autokorelasi.

4.1.2 Pengujian Secara Bersama

Hipotesis yang digunakan dalam uji ini yaitu:

$$H_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_1 = \text{paling sedikit ada satu } \beta_j \neq 0, \text{ dengan } j=1,2,3,4,5$$

Dengan taraf signifikansi $\alpha = 5\%$ diperoleh nilai $F = 4,35$ dan nilai $p\text{-value} = 0,0012$. Sehingga diperoleh hasil bahwa $p\text{-value} = 0,0012 < \alpha=0,05$ yang artinya terdapat pengaruh yang signifikan secara bersama-sama antara variabel bebas terhadap variabel tak bebas.

4.1.3 Pengujian Koefisien Secara Individual

Hipotesis yang digunakan yaitu:

$$H_0 = \beta_j = 0$$

$$H_1 = \beta_j \neq 0, \text{ dengan } j=1,2,3,4,5$$

Dengan taraf signifikansi $\alpha = 5\%$ dari diperoleh nilai t beserta peluangnya yang diringkas dalam Tabel 1 berikut.

Tabel 1 Nilai $t\text{-statistic}$ Variabel Bebas

Variabel Bebas	Koefisien	t	$p\text{-value}$
X_1	762721	2,73	0,008
X_2	-29356,64	-1,82	0,074
X_3	-2264,741	-1,88	0,065
X_4	29122,18	2,25	0,028
X_5	1241011	3,08	0,003
IMR	-0,0385442	-2,41	0,019

Menurut Tabel 1 dapat diketahui bahwa variabel bebas yang berpengaruh secara signifikan terhadap variabel tak bebas pengeluaran konsumsi susu adalah X_1 yaitu status bekerja kepala rumah tangga, X_4 yaitu banyaknya anggota keluarga usia balita, dan X_5 yaitu pendapatan rumah tangga tiap bulan.

4.2 Model Heckit Tahap Kedua

Model yang diperoleh adalah sebagai berikut.

$$\hat{Z} = -0,7933283 + 1,670364X_1 + 0,0040251X_4 - 0,00000581X_5$$

$$\hat{Y} = -619486,7 + 329247X_1 + 37797,02X_4 + 605537,9X_5 - 0,0285904IMR$$

Model regresi Heckit tahap pertama tersebut setelah diuji asumsi hasilnya adalah bahwa model tersebut memenuhi keempat asumsi yaitu residual berdistribusi normal, homoskedastisitas, non-multikolinieritas, dan non autokorelasi.

4.2.2 Pengujian Secara Bersama

Hipotesis yang digunakan dalam uji ini yaitu:

$$H_0 = \beta_1 = \beta_4 = \beta_5 = 0$$

$$H_1 = \text{paling sedikit ada satu } \beta_j \neq 0, \text{ dengan } j=1,4,5$$

dengan taraf signifikansi $\alpha = 5\%$ dapat dilihat diperoleh nilai $F = 4,10$ dan nilai $p\text{-value} = 0,0056$. Sehingga diperoleh hasil bahwa $p\text{-value} = 0,0056 < \alpha = 0,05$ yang artinya terdapat pengaruh yang signifikan secara bersama-sama antara variabel bebas terhadap variabel tak bebas.

4.2.3 Pengujian Koefisien Secara Individual

Hipotesis yang digunakan yaitu:

$$H_0 = \beta_j = 0$$

$$H_1 = \beta_j \neq 0, \text{ dengan } j=1,4,5$$

dengan taraf signifikansi $\alpha = 5\%$ dari diperoleh nilai t beserta peluangnya yang diringkas dalam Tabel 2 berikut.

Tabel 2 Nilai $t\text{-statistic}$ Variabel Bebas

Variabel Bebas	Koefisien	t	$p\text{-value}$
X_1	329247,9	2,24	0,029
X_4	37797,02	2,94	0,005
X_5	605537,9	3,53	0,001
IMR	-0,0285904	-2,85	0,006

Menurut Tabel 2 dapat diketahui bahwa variabel bebas yang berpengaruh secara signifikan terhadap variabel tak bebas pengeluaran konsumsi susu adalah X_1 yaitu status bekerja kepala rumah tangga, X_4 yaitu jumlah anggota keluarga usia balita, dan X_5 yaitu pendapatan rumah tangga tiap bulan.

