

ANALISIS REGRESI TOBIT SPASIAL : Studi Kasus Penggunaan Internet di Pulau Jawa

Andhie Surya Mustari¹, Ismaini Zain²

¹Badan Pusat Statistik, email: andhie@bps.go.id

²Institut Teknologi Sepuluh Nopember,
email: ¹andhie@bps.go.id, ²ismaini_z@statistika.its.ac.id

Abstrak

Dibutuhkan metode khusus untuk menganalisis data tersensor yang memiliki korelasi spasial. Jika menggunakan regresi linier, akan menghasilkan estimasi parameter yang tidak valid, tidak terpenuhinya asumsi normalitas dan mengaburkan interpretasi model. Model regresi Tobit spasial digunakan untuk menganalisis data penggunaan internet di Pulau Jawa. Estimasi parameter menggunakan metode MCMC Gibbs sampler dengan pendekatan inferensia Bayesian. Hasilnya, penggunaan internet di Pulau Jawa dipengaruhi oleh persentase penduduk yang tinggal di daerah perkotaan, persentase penduduk lulusan SMA ke atas, rata-rata lama sekolah, persentase rumah tangga yang memiliki telepon genggam, dan persentase desa/kelurahan yang mendapatkan sinyal telepon seluler.

Kata kunci: data tersensor, korelasi spasial, Tobit spasial, MCMC, penggunaan internet

Abstract

Special method is required for analyzing censored data with spatial dependence. Using linear regression will results in invalid parameter estimations, normality assumption violations, and obscure the model interpretation. Spatial Tobit regression model is used to analyze the data of internet usage in Java. MCMC Gibbs sampler method with Bayesian inference approach was used for parameter estimation. As a result, internet usage in Java Island is influenced by the percentage of population living in urban areas, the percentage of population graduated from senior high school, the average length of school, the percentage of households with mobile phones, and the percentage of villages receiving cell phone signal.

Keywords: *censored data, spatial dependence, spatial Tobit, MCMC, internet usage*

PENDAHULUAN

Menggunakan model regresi linier klasik untuk analisis data tersensor yang memiliki korelasi spasial merupakan keputusan yang kurang tepat. Istilah data tersensor digunakan untuk menjelaskan sekelompok data yang memiliki sejumlah nilai yang tidak diketahui pada batas atas atau bawahnya. Long (1997) menjelaskan bahwa jika menggunakan model regresi linier pada keseluruhan data tersensor, akan menghasilkan nilai parameter yang *overestimates* pada *slope* dan *underestimates* pada *intercept*. Sedangkan jika menghilangkan atau memotong observasi yang nilainya tidak diketahui, akan menghasilkan koefisien parameter yang *underestimates* pada *slope* dan *overestimates* pada *intercept*. Data terpotong menyebabkan terjadinya korelasi antara variabel prediktor dengan residual, sehingga menghasilkan estimasi yang tidak konsisten. Efek korelasi spasial dapat muncul pada pembentukan model regresi linier yang menggunakan data kewilayahan (*cross section data*). Hal tersebut mengakibatkan tidak terpenuhinya asumsi error yang independen dan identik berdistribusi normal, sehingga menghasilkan estimasi parameter yang tidak valid dan mengaburkan interpretasi model (Marsh, Mittelhammer, & Huffaker, 2000). Korelasi spasial dapat diamati dari mengelompoknya besaran nilai tertentu pada data yang berasal dari wilayah yang berdekatan, misalnya data tingkat penggunaan internet kabupaten/kota di Pulau Jawa.

Tingginya tingkat penggunaan internet di Pulau Jawa terutama ditemukan di kota-kota besar sebagai pusat jasa pendidikan dan hiburan, seperti DKI Jakarta, Yogyakarta, Bandung, dan Surabaya, kemudian diikuti oleh wilayah kabupaten/kota lain di sekitarnya. Fenomena dependensi spasial ini dapat diaplikasikan untuk analisis data tersensor yang memiliki korelasi spasial, dimana kabupaten/kota dengan penggunaan internet kategori tinggi dapat dianggap sebagai data yang tidak diketahui nilainya. Sebanyak 34 dari 118 kabupaten/kota di Pulau Jawa memiliki persentase pengguna internet yang lebih tinggi daripada 16 persen

(BPS, 2011), suatu nilai yang setara dengan jumlah akun *facebook* yang dibuat oleh penduduk Indonesia pada tahun 2011 (Socialbaker, 2011).

Dibutuhkan metode khusus untuk melakukan analisis penggunaan internet, dengan asumsi bahwa data penggunaan internet di Pulau Jawa merupakan data tersensor yang memiliki korelasi spasial. Fischer dan Getis (2010) mengatakan bahwa pemodelan data tersensor yang melibatkan wilayah sebaiknya menggunakan analisis spasial, metode yang paling sesuai adalah regresi Tobit spasial. Selain itu, Lee (2010) juga menyatakan bahwa pendekatan Tobit spasial lebih disarankan untuk analisis wilayah yang melibatkan data tersensor. Analisis regresi Tobit spasial digunakan apabila variabel respon pada model spasial melibatkan data yang diyakini memiliki nilai tersensor (LeSage & Pace, 2009).

Penelitian ini bermaksud untuk membentuk model regresi Tobit spasial dan mencari metode estimasi parameter dari model regresi Tobit spasial. Data yang digunakan sebagai variabel respon adalah persentase penduduk umur 5 tahun ke atas yang mengakses internet selama tiga bulan terakhir di pulau Jawa pada tahun 2010. Sensor diberikan kepada wilayah kabupaten/kota dengan persentase pengguna internet lebih besar dari 16 persen, yang dianggap sebagai batas minimal persentase pengguna internet yang ingin dicapai oleh suatu kabupaten/kota.

METODOLOGI

a. Tinjauan Referensi

Model Regresi Tobit

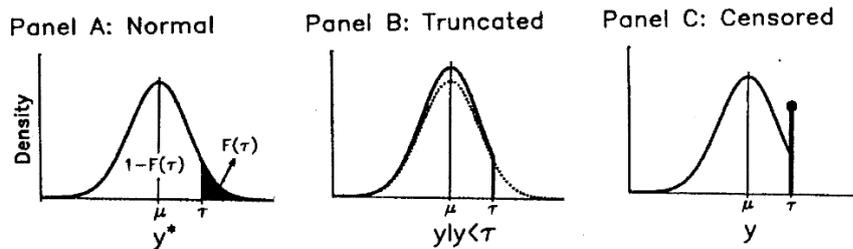
Misalkan Y^* adalah suatu variabel respon dengan informasi yang lengkap dan Y adalah data sampel dari Y^* , maka variabel respon Y yang tersensor dapat didefinisikan sebagai berikut:

$$y_i = \begin{cases} y_i^* & \text{jika } y_i^* < \tau \\ \tau & \text{jika } y_i^* \geq \tau \end{cases} \quad (1)$$

dimana $i = 1, 2, \dots, n$, τ merupakan suatu konstanta batasan, dan n adalah banyaknya observasi (Tobin, 1958)¹.

Jika nilai y_i^* tidak diketahui ketika $y_i^* \geq \tau$, maka y^* mengandung suatu variabel latent yang tidak dapat diamati pada seluruh *range*-nya. Panel A pada Gambar 2.1 memperlihatkan distribusi dari $y^* \sim N(\mu, \sigma^2)$, dengan nilai variabel latent

digambarkan sebagai wilayah gelap pada kurva. Jika nilai yang tidak diketahui tersebut dipotong, maka sebagian informasi yang dapat menjelaskan populasi akan hilang dimana kurva distribusi menjadi lebih runcing (panel B). Panel C menggambarkan data tersensor yang mengelompok pada nilai τ sehingga tidak merubah informasi berkaitan distribusi populasi.



Gambar 1. Ilustrasi Variabel Latent, Terpotong, dan Tersensor (Long, 1997)

Model Tobit dibentuk dengan terlebih dahulu mengasumsikan adanya hubungan linier antara Y^* dengan variabel prediktor X yang dinyatakan dengan:

$$y_i^* = x_i^T \beta + \varepsilon_i \quad (2)$$

dimana $i = 1, 2, \dots, n$, $\varepsilon_i \sim N(0, \sigma^2)$, $x_i^T = [1 \ x_{1i} \ x_{2i} \ \dots \ x_{pi}]$ adalah vektor variabel prediktor, $\beta = [\beta_0 \ \beta_1 \ \dots \ \beta_p]^T$ adalah vektor parameter, dan p merupakan banyaknya variabel, dengan y^* mengandung variabel latent yang mewakili nilai tersensor². Nilai tersensor tersebut bisa lebih kecil dari suatu batas bawah ($y_i^* \leq \tau < \bar{y}^*$), lebih besar dari batas atas ($y_i^* \geq \tau > \bar{y}^*$), atau keduanya.

