

Penggunaan Model Regresi Tobit Pada Data Tersensor

Defi Yusti Faidah¹, Resa Septiani Pontoh²

^{1,2}Departemen Statistika FMIPA Universitas Padjadjaran
defi.yusti@unpad.ac.id

Abstrak—Model regresi tobit digunakan ketika terdapat data yang bernilai nol untuk sebagian observasi, dan sisanya memiliki nilai yang beragam. Data yang memiliki struktur tersebut dinamakan data tersensor. Ciri lain dari data tersensor adalah sebagian nilai dari suatu rentang tertentu ditransformasikan sebagai suatu nilai tunggal. Penggunaan metode analisis regresi linier klasik tidak dapat digunakan untuk melihat hubungan variabel yang sifatnya tersensor. Sementara itu, penggunaan analisis regresi logistik tidak dapat menggambarkan nilai pengamatan yang bervariasi. Metode penaksiran parameter yang digunakan adalah Maximum Likelihood Estimation. Penaksiran parameter model tobit dilakukan dengan melakukan turunan parsial pertama fungsi ln likelihood terhadap parameter yang akan diestimasi dan kemudian disamakan dengan nol. Kajian penelitian ini adalah memodelkan Tingkat Pengangguran Terbuka (TPT) Perempuan di Pulau Jawa. Berdasarkan hasil pemodelan regresi tobit dapat diketahui faktor-faktor yang berpengaruh secara signifikan terhadap TPT Perempuan di Pulau Jawa yaitu persentase penduduk yang tinggal di daerah perkotaan, seks rasio, persentase penduduk yang berpendidikan diatas SLTP, dan tingkat pertumbuhan ekonomi. Diharapkan dengan diketahui faktor-faktor yang berpengaruh dapat dijabarkan sebagai masukan kepada pemerintah untuk menekan jumlah TPT Perempuan di Pulau Jawa.

Kata kunci: Data Tersensor, Regresi Tonit, TPT Perempuan

I. PENDAHULUAN

Masih tingginya jumlah pengangguran perempuan merupakan masalah yang cukup serius mengingat salah satu tujuan *Millenium Development Goals* adalah mendorong kesetaraan gender dan pemberdayaan perempuan. Akan tetapi pada kenyataannya, kesempatan perempuan untuk memperoleh lapangan pekerjaan masih kalah jika dibandingkan dengan laki-laki. Hal ini terlihat dari masih tingginya Tingkat Pengangguran Terbuka (TPT) perempuan di Pulau Jawa [1]. TPT Perempuan akan diberikan sensor pada suatu nilai tertentu sehingga sebagian data akan ditransformasi menjadi suatu nilai tunggal atau konstanta yang disebut dengan data tersensor.

Pendekatan yang dapat digunakan untuk mengatasi data tersensor tersebut adalah model regresi tobit. Penggunaan regresi tobit pada data campuran atau *mixture* akan mengurangi efek bias jika dibandingkan dengan data yang diolah menggunakan regresi linier klasik. Hal ini dikarenakan data yang bernilai konstan dapat diolah secara bersama dengan data kontinu sehingga tidak akan kehilangan informasi yang berasal dari data diskrit [2]. Model regresi tobit dapat menghasilkan *standard error* yang lebih *robust* dibandingkan model regresi linier klasik dan juga nilai prediksi yang kuat untuk kasus data tersensor [3]. Perbandingan antara regresi linier klasik bivariat dengan model tobit *bivariate* diperoleh kesimpulan bahwa model tobit menghasilkan R^2 lebih besar daripada regresi linier klasik pada data tersensor. [4]

Metode regresi tobit telah banyak digunakan dan dikembangkan dalam berbagai penelitian. Beberapa penelitian tersebut dilakukan untuk mengkaji tentang kepuasan konsumen untuk jasa pengangkutan barang [5]; permintaan dan penawaran kredit konsumsi rumah tangga [6]; faktor-faktor yang mempengaruhi konsumsi terhadap daging merah dan daging putih [7]; pengeluaran konsumsi daging dan susu [8]; dan penelitian tentang identifikasi nonparametrik dan estimasi untuk regresi tobit [9].

II. METODE PENELITIAN

A. Model Tobit

Model Tobit merupakan analisis regresi yang digunakan untuk menggambarkan hubungan antara variabel dependen (Y) dan variabel independen (X) dimana variabel dependen tersebut berskala campuran.

