MODEL *TRUNCATED SPATIAL* PADA DATA TERSENSOR

Defi Yusti Faidah
Departemen Statistika FMIPA Universitas Padjadjaran
defi.yusti@unpad.ac.id

Abstrak

Pada kasus tertentu seringkali ditemui data yang bernilai nol untuk sebagian observasi, dan sisanya memiliki nilai yang beragam. Data yang memiliki struktur tersebut dinamakan data tersensor. Dibutuhkan metode khusus untuk mengolah data tersebut. Penggunaan metode analisis regresi linier klasik untuk melihat hubungan variabel yang sifatnya tersensor dengan variabel prediktor tidak tepat. Untuk mengatasi hal tersebut, maka digunakan suatu model regresi untuk data tersensor yang dikenal dengan nama Model truncated. Akan tetapi seringkali data-data tersensor melibatkan aspek keterkaiatan antar wilayah. Oleh karena itu diperlukan suatu pendekatan yang mengkombinasikan antara model truncated dan spasial. Penelitian ini mengkaji data tersensor dengan pendekatan model truncated spatial. Metode penaksiran parameter yang digunakan adalah Maximum Likelihood Estimation. Penaksiran parameter dilakukan dengan melakukan turunan parsial pertama fungsi ln likelihood terhadap parameter yang akan diestimasi dan kemudian disamakan dengan nol. Penaksiran parameter model truncated spasial tidak bisa langsung diperoleh. Hal ini karena fungsinya berbentuk implisit sehingga diperlukan iterasi Newton Raphson untuk memperoleh estimasi parameternya

Kata Kunci: *Data Tersensor; Model Truncated; Spasial*

1. PENDAHULUAN

Pada umumnya untuk mengetahui hubungan antara beberapa variabel metode statistika yang sering digunakan adalah analisis regesi. Analisis regresi digunakan apabila variabel respon bersifat kontinu. Namun pada beberapa penelitian ditemukan hasil pengukuran pada variabel responnya merupakan data tersensor. Data tersensor merupakan data yang memuat nilai nol pada sebagian observasinya sedangkan untuk sebagian nilai lainnya mempunyai nilai tertentu yang bervariasi. Ciri lain dari data tersensor adalah sebagian nilai dari suatu rentang tertentu ditransformasikan sebagai suatu nilai tunggal atau konstanta (Greene, 2008). Seringkali dalam memodelkan data tersensor, data yang bernilai nol dianggap sebagai outlier dan dibuang dalam pemodelan. Hal ini akan menyebabkan banyak informasi yang hilang. Pendekatan yang bisa digunakan untuk mengatasi data tersensor ada model truncated. Penggunaan model truncated akan mengurangi efek bias jika dibandingkan dengan data yang diolah menggunakan regresi linier klasik.

Seringkali data tersensor melibatkan aspek wilayah sehingga diperlukan metode khusus yang menggabungkan model truncated dengan aspek spasial. Aspek spasial ini penting untuk dikaji, karena antara satu wilayah dengan wilayah lain mempunyai perbedaan karakteristik. Menurut Anselin (1988) pelibatan data spasial pada model regresi linier klasik bisa menyebabkan hasil akhir yang kurang tepat. Hal ini dikarenakan asumsi error yang saling bebas dan homegen tidak terpenuhi.

Penelitian tentang model truncated spasial pernah dilakukan oleh Langyintuo dan Merkuria (2008) yang diaplikasikan pada penggunaan varietas jagung unggulan oleh petani di Mozambique. Kaliba (2002) menggunakan pendekatan tobit spasial untuk menganalisis data pertanian di Tanzania. Kajian utama pada penelitian ini adalah melakukan penaksiran parameter model truncated spasial yang diaplikasikan pada kasus Tingkat Pengangguran Terbuka (TPT) Perempuan di Pulau Jawa. Hal ini karena pada penelitian sebelumnya belum dikaji secara rinci tentang penaksiran parameter model. Metode penaksiran parameter model menggunakan maximum likelihood estimation (MLE).