Ketika y^* tersensor pada batas atas $y_i^* \geq \tau$, maka model regresi Tobit dinyatakan dengan persamaan sebagai berikut:

$$y_i = \begin{cases} y_i^* = x_i^T \beta + \varepsilon_i & \text{jika } y_i^* < \tau \\ \tau & \text{jika } y_i^* \geq \tau \end{cases} \quad (3)$$

Model Regresi Spasial

Pada tahun 1988, Anselin mengembangkan bentuk umum dari model regresi spasial (*general spatial model*)

menggunakan data *cross section* sebagai berikut:

$$y = (I - \rho W)^{-1} X \beta + (I - \rho W)^{-1} (I - \lambda W)^{-1} \varepsilon \quad (4)$$

dimana $\varepsilon \sim N(0, \sigma^2 I)$, y merupakan vektor variabel respon yang memiliki korelasi spasial, X adalah matriks variabel prediktor, dan β adalah vektor parameter regresi. Adapun ρ adalah koefisien korelasi spasial lag dari variabel respon, λ merupakan koefisien korelasi spasial error, dan W merupakan matriks penimbang spasial dengan elemen diagonalnya bernilai nol. Persamaan ini juga biasa disebut sebagai model regresi *spatial autoregressive moving average* (SARMA).

$$\begin{aligned} y &= [y_1 \ y_2 \ \dots \ y_n]^T, \\ \beta &= [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_p]^T, \\ \varepsilon &= [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n]^T, \\ X &= \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \text{ dan} \end{aligned}$$

¹ Model regresi Tobit pertama kali diperkenalkan oleh James Tobin (1958), yang dijelaskan kembali oleh Long (1997), DeMaris (2004), Greene (2008), dan Lee (2010).

² Sebagaimana dijelaskan oleh Long (1997) dan Greene (2008).

$$W = \begin{bmatrix} 0 & w_{12} & \dots & w_{1n} \\ x_{21} & 0 & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & 0 \end{bmatrix}$$

Persamaan (4) di atas menjadi bentuk model regresi spasial lag ketika $\lambda = 0$, yang menjelaskan terjadinya efek korelasi spasial antar lag variabel respon. Ketika $\rho = 0$, persamaan (4) menjadi bentuk model regresi spasial error, yang menjelaskan terjadinya efek korelasi spasial antar lag variabel prediktor. Statistik uji Lagrange digunakan untuk menentukan bentuk dari model regresi spasial (Anselin, 1999).

Untuk melihat ada atau tidaknya efek korelasi spasial lag, digunakan hipotesis $H_0: \rho = 0$ lawan $H_1: \rho \neq 0$ dengan statistik uji *Lagrange Multiplier* atau LM-lag Test sebagai berikut:

$$LM_{LAG} = \frac{(e^T W y / \hat{\sigma}^2)^2}{D} \quad (5)$$

Untuk melihat adanya atau tidaknya efek spasial error, digunakan hipotesis $H_0: \lambda = 0$ lawan $H_1: \lambda \neq 0$ dengan statistik uji LM-err Test sebagai berikut:

$$LM_{ERR} = \frac{(e^T W e / \hat{\sigma}^2)^2}{T} \quad (6)$$

dimana $\hat{\sigma}^2 = e^T e / n$, $e = y - X\hat{\beta}$, $D = \left[(WX\hat{\beta})^T (I - X(X^T X)^{-1} X^T) (WX\hat{\beta}) \right] / \sigma^2 + T$

, dan $T = \text{tr} \{ (W^T + W)W \}$. Statistik uji ini mengikuti distribusi asimtotik $\chi^2_{(\alpha,1)}$, sehingga H_0 ditolak apabila $LM > \chi^2_{(\alpha,1)}$ atau p -value lebih kecil dari nilai α .

Markov Chain Monte Carlo (MCMC)

MCMC merupakan suatu teknik metode simulasi yang membangkitkan sejumlah sampel dari distribusi data yang telah diketahui (Chib & Greenberg, 1996). Ide dasar dari teknik MCMC adalah daripada menghitung suatu fungsi kepadatan peluang $p(\theta|y)$, lebih baik mengambil sampel random dalam jumlah besar dari $p(\theta|y)$ untuk mengetahui bentuk probabilitas tersebut secara tepat. Dengan ukuran sampel random yang cukup besar, nilai rata-rata dan standar deviasinya dapat dihitung secara akurat (Casella & George,

1992). LeSage (1999) menjelaskan bahwa algoritma MCMC *Gibbs sampler* akan memberi kemudahan estimasi parameter untuk model regresi Tobit spasial daripada harus memecahkan sejumlah persamaan integral pada metode maksimum likelihood. Metode MCMC *Gibbs sampler* bertujuan untuk mencari nilai estimasi dari θ_i menggunakan suatu distribusi posterior bersyarat, dimana nilai θ_{-i} lainnya diasumsikan telah diketahui. Distribusi posterior dari parameter θ ditentukan melalui prinsip dari teorema Bayes yang dinyatakan oleh:

$$p(\theta|y) = \frac{L(\theta|y)p(\theta)}{m(y)} \propto L(\theta|y)p(\theta) \quad (7)$$

dimana $L(\theta|y)$ merupakan fungsi likelihood dari y , $m(y)$ merupakan distribusi marginal dari y yang tidak melibatkan parameter, dan $p(\theta)$ merupakan distribusi prior dari θ yang diperoleh dari penelitian sebelumnya maupun berdasarkan kajian teoritis atas masalah yang sedang diteliti (Casella & Berger, 2002).

Hastings (1970) mengembangkan metode *Metropolis* untuk mencari estimasi parameter $\theta^{(t)}$ melalui suatu nilai inisiasi awal $\theta^{(0)}$, yang didasari oleh distribusi kandidat $f(\theta^{(*)}|\theta^{(t-1)})$ dimana nilai $\theta^{(t-1)}$ diketahui. Nilai dari $\theta^{(*)}$ dibangkitkan dari distribusi kandidat, kemudian terima $\theta^{(*)}$ sebagai $\theta^{(t)}$ jika $p(\theta^{(*)}, \theta^{(t-1)}) > q$. Dimana $q \sim \text{Uniform}(0,1)$ dan $p(\theta^{(*)}, \theta^{(t-1)}) =$

$$\min \left[1, \frac{p(\theta^{(*)}|y)f(\theta^{(t-1)}|\theta^{(*)})}{p(\theta^{(t-1)}|y)f(\theta^{(*)}|\theta^{(t-1)})} \right] \quad (8)$$

Penelitian Sebelumnya

Penelitian yang menggunakan model regresi Tobit spasial di antaranya dilakukan oleh Langyintuo dan Mekuria (2008) yang menggunakan metode maksimum likelihood untuk membentuk model Tobit SARMA pada data petani di Mozambique. Kaliba (2002) mengembangkan model Tobit SARMA menggunakan modul aplikasi Maximum Likelihood 4 dari paket program GAUSS (dikembangkan oleh Aptech Systems, 1995) pada data pedesaan di

Tanzania. Sementara LeSage dan Pace (2009) menggunakan data simulasi yang dibangkitkan oleh Koop untuk membentuk model Tobit spasial menggunakan pendekatan Bayesian MCMC (*Markov Chain Monte Carlo*) dengan algoritma Gibbs Sampling. Adapun penelitian yang menggunakan model Tobit spasial pada data teknologi informasi dan komunikasi (TIK) belum pernah ditemukan.

Pembangunan TIK suatu negara memiliki hubungan yang positif dengan pertumbuhan ekonomi. Artinya, pembangunan TIK akan memberikan efek berantai kepada meningkatnya pertumbuhan ekonomi (Kominfo, 2010). Rao dan Pattnaik (2006) menyatakan bahwa pertumbuhan TIK telah membuka kesempatan bagi masyarakat untuk lebih memanfaatkan fasilitas pembangunan sosial ekonomi dan budaya secara lebih modern. Pembangunan TIK memberikan pengaruh ekonomi yang luas, baik secara langsung maupun tidak langsung, meningkatkan kesejahteraan dan pembangunan fasilitas sosial ekonomi (ITU, 2010). Akses internet merupakan indikator yang paling merepresentasikan tingkat pembangunan TIK suatu negara, selain pertumbuhan ekonomi sektor telekomunikasi, kepemilikan telepon seluler atau penguasaan komputer.