Model tobit pertama kali dikemukakan oleh James Tobin pada 1958 yang digunakan untuk menganalisis pengeluaran para rumah tangga di Amerika Serikat untuk membeli mobil [4]. Pemodelan ini diawali dengan memperhatikan model sebagai berikut [2] :

$$y_i^* = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i \quad (1)$$

dengan y_i^* adalah variabel laten dependen yang diobservasi untuk nilai yang lebih besar dari c dan tersensor untuk nilai lainnya, \mathbf{x}_i^T adalah vektor variabel bebas $\mathbf{x}_i^T = [1 \quad X_{1i} \quad \cdots \quad X_{pi}]$, $\boldsymbol{\beta}$ adalah vektor parameter koefisien, dengan $\boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \cdots \quad \beta_k]^T$ dan ε_i adalah *error* yang diasumsikan berdistribusi $N(0, \sigma^2)$. Nilai observasi y_i , $i = 1, 2, \dots, n$ diperoleh dari persamaan (1) sehingga untuk persamaan data tersensor adalah

$$y_i = \begin{cases} c & \text{untuk } y_i^* \leq c \\ y_i^* = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i & \text{untuk lainnya} \end{cases} \quad (2)$$

dengan y_i^* berdistribusi normal dengan mean $\mathbf{x}_i^T \boldsymbol{\beta}$ dan varians σ^2

Persamaan yang mengandung beberapa variabel prediktor dan berpengaruh terhadap variabel respon dapat dilakukan pengujian dengan *likelihood ratio test* [10],[11]. *Likelihood ratio test* digunakan untuk menguji estimasi parameter secara serentak, sedangkan uji *wald* digunakan untuk pengujian secara individu.

B. Pengujian Model Tobit

Persamaan yang mengandung beberapa variabel prediktor dan berpengaruh terhadap variabel respon dapat dilakukan pengujian dengan *likelihood ratio test* [1]. *Likelihood ratio test* digunakan untuk menguji estimasi parameter secara serentak, sedangkan uji *wald* digunakan untuk pengujian secara individu.

Uji Serentak

Uji serentak digunakan untuk menguji parameter secara bersama-sama. Hipotesis yang digunakan adalah sebagai berikut :

$$H_0 = \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$H_1 = \text{minimal ada salah satu } \beta \text{ yang tidak sama dengan } 0$$

Statistik Uji yang digunakan adalah

$$G^2 = -2 \ln \left(\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right) \quad (3)$$

Dimana $L(\hat{\omega})$ = nilai maksimum likelihood tanpa variabel prediktor tertentu

$L(\hat{\Omega})$ = nilai maksimum likelihood dengan variabel prediktor tertentu

H_0 ditolak jika $G^2 > \chi_{(\alpha, k)}^2$, karena G^2 secara *asymptotically* mengikuti distribusi *chi-square*.

Dimana k adalah banyaknya variabel prediktor model atau jika $p - \text{value} < \alpha$ yang berarti ada salah satu atau lebih β_k yang berpengaruh pada model.

Uji Parsial

Uji Parsial digunakan untuk untuk pengujian individu yang menunjukkan apakah suatu variabel bebas signifikan atau layak untuk masuk model. Pengujian yang digunakan adalah *Wald test* [11].

$$H_0 : \beta_j = 0$$

$H_1 : \beta_j \neq 0$, dimana $j = 1, 2, \dots, k$

Statistik uji Wald yang digunakan adalah

$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \tag{4}$$

Berdasarkan (6) dan (7) H_0 ditolak jika $|W| > Z_{\alpha/2}$, atau jika $p - value < \alpha$ yang berarti bahwa parameter berpengaruh. Sampel besar mengikuti sebaran normal, maka kriteria pengujian dibandingkan dengan tabel normal Z.

C. Data Penelitian

Data yang digunakan adalah data sekunder yang yang diperoleh dari Badan Pusat Statistik (BPS) yaitu data Survei Sosial Ekonomi Nasional (Susenas) dan Survei Angkatan Kerja Nasional (Sakernas) Pada penelitian ini yang dijadikan unit observasi adalah 118 kabupaten/kota di Pulau Jawa. Variabel respon dalam penelitian ini adalah TPT perempuan dimana nilai C yang digunakan adalah 7,14. Hal ini karena pada umumnya nilai TPT perempuan Indonesia pada tahun 2010 mencapai 7,14. Terdapat tujuh variabel prediktor yang digunakan yaitu persentase penduduk yang tinggal di perkotaan (X_1), angka pertumbuhan penduduk (X_2), seks rasio (X_3), persentase penduduk berpendidikan SMP-Perguruan Tinggi (X_4), persentase penduduk yang mampu membaca dan menulis (X_5), tingkat pertumbuhan ekonomi (X_6) dan persentase partisipasi angkatan kerja (X_7).

D. Langkah Penelitian

Langkah-langkah yang dilakukan pada penelitian ini adalah

- Meregresikan variabel y terhadap semua variabel prediktor.
- Mencari nilai estimasi parameter dengan metode *MLE*.
- Melakukan pengujian terhadap estimasi parameter yang telah didapat dengan menggunakan *LR test* untuk keseluruhan model dan uji *wald* untuk menguji secara individu setiap estimasi parameter.
- Melakukan uji kebaikan model dengan menghitung nilai R^2 model.