2. METODE PENELITIAN

a. Sumber Data

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.

b. Variabel Penelitian

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 daerah perkotaan (X_1), angka pertumbuhan penduduk (X_2), seks rasio (X_3), persentase penduduk yang berpendidikan di atas SLTP (X_4), persentase penduduk yang mampu membaca dan menulis (X_5), tingkat pertumbuhan ekonomi (X_6), dan partisipasi angkatan kerja (X_7).

c. Metode Analisis

Langkah-langkah mengkaji penaksiran parameter model truncated spasial dengan metode *Maximum Likelihood Estimation* (MLE) dilakukan dengan langkah-langkah :

1) Membuat fungsi likelihood berdasarkan model regresi tobit spasial.
2) Membuat fungsi log likelihood dengan melakukan transformasi log pada fungsi likelihood.
3) Menaksir parameter $\boldsymbol{\beta}$, ρ, λ dan σ^2 dengan melakukan turunan pertama terhadap parameter yang akan diestimasi dan kemudian disamakan dengan nol.
4) Melakukan turunan parsial kedua terhadap parameter yang akan diestimasi.
5) Melakukan metode iterasi Newton-Raphson untuk mendapatkan estimasi parameter.
6) Mengaplikasikan pada data TPT Perempuan

3. HASIL PENELITIAN DAN PEMBAHASAN

Penaksiran parameter model tobit spasial dengan metode maksimum likelihood ini diawali dengan membuat fungsi likelihood. Fungsi likelihood model regresi tobit spasial terdiri dari data yang tersensor dan data yang tidak tersensor. Jika diambil sampel sebanyak *n*, maka fungsi likelihood untuk model tobit spasial adalah sebagai berikut.

$$L(\rho,\lambda,\boldsymbol{\beta}\,|\,y_i,\mathbf{x_i}) = \prod_{sensor}\Phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)\prod_{unsensor} f\left(y_i = y_i^* \mid \mathbf{x}_i\right)$$

$$= \prod_{sensor}\Phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)\prod_{unsensor}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}\left(y_i-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}\mathbf{u}\right)^2\right]\right]$$

Langkah selanjutnya adalah membentuk fungsi *ln-likelihood* dengan cara melakukan transformasi ln pada fungsi likelihood, sehingga diperoleh persamaan (1)

$$\ln L(\rho,\lambda,\boldsymbol{\beta}\,|\,y_i,\mathbf{x_i}) = \ln\left(\prod_{sensor}\Phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)\prod_{unsensor}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}\left(y_i-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}\mathbf{u}\right)^2\right]\right]\right)$$

$$= \sum_{sensor}\ln\Phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)+$$

$$-\frac{1}{2}\sum_{unsensor}\left(\frac{y_i-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)^2 \qquad (1)$$

Estimasi parameter dilakukan dengan melakukan turunan parsial pertama persamaan (1) terhadap parameter yang akan diestimasi dan kemudian disamakan dengan nol sehingga diperoleh :

a. Untuk menaksir koefisien ρ

$$\frac{\partial\ln(L)}{\partial\rho}=0$$

$$\sum_{sensor}\frac{-\frac{\mathbf{w}_{1i}^T\mathbf{y}}{\sigma}\phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}{\Phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}+\frac{\mathbf{w}_{1i}^T\mathbf{y}}{\sigma}\sum_{unsensor}\left(\frac{y_i-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)=0$$

$$-\frac{\mathbf{w}_{1i}^T\mathbf{y}}{\sigma}\left(\sum_{sensor}\frac{\phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}{\Phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}-\sum_{unsensor}\left(\frac{y_i-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)\right)=0$$