4.3 Pemilihan Model Terbaik

Tabel 3 Nilai AIC dan MSE tiap model

Model	AIC	MSE
Model Pertama (5 variabel)	25,82760	0,881318
Model Kedua (3 variabel)	25,70569	0,591008

Dari Tabel 3 dapat dilihat bahwa nilai AIC dan MSE yang dihasilkan oleh model kedua dengan 3 variabel yang signifikan lebih kecil dibandingkan dengan model pertama dengan 5 variabel. Sehingga model terbaik yang digunakan adalah model kedua dengan 3 variabel.

5. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan, maka diperoleh kesimpulan sebagai berikut:

1. Model hasil analisis menggunakan Regresi Heckit adalah:

Model seleksi :

$$\hat{Z} = -1,118502 + 1,663954X_1 + 0,4230281X_2 + 0,1171111X_3 - 0,1042591X_4 - 0,00000824X_5$$

Model hasil :

$$\hat{Y} = -1240320 + 762721,9X_1 - 29356,64X_2 - 2264,741X_3 + 29122,18X_4 + 1241011X_5 - 0,0385442IMR$$

2. Berdasarkan model Regresi Heckit tahap pertama, variabel bebas yang berpengaruh terhadap pengeluaran konsumsi susu rumah tangga adalah status bekerja kepala rumah tangga, banyaknya anggota rumah tangga usia balita, dan pendapatan rumah tangga. Sehingga dilakukan pemodelan Regresi Heckit dengan tiga variabel yang signifikan dan hasilnya sebagai berikut. adalah:

Model seleksi :

$$\hat{Z} = -0,7933283 + 1,670364X_1 + 0,0040251X_4 - 0,00000581X_5$$

Model hasil :

$$\hat{Y} = -619486,7 + 329247X_1 + 37797,02X_4 + 605537,9X_5 - 0,0285904IMR$$

3. Dari kedua model yang didapatkan model kedua dengan tiga variabel merupakan model yang terbaik karena memiliki nilai *AIC* dan *MSE* yang lebih rendah dibandingkan dengan model pertama dengan lima variabel.

DAFTAR PUSTAKA

- Ariefianto, M.D. 2012. *Ekonometrika Esensi dan Aplikasi dengan Menggunakan Eviews*. Briggs, D. C. (2004). Causal Inference and Heckman Model. *Journal of Educational and Behavioral Statistics*, 5-15.
- Dimitriou, L. (2008). Modeling Participation and Consumption in the Greek Interrurban Public. *Journal of Public Transportation*.
- Firdaus, M., & Suryaningsih, T. (2010). Kemiskinan dan Tingginya Konsumsi Rokok: Faktor Penyebab Sulitnya Implementasi Green Economic di Pulau Jawa.
- Greene, W. H. (2002). *Econometric Analysis 5th-Edition*. New Jersey : Pearson Education Inc.
- Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample-Selection, and Limited Dependent Variables and A Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, 475-478.
- Heckman, J. J. (1979). Sample Selection Bias As A Specification Error. *Econometrica*, 153-161.

- Hills, R. C., Adkins, L. C., & Bender, K. A. (2002). Test Statistics and Critical Values in Selectivity Models. *Advance in Econometric Conference*, 1-11.
- Montgomery, D. C., & Runger, G. C. (2011). *Applied Statistics and Probability for Engineers*. United States.
- N.Gujarati, D., & D.C., P. (2010). *Dasar-Dasar Ekonometrika*. Jakarta: Sealemba Empat.
- Ritov, Y. (1990). Estimation in A Linear Regression Model With Censored Data. *Annals of Statistics*, 303-328.
- Schmidheiny, D. K. (2007). Limited Dependent Variable Models. *Lecture Notes in Microeconometrics*, 11-18.
- Siegelmen, Lee ; Zeng, Langche;. (1999). Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models. *Political Analysis*, 177-181.
- Tobin, J. (1958). Estimation of relationships for limited dependent. *Econometrica Vol 26*., 24-36.
- Vance, M. F. (2012). *On Interaction Effects: The Case of Heckit and Two Part Models*. Essen: RUB.