Menggunakan data dari 154 negara, Howard dan Mazaheri (2009) menemukan bahwa kesenjangan penggunaan TIK (telepon seluler, komputer, dan bandwidth internet) dipengaruhi oleh; investasi asing, perdagangan, jumlah penduduk, populasi perkotaan, literacy rate, konsumsi, telepon kabel, serta sembilan variabel lain yang menjelaskan regulasi pemerintah. Andonova dan Serrano (2007) menjelaskan bahwa perkembangan TIK dan pertumbuhan pemanfaatan internet lebih banyak dipengaruhi oleh faktor perhatian pemerintah dan regulasi yang berlaku di wilayah tersebut. Michailidis dkk. mengungkapkan bahwa pengguna internet di pedesaan Yunani dipengaruhi oleh tingkat pendapatan, harga dari akses internet, kepemilikan PC, tempat tinggal, serta variabel sosial demografi seperti; jenis kelamin, jumlah penduduk muda yang

tinggal satu rumah, umur, tingkat pendidikan, dan status pekerjaan (Michailidis, Partalidou, Nastis, Klavdianou, & Charatsari, 2011).

Berdasarkan hasil-hasil pada penelitian terdahulu, model regresi Tobit spasial yang dibangun pada penelitian ini akan menggunakan algoritma MCMC sebagai metode estimasi parameternya. Tingkat penggunaan internet di kabupaten/kota di Pulau Jawa digunakan sebagai variabel respon dengan variabel-variabel prediktor sebagai berikut; persentase penduduk yang tinggal di daerah perkotaan, persentase penduduk usia muda, persentase penduduk lulusan SMA ke atas, rata-rata lama sekolah, persentase rumah tangga yang memiliki komputer, persentase rumah tangga yang memiliki telepon seluler, dan persentase desa/kelurahan yang mendapatkan sinyal telepon seluler.

b. Metode Analisis

Sumber Data dan Variabel Penelitian

Sumber data yang digunakan dalam penelitian ini adalah data olahan hasil Susenas 2010 dan Podes 2008 yang dikumpulkan oleh Badan Pusat Statistik (BPS). Matriks penimbang spasial W disusun menggunakan metode *queen contiguity*, yaitu daerah kabupaten/kota yang saling berbatasan wilayah akan memiliki korelasi spasial sedangkan yang saling terpisah tidak memiliki korelasi. Nilai $w_{i,j} = 1$ jika daerah i dan j saling berbatasan wilayah, menjadi $w_{i,j} = 0$ jika tidak saling berbatasan. Adapun peta digital yang digunakan berdasarkan hasil kegiatan *updating* peta Sensus Penduduk 2010.

Objek penelitian yang dijadikan sebagai variabel respon adalah tingkat penggunaan internet di 118 kabupaten/kota di Pulau Jawa, yaitu persentase penduduk usia 5 tahun ke atas yang pernah mengakses internet dalam tiga bulan terakhir. Sensor diberikan kepada wilayah kabupaten/kota dengan persentase penduduk pengguna internet di atas 16 persen dengan menganggap nilai $y_i = 16$ pada nilai $y_i^* \geq 16$. Adapun variabel prediktor yang digunakan adalah sebagai berikut:

- X_1 : Persentase penduduk yang tinggal di daerah perkotaan.
- X_2 : Persentase penduduk usia muda (13-24 tahun).
- X_3 : Persentase penduduk lulusan SMA ke atas.
- X_4 : Rata-rata lama sekolah.
- X_5 : Persentase rumah tangga yang memiliki komputer.
- X_6 : Persentase rumah tangga yang memiliki telepon seluler.
- X_7 : Persentase desa/kelurahan yang mendapatkan sinyal telepon seluler.

Model Regresi Tobit Spasial

Model regresi Tobit spasial merupakan penerapan model regresi spasial pada data yang tersensor. Sehingga dengan menggabungkan persamaan (4) ke dalam (1), akan diperoleh suatu model umum regresi Tobit spasial sebagaimana berikut.

$$y_i = \begin{cases} y_i^* = (I - \rho W)^{-1} X\beta \\ \quad + (I - \rho W)^{-1} (I - \lambda W)^{-1} \varepsilon & \text{jika } y_i^* < \tau \\ \tau & \text{jika } y_i^* \geq \tau \end{cases} \quad (9)$$

dimana τ merupakan suatu nilai konstanta batasan yang diberikan kepada y_i jika y_i^* tersensor, $i = 1, 2, \dots, n$. Persamaan $y^* = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} (I - \lambda W)^{-1} \varepsilon$ merupakan model umum dari regresi spasial yang menjadi spasial lag ketika $\lambda = 0$ dan menjadi spasial error ketika $\rho = 0$, dimana $\varepsilon \sim N(0, \sigma^2 V)$ dengan $\text{diag}(V)$ merupakan pengali yang menyatakan heterogenitas dari varians error. Pada kondisi homoskedastisitas, matriks V menjadi $V = I$ (LeSage, 2000).

Apabila nilai $\rho \neq 0$ dan $\lambda = 0$, persamaan (9) menjadi model regresi Tobit spasial lag sebagai berikut:

$$y_i = \begin{cases} \rho w_i^T y + x_i^T \beta + \varepsilon_i & \text{jika } y_i < \tau \\ \tau & \text{jika } y_i \geq \tau \end{cases} \quad (10)$$

Apabila nilai $\rho = 0$ dan $\lambda \neq 0$, menjadi model regresi Tobit spasial error:

$$y_i = \begin{cases} \lambda w_i^T y + x_i^T \beta - \lambda (w_i^T X) \beta + \varepsilon_i & \text{jika } y_i < \tau \\ \tau & \text{jika } y_i \geq \tau \end{cases} \quad (11)$$

Pengujian heteroskedastisitas untuk menentukan matrik $\text{diag}(V)$ menggunakan

statistik uji Breusch-Pagan (BP) dengan hipotesis sebagai berikut:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

(homoskedastisitas)

H_1 : minimal ada satu $\sigma_i^2 \neq \sigma^2$ (heteroskedastisitas)

Nilai dari BP-Test adalah sebagai berikut:

$$BP = \frac{1}{2} \mathbf{f}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{f} \quad (12)$$

dengan elemen vektor \mathbf{f} adalah

$f_i = \left(\frac{e_i^2}{\hat{\sigma}^2} - 1 \right)$, dimana e_i adalah residual observasi ke- i hasil regresi linier, $i = 1, 2, \dots, n$, $\hat{\sigma}^2 = \mathbf{e}^T \mathbf{e} / n$, \mathbf{X} adalah matriks $n \times (p + 1)$ dari observasi dengan elemen kolom pertama merupakan vektor satu, dan p adalah jumlah variabel prediktor (Breusch & Pagan, 1979). H_0 ditolak apabila $BP > \chi^2_{(\alpha, p)}$.

Melengkapi Data Tersensor

Estimasi parameter model regresi Tobit spasial dilakukan dengan asumsi awal bahwa variabel respon Y merupakan data dengan informasi yang lengkap, tidak tersensor, dan memiliki korelasi spasial. Padahal sesuai persamaan (1), data yang memiliki informasi lengkap adalah variabel respon Y^* yang mengikuti distribusi normal $y_i^* \sim N(\mu, \sigma^2)$. Nilai dari y_i^* ketika $y_i^* \geq \tau$ merupakan observasi yang tidak diketahui atau dianggap sebagai variabel latent. Sehingga nilainya harus dilengkapi menggunakan suatu nilai $y_i^* = z_i$. Variabel respon Y yang lengkap kemudian didefinisikan sebagai berikut:

$$y_i = \begin{cases} y_i^* & \text{jika } y_i^* < \tau \\ z_i & \text{jika } y_i^* \geq \tau \end{cases} \quad (13)$$

Melengkapi observasi tersensor dilakukan dengan cara membangkitkan variabel random Z yang berdistribusi normal $\mathbf{z} \sim N(\mu, \Sigma)$. Nilai z_i merupakan elemen ke- i dari vektor $\mathbf{z} = [z_1 \ z_2 \ \dots \ z_n]^T$ yang digunakan untuk mengganti observasi

tersensor ketika $z_i \geq \tau$. Menurut LeSage (1999), nilai rata-rata dari \mathbf{z} ³:

$$\boldsymbol{\mu} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} \quad (14)$$

dengan nilai varians dari \mathbf{z} adalah:

$$\boldsymbol{\Sigma} = (\mathbf{I} - \rho\mathbf{W})^{-1} (\mathbf{I} - \lambda\mathbf{W})^{-1} \sigma_\varepsilon^2 \mathbf{V} \times [(\mathbf{I} - \rho\mathbf{W})^{-1}]^T [(\mathbf{I} - \lambda\mathbf{W})^{-1}]^T \quad (15)$$