(2)

b. Untuk menaksir koefisien λ

$$\frac{\partial \ln(L)}{\partial \lambda}=0$$

$$\sum_{sensor}\frac{-\frac{\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}{\Phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}+\frac{\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\sum_{unsensor}\left(\frac{y_i-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)=0$$

$$-\frac{\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\left(\sum_{sensor}\frac{\phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}{\Phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}-\sum_{unsensor}\left(\frac{y_i-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)\right)=0$$

(3)

c. Untuk menaksir koefisien $\boldsymbol{\beta}$

$$\frac{\partial \ln(L)}{\partial \boldsymbol{\beta}}=0$$

$$\sum_{sensor}\frac{-\frac{\mathbf{x}_i^T}{\sigma}\phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}{\Phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}+\frac{\mathbf{x}_i^T}{\sigma}\sum_{unsensor}\left(\frac{y_i-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)=0$$

$$-\frac{\mathbf{x}_i^T}{\sigma}\left(\sum_{sensor}\frac{\phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}{\Phi\left(\frac{c-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)}-\sum_{unsensor}\left(\frac{y_i-\rho\mathbf{w}_{1i}^T\mathbf{y}-\mathbf{x}_i^T\boldsymbol{\beta}-\lambda\mathbf{w}_{2i}^T\mathbf{u}}{\sigma}\right)\right)=0 \quad (4)$$

Berdasarkan hasil turunan pertama pada persamaan (2), (3) dan (4) ternyata penaksir untuk $\rho,\lambda,$ dan $\boldsymbol{\beta}$ tidak bisa langsung diperoleh karena fungsinya berbentuk implisit sehingga diperlukan suatu metode numerik untuk memperoleh estimasi parameternya. Metode numerik yang dapat digunakan adalah metode Iterasi Newton Raphson.Persamaan yang digunakan dalam proses iterasi Newton Rahpson untuk mendapatkan nilai $\hat{\boldsymbol{\theta}}^{(t)}$ adalah

$$\boldsymbol{\theta}^{(t+1)}=\boldsymbol{\theta}^{(t)}-\left[\mathbf{H}^{(t)}\right]^{-1}\mathbf{q}^{(t)}$$

Proses iterasi Newton-Raphson ini akan berhenti jika terpenuhi kondisi konvergen, yaitu selisih $\left\|\boldsymbol{\theta}^{(t+1)}-\boldsymbol{\theta}^{(t)}\right\|<\varepsilon$, dimana ε adalah bilangan yang sangat kecil.

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 tersbut memiliki TPT perempuan yang berkisar antara 7,14-8,58 persen. Sementara itu wilayah yang memiliki TPT perempuan yang tinggi mengelompok di Provinsi Jawa Tengah bagian barat, Jawa Barat, Banten, dan DKI Jakarta. Kabupaten/kota dengan TPT perempuan yang cukup tinggi adalah Kota Cilegon (28,88 persen).

Gambar 1. Persentase TPT perempuan di Pulau Jawa

b. Pemodelan TPT Perempuan

Pada pemodelan TPT perempuan di Pulau Jawa ini model yang digunakan adalah model truncated spasial dengan dependensi spasial lag. Dari hasil pengolahan disajikan pada Tabel 1.

Tabel 1. Estimasi Parameter dengan Menggunakan Tobit Spasial Dependensi Lag

Tobit Spasial Dependensi Lag			
Variabel	**Estimasi**	**SE**	**Z**
Konstan	-70,5494	15,8761	-4,4438*
X_1	0,1011	0,0299	3,3799*
X_2	0,6155	0,5358	1,1487
X_3	0,6419	0,1758	3,6516*
X_4	0,0921	0,0662	1,3923
X_5	0,0819	0,0579	1,4142
X_6	-0,8887	0,4636	-1,9167**
X_7	-0,0104	0,0557	- 0,1861
rho	0,9689	0,1224	7,9156*