III. HASIL DAN PEMBAHASAN

Pada bagian ini diuraikan tentang deskripsi TPT Perempuan di Pulau Jawa. Selain itu juga diuraikan pemodelan TPT Perempuan beserta faktor-faktor yang mempengaruhinya dengan menggunakan model tobit.

A. Deskripsi TPT Perempuan

Hampir semua kabupaten/kota di Provinsi Jawa Timur memiliki TPT perempuan yang berkisar antara 7,14 – 8,58 persen. Hal yang sama terjadi pada D.I Yogyakarta dan sebagian besar kabupaten/kota di Provinsi Jawa Tengah, dimana sebagian besar wilayah di kedua provinsi tersebut memiliki TPT perempuan yang berkisar antara 7,14-8,58 persen.



Gambar 1 Persentase TPT perempuan di Pulau Jawa

B. Pemodelan TPT Perempuan

Langkah awal dalam model tobit adalah melakukan analisis korelasi antara variabel prediktor dan respon. Hal ini digunakan untuk mengetahui hubungan antara variabel prediktor dan respon. Berdasarkan Tabel 1 dapat diketahui bahwa hasil pengujian dengan menggunakan $\alpha = 5\%$ diperoleh ketujuh variabel prediktor, yaitu persentase penduduk yang tinggal di perkotaan (X_1), angka pertumbuhan penduduk (X_2), seks rasio (X_3), persentase penduduk berpendidikan SMP-Perguruan Tinggi (X_4), persentase penduduk yang mampu membaca dan menulis (X_5), tingkat pertumbuhan ekonomi (X_6) dan persentase partisipasi angkatan kerja (X_7) memiliki hubungan yang nyata terhadap TPT perempuan (Y).

Tabel 1 Korelasi antara Variabel Prediktor dan Variabel Respon

Variabel Prediktor	Koefisien Korelasi	P-Value	Keterangan
X_1	0,411	0,00	Ada Korelasi
X_2	0,437	0,00	Ada Korelasi
X_3	0,579	0,00	Ada Korelasi
X_4	0,216	0,019	Ada Korelasi
X_5	0,374	0,00	Ada Korelasi
X_6	-0,406	0,00	Ada Korelasi
X_7	-0,440	0,00	Ada Korelasi

Terdapat lima variabel prediktor yang berkorelasi positif terhadap TPT perempuan yaitu X_1 , X_2 , X_3 , X_4 , dan X_5 . Korelasi positif ini berarti bahwa jika terjadi peningkatan pada variabel X_1 , X_2 , X_3 , X_4 , dan X_5 maka akan mengakibatkan semakin tingginya persentase TPT perempuan begitu juga sebaliknya. Sementara itu variabel X_6 dan X_7 berkorelasi negatif yang berarti bahwa jika terjadi penurunan pada variabel tersebut maka akan berakibat pada peningkatan persentase TPT perempuan.

Pada pemodelan TPT perempuan di Pulau Jawa ini model yang digunakan adalah model tobit. Dari hasil pengolahan disajikan pada Tabel 1. Hasil statistik uji LR pada model kedua adalah 167,78 yang lebih besar dari 14,067 sehingga dengan menggunakan taraf $\alpha = 5\%$, paling sedikit terdapat satu $\beta_k \neq 0$ Variabel yang berpengaruh secara signifikan pada taraf $\alpha = 5\%$ adalah persentase penduduk yang tinggal di perkotaan (X_1), seks rasio (X_3) dan tingkat pertumbuhan ekonomi (X_6).

Tabel 2. Penaksiran Parameter Model Tobit

Variabel	Estimasi	SE	Z
Konstan	-51,91	18,13	-2,86*
X_1	0,0751	0,0267	2,81*
X_2	0,0593	0,5279	0,11
X_3	0,7461	0,1677	4,45*
X_4	0,0212	0,0638	0,33
X_5	0,0186	0,1258	0,15
X_6	-1,0022	0,4848	-2,07*
X_7	-0,1499	0,1017	-1,47
LR = 167,7799 $\chi^2_{(0,05;7)} = 14,067$ $R^2 = 65,60\%$			

Terdapat beberapa variabel prediktor yang tidak signifikan. Oleh karena itu diperlukan tahapan seleksi variabel prediktor untuk mendapatkan model terbaik. Seleksi variabel dilakukan dengan menggunakan metode *Backward Elimination*. Hasil pengolahan disajikan pada Tabel 3.