Estimasi Parameter

Sesuai dengan asumsi awal bahwa variabel respon Y merupakan data dengan informasi yang lengkap, tidak tersensor dan memiliki korelasi spasial, maka hubungannya dengan variabel prediktor X diwakili oleh model regresi spasial sesuai dengan persamaan (4) di atas. LeSage (2000) dan Lacombe (2008) merumuskan distribusi posterior bersyarat dari masing-masing parameter adalah sebagai berikut:

$$\left(\frac{1}{\sigma^2} \boldsymbol{\varepsilon}^T \mathbf{V}^{-1} \boldsymbol{\varepsilon}\right) \sim \chi^2_{(n+4)} \quad (16)$$

$$\left\{\frac{1}{v_{ii}} \left(\frac{\varepsilon_i^2}{\sigma^2} + r\right)\right\} \sim \chi^2_{(r+1)} \quad (17)$$

$$p(\boldsymbol{\beta} | \sigma^2, \mathbf{V}, \rho, \lambda) \propto$$

$$\exp\left\{-\frac{1}{2\sigma^2} [\mathbf{B}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]^T \mathbf{V}^{-1} [\mathbf{B}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]\right\} \quad (18)$$

$$p(\rho, \lambda | \sigma^2, \mathbf{V}, \boldsymbol{\beta}) \propto |\mathbf{I} - \rho\mathbf{W}| |\mathbf{I} - \lambda\mathbf{W}| \times \exp\left\{-\frac{1}{2\sigma^2} \boldsymbol{\varepsilon}^T \mathbf{V}^{-1} \boldsymbol{\varepsilon}\right\} \quad (19)$$

dimana $\mathbf{A} = (\mathbf{I} - \rho\mathbf{W})$, $\mathbf{B} = (\mathbf{I} - \lambda\mathbf{W})$, dan $\boldsymbol{\varepsilon} = \mathbf{B}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

Estimasi parameter metode MCMC *Gibbs sampler* dilakukan dengan cara membangkitkan angka random yang mengikuti distribusi posterior bersyarat dari masing-masing parameter sebanyak jumlah iterasi yang diinginkan. Adapun algoritma *Metropolis within Gibbs* digunakan pada bentuk distribusi posterior yang tidak standar, yaitu untuk parameter ρ atau parameter λ (LeSage, 2000). Untuk menentukan layak atau tidaknya suatu

variabel prediktor dimasukkan ke dalam model, digunakan statistik uji Wald dengan hipotesis sebagai berikut:

$$H_0: \boldsymbol{\beta}_k = \mathbf{0}$$

$$H_1: \boldsymbol{\beta}_k \neq \mathbf{0}, k = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah:

$$W_k = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \quad (20)$$

dimana $\hat{\beta}_k$ merupakan parameter hasil estimasi, dan $SE(\hat{\beta}_k) = \sqrt{\widehat{Var}(\hat{\beta}_k)}$.

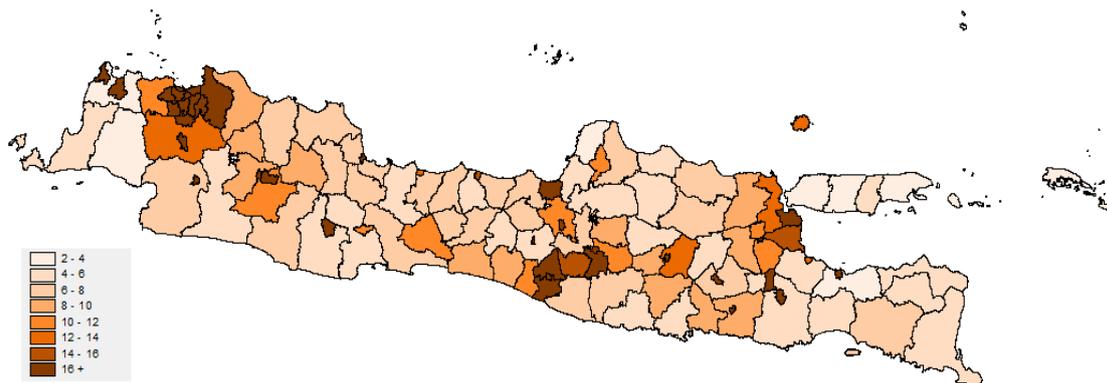
Keputusan untuk menolak H_0 diambil apabila nilai $|W_k| > Z_{\alpha/2}$ atau p-value lebih kecil dari nilai α .

HASIL DAN PEMBAHASAN

Penggunaan Internet di Pulau Jawa

Tingkat penggunaan internet dilihat dari nilai persentase penduduk usia lima tahun ke atas yang pernah mengakses internet dalam dalam tiga bulan terakhir. Penggunaan internet terbesar pada tahun 2010 dapat ditemukan di kota pelajar Yogyakarta yaitu sebanyak 36,20 persen, disusul oleh Jakarta Selatan, Kabupaten Sleman, dan kota industri Tangerang Selatan, masing-masing sebanyak 33,20 persen, 29,52 persen dan 29,29 persen. Rata-rata penggunaan internet per kabupaten/kota di Pulau Jawa sebesar 12,02 persen. Wilayah dengan nilai penggunaan internet di sekitar rata-rata adalah Kabupaten Mojokerto, Bandung, Madiun, dan Kota Pasuruan. Sementara wilayah dengan persentase pengguna internet terkecil ada di Kabupaten Sampang dan Bangkalan, masing-masing sebesar 2,23 persen dan 2,54 persen.

³ Informasi nilai rata-rata dan varians dari variabel random latent \mathbf{Z} ini dijelaskan pula dalam LeSage (2000) serta LeSage dan Pace (2009).



Gambar 2. Penggunaan Internet Kabupaten/Kota di Pulau Jawa Tahun 2010

Peta tematik penggunaan internet pada Gambar 2 memperlihatkan bahwa penggunaan internet kategori tinggi banyak terdapat pada wilayah perkotaan seperti DKI Jakarta dan sekitarnya, deretan Yogyakarta ke arah Surakarta, serta wilayah Bandung, Semarang, Surabaya dan sekitarnya. Tingginya penggunaan internet di kota-kota pusat pelajar, industri, dan bisnis tersebut kemudian diikuti oleh wilayah lain di sekitarnya. Wilayah yang bersinggungan langsung dengan kota-kota tersebut memiliki persentase pengguna internet yang sedikit lebih rendah, sementara wilayah berikutnya yang bersinggungan secara tidak langsung memiliki nilai yang lebih rendah lagi.

Untuk pemodelan regresi Tobit spasial, nilai persentase pengguna internet dari 34 kabupaten/kota dianggap tidak diketahui. Mereka adalah wilayah dengan tingkat penggunaan internet yang lebih besar dari 16 persen, yaitu sebagai batas minimal persentase pengguna internet yang ingin dicapai oleh suatu kabupaten/kota. Nilai-

nilai tingkat penggunaan internet yang tidak diketahui tersebut dianggap sama dengan 16 persen, sehingga diperoleh variabel persentase pengguna internet sebagai data tersensor. Hal ini sesuai dengan konsep pada persamaan (1) di atas.

Deskriptif data tingkat penggunaan internet sebagai variabel respon yang tersensor dapat dilihat pada Tabel 1 di bawah ini. Nilai maksimum dari persentase pengguna internet per kabupaten/kota sama dengan 16 persen, dengan rata-rata dan standar deviasi masing-masing sebesar 9,97 persen dan 4,63. Variabel yang memiliki variasi nilai terbesar adalah persentase penduduk perkotaan dengan standar deviasi sebesar 30,84 dan panjang range data dari 9,27 persen hingga 100 persen. Variabel rata-rata lama sekolah memiliki variasi terkecil dengan standar deviasi sebesar 1,52. Hal itu karena satuan dari variabel tersebut dalam ukuran tahun, sementara variabel lainnya dalam satuan persentase.