LR = 176,7799

$\chi^2_{(0,05;7)} = 14,067$

R^2= 65,60 %

Ket: *)Signifikan pada α = 5%, **)Signifikan pada α = 10%

Berdasarkan Tabel. 1 dapat diketahui bahwa menghasilkan koefisien lag yang signifikan pada taraf $\alpha = 5\%$. Variabel yang berpengaruh secara signifikan pada taraf $\alpha = 5\%$ adalah persentase penduduk yang tinggal di perkotaan (X_1), dan seks rasio (X_3), sementara itu tingkat pertumbuhan ekonomi (X_6) berpengaruh pada taraf $\alpha = 10\%$. Akan tetapi terdapat beberapa variabel prediktor yang tidak signifikan. Oleh karena itu diperlukan tahapan seleksi variabel prediktor untuk mendapatkan model terbaik. Hasil pengolahan disajikan pada Tabel 2.

Tabel 2. Estimasi Parameter Tobit Spasial Dependensi Lag dengan *Bacward Elimination*

Variabel	Estimasi	SE	Z
Konstan	-58,5047	11,9008	-4,9160*
X_1	0,0981	0,0273	3,5930*
X_3	0,5138	0,1364	3,7670*
X_4	0,1079	0,0612	1,7642**
X_5	0,0969	0,0342	2,8375*
X_6	-0,8352	0,4592	-1,8187**
Rho	0,9559	0,1225	7,8049*

LR = 143,2813

$\chi^2_{(0,05;5)} = 11{,}0705$

R^2= 65,14 %

Ket: *)Signifikan pada α = 5%, **)Signifikan pada α = 10%

Berdasarkan hasil *Backward Elimination* pada Tabel 3 dapat diketahui bahwa semua variabel prediktor dalam model sudah signifikan. Untuk kabupaten/kota yang memiliki TPT perempuan tergolong tinggi mengikuti persamaan (5).

$$\hat{y}_i = -58{,}505 + 0{,}956 \sum_{i=1\, j\neq 1}^{n} w_{ij} y_i + 0{,}098 x_{i1} + 0{,}514 x_{3i} + 0{,}108 x_{4i} + 0{,}097 x_{5i} - 0{,}835 x_{6i} \quad (5)$$

Untuk kabupaten/kota yang memiliki TPT perempuan yang tergolong rendah adalah $\hat{y}_i = 7{,}14$

Variabel yang berpengaruh secara signifikan pada taraf $\alpha = 5\%$ adalah persentase penduduk yang tinggal di perkotaan (X_1), seks rasio (X_3), dan persentase penduduk yang mampu membaca dan menulis (X_5). Sementara itu persentase penduduk yang yang berpendidikan di atas SLTP (X_4) dan tingkat pertumbuhan ekonomi (X_6) berpengaruh pada taraf $\alpha = 10\%$. Koefisien determinasi (R^2) yang dihasilkan adalah 65,14 persen. Hal ini berarti bahwa model regresi tobit SAR mampu menjelaskan variasi dari TPT perempuan sebesar 65,14 persen, sedangkan sisanya dijelaskan oleh variabel lain diluar model.

4. SIMPULAN

Berdasarkan hasil dan pembahasan diperoleh kesimpulan bahwa penaksiran parameter model truncated spasial dengan menggunakan MLE tidak bisa langsung diperoleh. Hal ini karena fungsinya berbentuk implisit sehingga diperlukan iterasi Newton Raphson untuk memperoleh estimasi parametrnya.

5. DAFTAR PUSTAKA

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Dordrecht: Kluwer Academic.

Greene, W. H. (2008). *Econometrics Analysis*. 6th edition. New Jersey, NJ:Prentice Hall.

Kaliba, A.R. (2002). *Dissertation: Participatory Evaluation of Community Based Water and Sanitation Programes: The Case of Central Tanzania*. Mahattan: Kansas State University.

Langyintuo, A.S. & Merkuria, M. (2008). Assesing the Influence of Neighborhood Effects on the Adoption of Improved Agricultural Technologies in Developing Agriculture. *AfJARE*, **2**(2). 151-169.