Tabel 3. Penaksiran Parameter Model Tobit Hasil *Backward Elimination*

Variabel	Estimasi	SE	Z
Konstan	-69,72	12,41	-5,62*
X ₁	0,0903	0,0242	3,72*
X ₃	0,7868	0,1388	5,67*
X ₄	0,0519	0,0190	2,72*
X ₅	0,0277	0,0102	2,74*
X ₆	-0,8898	0,3776	-2,36*
LR = 140,813 $\chi^2_{(0,05;7)} = 11,075$ R ² = 75,87 %			

Berdasarkan hasil *Backward Elimination* pada Tabel 2 dapat diketahui bahwa semua variabel prediktor dalam model sudah signifikan. Model tobit untuk TPT Perempuan adalah sebagai berikut

$$\hat{y} = -69,72 + 0,0903X_1 + 0,7868X_3 + 0,0519X_4 + 0,0277X_5 - 0,8898X_6$$

Hasil statistik uji LR adalah 140,813 yang lebih besar dari 11,0705 sehingga dengan menggunakan taraf $\alpha = 5\%$ paling sedikit terdapat satu $\beta_k \neq 0$ Variabel yang berpengaruh secara signifikan pada taraf $\alpha = 5\%$ adalah persentase penduduk yang tinggal di perkotaan (X₁), seks rasio (X₃), persentase penduduk yang yang berpendidikan di atas SLTP (X₄), persentase penduduk yang mampu membaca dan menulis (X₅) dan dan tingkat pertumbuhan ekonomi (X₆). Variabel yang bertanda positif yaitu persentase penduduk yang tinggal didaerah perkotaan, seks rasio, persentase penduduk yang berpendidikan di atas SLTP, persentase penduduk yang bisa membaca dan menulis, akan mengakibatkan meningkatnya TPT Perempuan sedangkan yang bertanda negatif yaitu pertumbuhan ekonomi akan menurunkan TPT Perempuan

IV. SIMPULAN DAN SARAN

Hasil pemodelan TPT Perempuan dengan menggunakan *tobit* adalah sebagai berikut

$$\hat{y} = -69,72 + 0,0903X_1 + 0,7868X_3 + 0,0519X_4 + 0,0277X_5 - 0,8898X_6$$

Faktor-faktor yang mempengaruhi TPT Perempuan di Pulau Jawa adalah persentase penduduk yang tinggal didaerah perkotaan, seks rasio, persentase penduduk yang berpendidikan di atas SLTP, persentase penduduk yang bisa membaca dan menulis, serta pertumbuhan ekonomi. Penelitian ini belum memperhatikan adanya keterkaitan antar wilayah dalam pemodelan. Perlu dilakukan pemodelan tobit spasial untuk mengatasi adanya keterkaitan antar wilayah yang saling berdekatan.

DAFTAR PUSTAKA

- [1] BPS, "Booklet Agustus 2012", Jakarta: BPS RI, 2013.
- [2] W.H. Greene, "Econometrics Analysis, 6th edition," New Jersey: Prentice Hall, 2008.
- [3] S.H. Cox, and Y. Linn, "Annuity Lapse Rate Modeling: Tobit Or Not Tobit", Journal of economic and social measurement, vol.38, pp: 432, 2006.
- [4] J. Tobin, "Estimation of Relationships for Limited Dependent Variables", Econometrica, vol.26(1); pp: 24-36, 1958.
- [5] I.Y. Suhardi, and R. Llewlyn, "Penggunaan Model Regresi Tobit untuk Menganalisa Faktor-Faktor yang Berpengaruh terhadap Kepuasan Konsumen untuk Jasa Pengangkutan Barang", Jurnal Manajemen & Kewirausahaan, vol.3(2), pp: 106-112, 2001.
- [6] M.D. Hadad, W. Santoso, and A. Alisjahbana, "Model dan Estimasi Permintaan dan Penawaran Kredit Konsumsi Rumah Tangga di Indonesia", Jurnal Bank Indonesia, pp: 1-25, 2004.

- [7] B. Karli, & A. Bilgic, "Factors Affecting Meat And Meat Products Consumption Quantities In Sanliurfa Province", *Mediterranean Agricultural Sciences*, vol.20(1), pp: 127-136, 2007
- [8] D. Purnomo, "Fenomena Migrasi Tenaga Kerja dan Perannya Bagi Pembangunan Daerah Asal: Studi Empiris di Kabupaten Wonogiri", *Jurnal Ekonomi Pembangunan*, vol.10(1), p: 84-102, 2009.
- [9] S. Chen. "Nonparametric Identification and Estimation of Truncated Regression Models", *Oxford Journals*, vol.77(1), pp: 127-153, 2009.
- [10] Myoung-jae Lee. "Micro-Econometrics: Methods of Moments and Limited Dependent Variables, 2th edition", New York: Springer, 2010.
- [11] J.S. Long, "Regression Models for Categorical and Limited Dependent Variables", California: Sage Publications Inc, 1997.