Tabel 1. Deskriptif Variabel Penelitian

| Variabel Penelitian | Deskripsi | Minimum | Maksimum | Rata-rata | Standar Deviasi |
|---------------------|---|---------|----------|-----------|-----------------|
| (1) | (2) | (3) | (4) | (5) | (6) |
| Y | Persentase pengguna internet | 2,23 | 16,00 | 9,968 | 4,626 |
| X_1 | Persentase penduduk perkotaan | 9,27 | 100,00 | 57,991 | 30,837 |
| X_2 | Persentase penduduk usia muda | 12,44 | 25,00 | 17,395 | 2,567 |
| X_3 | Persentase penduduk lulusan SMA ke atas | 5,30 | 50,26 | 21,349 | 11,105 |
| X_4 | Rata-rata lama sekolah | 4,21 | 11,55 | 8,052 | 1,518 |
| X_5 | Persentase rumah tangga yang memiliki komputer | 2,20 | 40,22 | 11,433 | 9,127 |
| X_6 | Persentase rumah tangga memiliki telepon seluler | 44,87 | 94,89 | 71,923 | 12,050 |
| X_7 | Persentase desa/kelurahan mendapat sinyal telepon | 50,00 | 100,00 | 88,892 | 10,026 |

Pemodelan regresi linier berganda dengan metode *ordinary least squared* (OLS) dilakukan untuk menjelaskan hubungan antara variabel prediktor terhadap tingkat penggunaan internet di Pulau Jawa. Pada derajat kepercayaan 95 persen, hasil uji parameter hanya menghasilkan dua dari tujuh variabel prediktor yang berpengaruh terhadap variabel respon. Nilai *variance inflation factor* (VIF) yang sangat tinggi pada variabel X_3 dan X_4 mengindikasikan

adanya kondisi multikolinieritas antar variabel prediktor. Meskipun menghasilkan model yang fit dengan nilai koefisien determinasi yang sangat tinggi, model regresi linier berganda yang diperoleh kurang tepat untuk digunakan sebagai dasar analisis. Hal ini karena tidak terpenuhinya asumsi non multikolinieritas dan banyaknya informasi dari variabel prediktor yang terbuang.

Tabel 2. Model Regresi Linier Berganda dan Nilai *Variance Inflation Factor* (VIF)

| Parameter | Koefisien | Uji Parameter | | Statistik VIF |
|--|-----------|-----------------|----------------|---------------|
| | | Statistik Uji t | <i>p-Value</i> | |
| (1) | (2) | (3) | (4) | (5) |
| $\hat{\beta}_0$ | -13,729 | -4,158 | 0,000 | – |
| $\hat{\beta}_1$ | 0,022 | 1,960 | 0,053 | 5,705 |
| $\hat{\beta}_2$ | -0,022 | -0,298 | 0,766 | 1,701 |
| $\hat{\beta}_3$ | 0,107 | 1,585 | 0,116 | 27,368 |
| $\hat{\beta}_4$ | 0,797 | 1,936 | 0,055 | 18,952 |
| $\hat{\beta}_5$ | -0,026 | -0,531 | 0,597 | 9,827 |
| $\hat{\beta}_6$ | 0,083 | 2,671 | 0,009 | 6,872 |
| $\hat{\beta}_7$ | 0,094 | 4,582 | 0,000 | 2,072 |
| <i>Analysis of Variance</i> (uji F) | – | 132,47 | 0,000 | – |
| Koefisien Determinasi (R^2) | 0,894 | – | – | – |

Menurut Gujarati (2004), kondisi multikolinieritas dapat ditangani dengan cara melakukan pemilihan variabel, baik dengan menambah variabel prediktor baru maupun dengan mengurangi yang sudah ada. Cara lain yang dapat dilakukan adalah dengan menggunakan model lain yang lebih sesuai untuk menjelaskan hubungan antara variabel prediktor terhadap respon. Model regresi Tobit spasial lebih tepat digunakan untuk menjelaskan faktor-faktor yang mempengaruhi tinggi serta keragaman rendahnya penggunaan internet antar kabupaten/kota di Pulau Jawa yang dipengaruhi secara spasial oleh daerah-daerah di sekitarnya.

Pemodelan Regresi Tobit Spasial

Berdasarkan persamaan (9) di atas, model regresi Tobit spasial untuk penggunaan internet kabupaten/kota di Pulau Jawa adalah sebagai berikut:

$$y_i = \begin{cases} \rho w_i^T y + \lambda w_i^T y + (\rho W^T \lambda W)^T y \\ + x_i^T \beta - \lambda (w_i^T X \beta) + \varepsilon_i & \text{jika } y_i < 16 \\ 16 & \text{jika } y_i \geq 16 \end{cases} \quad (21)$$

Pemodelan Tobit spasial dilakukan dengan terlebih dahulu menguji efek korelasi spasial dan heteroskedastisitas, untuk menentukan bentuk korelasi spasial yang ada serta kondisi heteroskedastisitas yang akan diperhitungkan dalam algoritma MCMC. Metode *backward elimination* dilakukan untuk menghasilkan model regresi Tobit spasial terbaik dengan cara menggugurkan

satu persatu variabel yang tidak berpengaruh secara nyata (Draper & Smith, 1998).

Hasil uji efek korelasi spasial dan heteroskedastisitas memperlihatkan bahwa bentuk model yang akan dibentuk adalah regresi Tobit spasial lag dalam kondisi homoskedastisitas. Menggunakan hipotesis $H_0: \rho = 0$, nilai statistik uji LM-lag memperlihatkan kesimpulan untuk menolak H_0 pada derajat kepercayaan 95 persen. Sementara nilai statistik uji LM-err memperlihatkan bahwa $H_0: \lambda = 0$ gagal ditolak pada nilai $\alpha = 0,05$. Adapun hasil uji heteroskedastisitas yang menggunakan statistik uji Breusch Pagan menyimpulkan untuk tidak menolak hipotesis H_0 : homoskedastisitas pada derajat kepercayaan 95 persen. Kondisi tersebut berlaku pada setiap tahapan pembentukan model menggunakan metode *backward elimination*.

Estimasi parameter dilakukan dengan terlebih dahulu membangkitkan sejumlah angka random $z \sim N(\mu, \Sigma)$, di mana nilai z_i ketika $z_i \geq 16$ digunakan untuk mengganti data yang tersensor. Algoritma MCMC digunakan pada data yang telah lengkap untuk estimasi parameter model regresi Tobit spasial lag, yaitu *Gibbs sampler* untuk estimasi $\hat{\sigma}^2$ dan $\hat{\beta}$ serta *Metropolis within Gibbs* untuk estimasi $\hat{\rho}$. Nilai $V = I$ digunakan karena kondisi homoskedastisitas yang terpenuhi. Eliminasi variabel diputuskan berdasarkan p-value dari statistik uji Wald yang lebih besar daripada nilai $\alpha = 0,05$.

Tabel 3. Pengujian Efek Korelasi Spasial, Heteroskedastisitas, dan Hasil Estimasi Parameter untuk Pemodelan Regresi Tobit Spasial menggunakan Metode *Backward Elimination*

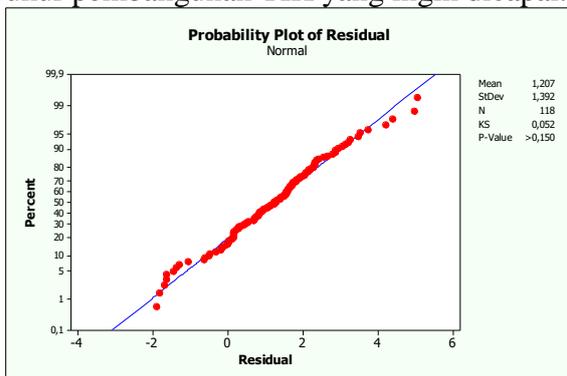
| Statistik Uji/ Parameter | Tahap I (7 Variabel) | | Tahap II (6 Variabel) | | Tahap III (5 Variabel) | |
|-----------------------------|----------------------|---------------|-----------------------|---------------|------------------------|---------------|
| | Nilai | p-value | Nilai | p-value | Nilai | p-value |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| LM-lag | 12,1942 | 0,0000 | 12,2410 | 0,0000 | 11,8122 | 0,0000 |
| LM-err | 1,5595 | 0,2117 | 1,5765 | 0,2093 | 1,4396 | 0,2302 |
| Breusch Pagan | 9,7859 | 0,2010 | 9,4602 | 0,1493 | 7,9076 | 0,1614 |
| $\hat{\rho}$ | -0,1662 | 0,0000 | -0,1693 | 0,0000 | -0,1700 | 0,0000 |
| $\hat{\beta}_0$ | -13,9281 | 0,0000 | -13,8073 | 0,0000 | -13,7268 | 0,0000 |
| $\hat{\beta}_1$ | 0,0115 | 0,0000 | 0,0122 | 0,0000 | 0,0117 | 0,0000 |
| $\hat{\beta}_2$ | 0,0107 | 0,2779 | | | | |
| $\hat{\beta}_3$ | 0,1549 | 0,0000 | 0,1529 | 0,0000 | 0,1435 | 0,0000 |

| | | | | | | |
|------------------|---------|---------------|---------|---------------|--------|--------|
| $\hat{\beta}_4$ | 0,7565 | 0,0000 | 0,7574 | 0,0000 | 0,7548 | 0,0000 |
| $\hat{\beta}_5$ | -0,0143 | 0,2102 | -0,0139 | 0,2067 | | |
| $\hat{\beta}_6$ | 0,0918 | 0,0000 | 0,0929 | 0,0000 | 0,0931 | 0,0000 |
| $\hat{\beta}_7$ | 0,0831 | 0,0000 | 0,0828 | 0,0000 | 0,0827 | 0,0000 |
| $\hat{\sigma}^2$ | 1,9561 | – | 1,9671 | – | 1,9580 | – |
| R^2 | 0,8420 | – | 0,8403 | – | 0,8394 | – |

Variabel prediktor yang tersisih dari model adalah X_2 (persentase penduduk usia muda) dan X_5 (persentase rumah tangga yang memiliki komputer). Pada tahap akhir *backward elimination* diperoleh lima variabel prediktor yang secara nyata mempengaruhi keragaman penggunaan internet di Pulau Jawa. Berdasarkan persamaan (21) dan hasil estimasi parameter yang disajikan pada tabel 3 di atas, model regresi Tobit spasial lag yang terbentuk adalah:

$$\hat{y}_i = \begin{cases} -13,727 - 0,17 \sum_{j=1, j \neq i}^n w_{ij} y_j + 0,012 x_{1i} \\ + 0,144 x_{3i} + 0,755 x_{4i} + 0,093 x_{6i} + 0,083 x_{7i} ; y_i < 16 \\ 16 ; y_i \geq 16 \end{cases} \quad (22)$$

Model ini digunakan untuk menjelaskan faktor-faktor dan daerah di sekitarnya yang mempengaruhi tingkat penggunaan internet di suatu kabupaten/kota, ketika nilainya lebih kecil dari 16 persen. Adapun untuk kabupaten/kota dengan tingkat penggunaan internet yang tinggi dianggap sebagai tolok ukur pembangunan TIK yang ingin dicapai.



Gambar 3. Grafik Plot Probabilitas Normal dari Residual Model Regresi Tobit Spasial Lag

Grafik plot probabilitas normal dari residual digunakan untuk menguji asumsi normalitas dari error model. Plot residual terlihat berada di sekitar garis probabilitas

normal, sehingga dapat disimpulkan bahwa asumsi normalitas dari error model terpenuhi secara nyata. Asumsi berikutnya terkait masalah varians error yang homogen dianggap telah terpenuhi, mengingat hasil uji Breusch Pagan sebelumnya yang memperlihatkan tidak terpenuhinya kondisi heteroskedastisitas serta proses simulasi MCMC yang didasari oleh kondisi homoskedastisitas. Adapun asumsi tidak adanya autokorelasi dalam error juga telah terpenuhi, karena model yang terbentuk bukanlah model regresi Tobit spasial error.

Interpretasi Model

Persamaan (22) di atas menjelaskan bahwa untuk kabupaten/kota dengan persentase pengguna internet yang kurang dari 16 persen, tingkat penggunaan internet di wilayah tersebut dipengaruhi oleh daerah lain di sekitarnya serta variabel-variabel sebagai berikut:

- Persentase penduduk yang tinggal di perkotaan (X_1). Apabila variabel lain dianggap konstan, maka setiap kenaikan persentase penduduk yang tinggal di perkotaan sebesar satu persen akan mengakibatkan kenaikan persentase pengguna internet sebesar 0,012 persen.
- Persentase penduduk lulusan SMA ke atas (X_3). Apabila variabel lain dianggap konstan, maka setiap kenaikan persentase penduduk lulusan SMA ke atas sebesar satu persen akan mengakibatkan kenaikan persentase pengguna internet sebesar 0,144 persen.
- Rata-rata lama sekolah (X_4). Apabila variabel lain dianggap konstan, maka setiap kenaikan rata-rata lama sekolah selama satu tahun akan mengakibatkan kenaikan penggunaan internet suatu kabupaten/kota sebesar 0,755 persen.

- Persentase rumah tangga yang memiliki telepon genggam (X_6). Apabila variabel lain dianggap konstan, maka setiap kenaikan persentase rumah tangga yang memiliki telepon genggam sebesar satu persen akan mengakibatkan kenaikan persentase pengguna internet sebesar 0,093 persen.
- Persentase desa/kelurahan yang mendapat sinyal telepon seluler (X_7).

Apabila variabel lain dianggap konstan, maka setiap kenaikan persentase desa/kelurahan yang mendapat sinyal telepon seluler sebesar satu persen akan menyebabkan kenaikan penggunaan internet sebesar 0,083 persen.

Nilai koefisien determinasi memperlihatkan bahwa sebesar 83,94 persen variasi penggunaan internet di Pulau Jawa dijelaskan oleh lima variabel prediktor dalam model, sisanya oleh variabel lain. Pengaruh spasial lag dari daerah lain yang saling bersinggungan wilayah dapat terlihat pada model regresi Tobit spasial lag masing-masing kabupaten/kota, yaitu model \hat{y}_i ketika $y_i < 16$. Adapun untuk kabupaten/kota dengan kategori penggunaan internet yang tinggi, maka nilai persentase penduduk umur 5 tahun ke atas yang pernah mengakses internet dalam tiga bulan terakhir dianggap sama dengan 16 persen atau $\hat{y}_i = 16$ ketika $y_i \geq 16$.

Persentase penduduk yang tinggal di perkotaan mengindikasikan tingkat kemajuan dan kelengkapan fasilitas umum di daerah tersebut. Adapun presentase penduduk lulusan SMA ke atas dan rata-rata lama sekolah mencerminkan kualitas sumber daya manusia di daerah tersebut. Dengan demikian, meningkatkan persentase pengguna internet dapat dilakukan melalui upaya peningkatan kualitas sumber daya manusia dari aspek pendidikan. Selain itu, pembangunan kelengkapan fasilitas umum di daerah pedesaan juga dapat mendorong tingkat penggunaan internet di kabupaten/kota.

Variabel karakteristik perangkat dan jaringan di daerah mengindikasikan pentingnya perkembangan teknologi telepon seluler bagi pertumbuhan internet. Berbagai

kemudahan akses internet yang disediakan melalui perangkat telepon genggam dan keluasan jaringan telepon seluler telah secara nyata mendorong peningkatan tingkat penggunaan internet. Di sisi lain, penggunaan internet ternyata tidak secara nyata diakses melalui komputer atau didominasi oleh penduduk usia muda. Internet dapat diakses oleh siapapun dan melalui media apapun, terutama telepon seluler.

Selain dipengaruhi oleh kelima variabel di atas, tingkat penggunaan internet kabupaten/kota di Pulau Jawa juga dipengaruhi oleh daerah lain yang bersinggungan wilayah. Sebagai contoh, tingkat penggunaan internet di Kabupaten Kepulauan Seribu dapat dijelaskan melalui model regresi Tobit spasial lag berikut ini:

$$\hat{y}_{3101} = -0.085(y_{3175} + y_{3603}) + x_{3101}^T \beta \quad (23)$$

dimana

$$x_{3101}^T = [1 \quad x_{1_{3101}} \quad x_{3_{3101}} \quad x_{4_{3101}} \quad x_{6_{3101}} \quad x_{7_{3101}}]$$

adalah vektor variabel prediktor dari Kepulauan Seribu dan

$$\beta = [-13,727 \quad 0,012 \quad 0,144 \quad 0,755 \quad 0,093 \quad 0,083]^T$$

adalah vektor parameter. Tingkat penggunaan internet di Kepulauan Seribu dipengaruhi pula oleh penggunaan internet di Kota Jakarta Utara dan Kabupaten Tangerang. Jika variabel lain dianggap konstan, maka penggunaan internet di Kepulauan Seribu adalah sebesar -0,085 kali dari gabungan penggunaan internet di Tangerang (y_{3603}) dan Kota Jakarta Utara (y_{3175}). Secara detail, masing-masing 84 model regresi Tobit spasial lag ketika nilai persentase pengguna internet kurang dari 16 persen dapat dilihat pada lampiran.

KESIMPULAN DAN SARAN

Model regresi Tobit spasial merupakan suatu model regresi spasial yang diterapkan pada data tersensor, dengan bentuk model umum dari regresi Tobit spasial adalah:

$$y_i = \begin{cases} \rho w_i^T y + \lambda w_i^T y + (\rho W^T \lambda W)_i^T y \\ + x_i^T \beta - \lambda (w_i^T X \beta) + \varepsilon_i & \text{jika } y_i < \tau \\ \tau & \text{jika } y_i \geq \tau \end{cases} \quad (24)$$

Metode estimasi parameter yang dapat digunakan adalah *Markov Chain Monte Carlo* (MCMC) yang dilengkapi dengan algoritma *Gibbs sampler* dan *Metropolis within Gibbs*. Metode ini lebih mengedepankan teknik simulasi komputasi untuk membangkitkan sejumlah besar variabel random menggunakan pendekatan inferensia Bayesian.

Menggunakan penggunaan internet di Pulau Jawa sebagai studi kasus, diketahui bahwa model regresi Tobit Spasial lag menghasilkan informasi yang lebih kaya daripada model regresi linier berganda. Faktor-faktor yang mempengaruhi penggunaan internet kabupaten/kota di Pulau Jawa adalah persentase penduduk yang tinggal di daerah perkotaan, persentase penduduk lulusan SMA ke atas, rata-rata lama sekolah, persentase rumah tangga yang memiliki telepon genggam, dan persentase desa/kelurahan yang mendapatkan sinyal telepon seluler. Selain dipengaruhi oleh kelima variabel tersebut, tingkat penggunaan internet kabupaten/kota di Pulau Jawa juga dipengaruhi oleh daerah lainnya yang bersinggungan wilayah.

Berdasarkan hasil penelitian yang telah diperoleh, pengembangan lebih lanjut dapat dilakukan dengan menggunakan *highest posterior density* (HPD) dan Bayes Faktor sebagai metode pengujian parameter dan model. Penelitian ini masih menggunakan matriks penimbang *queen contiguity*, sehingga pada penelitian selanjutnya dapat dikembangkan menggunakan matriks penimbang lain misalnya jarak. Lebih lanjut, metode MCMC *Gibbs sampler* untuk pemodelan regresi Tobit spasial ini dapat digunakan untuk data dan kasus lain yang lebih aplikatif.

DAFTAR PUSTAKA

Andonova, V., & Serrano, L. D. 2007. *Political Institutions and the*

Development of Telecommunications. Bonn: IZA Discussion Paper.

- Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.
- Anselin, L. 1999. *Spatial Econometrics*. Dallas: University of Texas.
- BPS. 2011. *Sensus Penduduk 2010*. Hämtat från Sensus Penduduk 2010: <http://sp2010.bps.go.id/index.php/site/index> (diakses 4 November 2011)
- BPS. 2011. *Statistik Komunikasi dan Teknologi Informasi Tahun 2010*. Jakarta: Badan Pusat Statistik.
- Breusch, T., & Pagan, A. 1979. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, Vol. 47, No. 5, 1287-1294.
- Casella, G. dan Berger, R. 2002. *Statistical Inference*. Duxbury, Thomson Learning.
- Casella, G. dan George, E. I. 1992. Explaining the Gibbs Sampler. *The American Statistician*, Vol. 46, No. 3, 167-335.
- Chib, S. dan Greenberg, E. 1996. Markov Chain Monte Carlo Simulation Methods in Econometrics. *Econometrics Theory*, Vol. 12, 409-431.
- DeMaris, A. 2004. *Regression with Social Data: Modelling Continuous and Limited Response Variable*. New Jersey: John Wiley and Sons, Inc.
- Draper, N. R. dan Smith, H. 1998. *Applied Regression Analysis*. New York: John Wiley and Sons, Inc.
- Fischer, M. M. dan Getis, A. 2010. *Handbook of Applied Spatial Analysis: Software Tools, Methods, and Application*. New York: Springer.
- Greene, W. H. 2008. *Econometric Analysis, Sixth Edition*. New York: Pearson - Prentice Hall.
- Hastings, W. 1970. Monte Carlo Sampling Methods using Markov Chains and Their Applications. *Biometrika*, Vol. 57, No. 1, 97-109.
- Howard, P. N. dan Mazaheri, N. 2009. Telecommunications Reform, Internet Use, and Mobile Phone Adoption in

- Developing World. *World Development*, Vol. 37, No. 7, 1159-1169.
- ITU. 2010. *Measuring the Information Society*. Geneva: International Telecommunications Union.
- Kaliba, A. R. 2002. Participatory Evaluation of Community Based Water and Sanitation Programmes: The Case of Central Tanzania. *Dissertation*. Mahattan: Kansas State University.
- Kominfo. 2010. *Buku Putih Komunikasi dan Informatika Indonesia*. Jakarta: Pusat Data Kementerian Komunikasi dan Informatika.
- Lacombe, D. J. (2008, Juli 24). *An Introduction to Bayesian Inference in Spatial Econometrics*. Hämtat från <http://ssrn.com/abstract=1244261>. (diakses 13 November 2011).
- Langyintuo, A. S. dan Mekuria, M. 2008. Assessing the Influence of Neighborhood Effects on the Adoption of Improved Agricultural Technologies in Developing Agriculture. *AfJARE*, Vol. 2, No. 2, 151-169.
- Lee, M. J. 2010. *Micro-Econometrics: Methods of Moments and Limited Dependent Variables, Second Edition*. New York: Springer.
- LeSage, J. P. 1999. *The Theory and Practice of Spatial Econometrics*. Ohio: University of Toledo.
- LeSage, J. P. 2000. Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models. *Geographical Analysis*, Vol. 32, No. 1, 19-35.
- LeSage, J. dan Pace, R. K. 2009. *Introduction to Spatial Econometrics*. New York: CRC Press.
- Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. California: Sage Publications, Inc.
- Marsh, T. L., Mittelhammer, R. C., & Huffaker, R. G. 2000. Probit with Spatial Correlation by Field Plot: Potato Leafroll Virus Net Necrosis in Potatoes. *Journal of Agricultural, Biological, and Environmental Statistics*, Volume 5, Number 1, Pages 22-36.
- Michailidis, A., Partalidou, M., Nastis, S. A., Klavdianou, A. P. dan Charatsari, C. 2011. Who Goes Online? Evidence of Internet Use Patterns from Rural Greece. *Telecommunications Policy*, Vol. 35, 333-343.
- Rao, J. G. dan Pattnaik, S. 2006. Technology for Rural Development Role of Telecommunication Media in India. *Indian Media Studies Journal*, Vol. 1, No. 1, 85-92.
- Socialbaker. (2011, Agustus 17) . *Facebook Statistics by Country*. Hämtat från [www.socialbaker.com: http://www.socialbakers.com/facebook-statistics/?interval=last-3-months#chart-intervals](http://www.socialbakers.com/facebook-statistics/?interval=last-3-months#chart-intervals).
- Tobin, J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica*, Vol. 26, No. 1, 24-36.

LAMPIRAN

Model regresi Tobit spasial lag untuk 84 kabupaten/kota dengan nilai $y_i < 16$:

1. Kepulauan Seribu : $\hat{y}_{3101} = -0,085(y_{3175} + y_{3603}) + x_{3101}^T \beta$
2. Bogor : $\hat{y}_{3201} = -0,015(y_{3202} + y_{3203} + y_{3214} + y_{3215} + y_{3216} + y_{3271} + y_{3275} + y_{3276} + y_{3602} + y_{3603} + y_{3674}) + x_{3201}^T \beta$
3. Sukabumi : $\hat{y}_{3202} = -0,043(y_{3201} + y_{3203} + y_{3272} + y_{3602}) + x_{3202}^T \beta$
4. Cianjur : $\hat{y}_{3203} = -0,028(y_{3201} + y_{3202} + y_{3214} + y_{3217} + y_{3274} + y_{3605}) + x_{3203}^T \beta$
5. Bandung : $\hat{y}_{3204} = -0,024(y_{3203} + y_{3205} + y_{3211} + y_{3213} + y_{3217} + y_{3273} + y_{3277}) + x_{3204}^T \beta$
6. Garut : $\hat{y}_{3205} = -0,043(y_{3203} + y_{3204} + y_{3206} + y_{3211}) + x_{3205}^T \beta$

7. Tasikmala ya : $\hat{y}_{3206} = -0,034(y_{3205} + y_{3207} + y_{3210} + y_{3211} + y_{3278}) + x_{3206}^T \beta$
8. Ciamis : $\hat{y}_{3207} = -0,028(y_{3206} + y_{3208} + y_{3210} + y_{3278} + y_{3279} + y_{3301}) + x_{3207}^T \beta$
9. Kuningan : $\hat{y}_{3208} = -0,034(y_{3207} + y_{3209} + y_{3210} + y_{3301} + y_{3329}) + x_{3208}^T \beta$
10. Cirebon : $\hat{y}_{3209} = -0,034(y_{3208} + y_{3210} + y_{3212} + y_{3274} + y_{3329}) + x_{3209}^T \beta$
11. Majalengka : $\hat{y}_{3210} = -0,028(y_{3206} + y_{3207} + y_{3208} + y_{3209} + y_{3211} + y_{3212}) + x_{3210}^T \beta$
12. Sumedang : $\hat{y}_{3211} = -0,028(y_{3204} + y_{3205} + y_{3206} + y_{3210} + y_{3212} + y_{3213}) + x_{3211}^T \beta$
13. Indramayu : $\hat{y}_{3212} = -0,043(y_{3209} + y_{3210} + y_{3211} + y_{3213}) + x_{3212}^T \beta$
14. Subang : $\hat{y}_{3213} = -0,028(y_{3204} + y_{3211} + y_{3212} + y_{3214} + y_{3215} + y_{3217}) + x_{3213}^T \beta$
15. Purwakarta : $\hat{y}_{3214} = -0,034(y_{3201} + y_{3203} + y_{3213} + y_{3215} + y_{3217}) + x_{3214}^T \beta$
16. Karawang : $\hat{y}_{3215} = -0,043(y_{3201} + y_{3213} + y_{3214} + y_{3216}) + x_{3215}^T \beta$
17. Bandung Barat : $\hat{y}_{3217} = -0,028(y_{3203} + y_{3204} + y_{3213} + y_{3214} + y_{3273} + y_{3277}) + x_{3217}^T \beta$
18. Kota Banjar : $\hat{y}_{3279} = -0,085(y_{3207} + y_{3301}) + x_{3279}^T \beta$
19. Cilacap : $\hat{y}_{3301} = -0,028(y_{3207} + y_{3208} + y_{3279} + y_{3302} + y_{3305} + y_{3329}) + x_{3301}^T \beta$
20. Banyumas : $\hat{y}_{3302} = -0,024(y_{3301} + y_{3303} + y_{3304} + y_{3305} + y_{3327} + y_{3328} + y_{3329}) + x_{3302}^T \beta$
21. Purbalingga : $\hat{y}_{3303} = -0,043(y_{3302} + y_{3304} + y_{3326} + y_{3327}) + x_{3303}^T \beta$
22. Banjarnegara : $\hat{y}_{3304} = -0,028(y_{3302} + y_{3303} + y_{3305} + y_{3307} + y_{3325} + y_{3326}) + x_{3304}^T \beta$
23. Kebumen : $\hat{y}_{3305} = -0,034(y_{3301} + y_{3302} + y_{3304} + y_{3306} + y_{3307}) + x_{3305}^T \beta$
24. Purworejo : $\hat{y}_{3306} = -0,043(y_{3305} + y_{3307} + y_{3308} + y_{3401}) + x_{3306}^T \beta$
25. Wonosobo : $\hat{y}_{3307} = -0,024(y_{3304} + y_{3305} + y_{3306} + y_{3308} + y_{3323} + y_{3324} + y_{3325}) + x_{3307}^T \beta$
26. Magelang : $\hat{y}_{3308} = -0,019(y_{3306} + y_{3307} + y_{3309} + y_{3310} + y_{3322} + y_{3323} + y_{3371} + y_{3401} + y_{3404}) + x_{3308}^T \beta$
27. Boyolali : $\hat{y}_{3309} = -0,019(y_{3308} + y_{3310} + y_{3311} + y_{3313} + y_{3314} + y_{3315} + y_{3322} + y_{3372} + y_{3404}) + x_{3309}^T \beta$
28. Klaten : $\hat{y}_{3310} = -0,034(y_{3308} + y_{3309} + y_{3311} + y_{3403} + y_{3404}) + x_{3310}^T \beta$
29. Wonogiri : $\hat{y}_{3312} = -0,028(y_{3311} + y_{3313} + y_{3403} + y_{3501} + y_{3502} + y_{3520}) + x_{3312}^T \beta$
30. Karanganyar : $\hat{y}_{3313} = -0,024(y_{3309} + y_{3311} + y_{3312} + y_{3314} + y_{3372} + y_{3520} + y_{3521}) + x_{3313}^T \beta$
31. Sragen : $\hat{y}_{3314} = -0,043(y_{3309} + y_{3313} + y_{3315} + y_{3521}) + x_{3314}^T \beta$
32. Grobogan : $\hat{y}_{3315} = -0,021(y_{3309} + y_{3314} + y_{3316} + y_{3318} + y_{3319} + y_{3321} + y_{3322} + y_{3521}) + x_{3315}^T \beta$

- 33 Blora : $\hat{y}_{3316} = -0,028(y_{3315} + y_{3317} + y_{3318} + y_{3521} + y_{3522} + y_{3523}) + \mathbf{x}_{3316}^T \boldsymbol{\beta}$
 .
- 34 Rembang : $\hat{y}_{3317} = -0,057(y_{3316} + y_{3318} + y_{3523}) + \mathbf{x}_{3317}^T \boldsymbol{\beta}$
 .
- 35 Pati : $\hat{y}_{3318} = -0,034(y_{3315} + y_{3316} + y_{3317} + y_{3319} + y_{3320}) + \mathbf{x}_{3318}^T \boldsymbol{\beta}$
 .
- 36 Kudus : $\hat{y}_{3319} = -0,043(y_{3315} + y_{3318} + y_{3320} + y_{3321}) + \mathbf{x}_{3319}^T \boldsymbol{\beta}$
 .
- 37 Jepara : $\hat{y}_{3320} = -0,057(y_{3318} + y_{3319} + y_{3321}) + \mathbf{x}_{3320}^T \boldsymbol{\beta}$
 .
- 38 Demak : $\hat{y}_{3321} = -0,034(y_{3315} + y_{3319} + y_{3320} + y_{3322} + y_{3374}) + \mathbf{x}_{3321}^T \boldsymbol{\beta}$
 .
- 39 Semarang : $\hat{y}_{3322} = -0,021(y_{3308} + y_{3309} + y_{3315} + y_{3321} + y_{3323} + y_{3324} + y_{3373} + y_{3374}) + \mathbf{x}_{3322}^T \boldsymbol{\beta}$
 .
- 40 Temanggung : $\hat{y}_{3323} = -0,043(y_{3307} + y_{3308} + y_{3322} + y_{3324}) + \mathbf{x}_{3323}^T \boldsymbol{\beta}$
 .
 ng
- 41 Kendal : $\hat{y}_{3324} = -0,034(y_{3307} + y_{3322} + y_{3323} + y_{3325} + y_{3374}) + \mathbf{x}_{3324}^T \boldsymbol{\beta}$
 .
- 42 Batang : $\hat{y}_{3325} = -0,034(y_{3304} + y_{3307} + y_{3324} + y_{3326} + y_{3375}) + \mathbf{x}_{3325}^T \boldsymbol{\beta}$
 .
- 43 Pekalonga : $\hat{y}_{3326} = -0,034(y_{3303} + y_{3304} + y_{3325} + y_{3327} + y_{3375}) + \mathbf{x}_{3326}^T \boldsymbol{\beta}$
 .
 n
- 44 Pemasang : $\hat{y}_{3327} = -0,043(y_{3302} + y_{3303} + y_{3326} + y_{3328}) + \mathbf{x}_{3327}^T \boldsymbol{\beta}$
 .
- 45 Tegal : $\hat{y}_{3328} = -0,043(y_{3302} + y_{3327} + y_{3329} + y_{3376}) + \mathbf{x}_{3328}^T \boldsymbol{\beta}$
 .
- 46 Brebes : $\hat{y}_{3329} = -0,028(y_{3208} + y_{3209} + y_{3301} + y_{3302} + y_{3328} + y_{3376}) + \mathbf{x}_{3329}^T \boldsymbol{\beta}$
 .
- 47 Kota : $\hat{y}_{3375} = -0,085(y_{3325} + y_{3326}) + \mathbf{x}_{3375}^T \boldsymbol{\beta}$
 .
 Pekalonga
 n
- 48 Kota : $\hat{y}_{3376} = -0,085(y_{3328} + y_{3329}) + \mathbf{x}_{3376}^T \boldsymbol{\beta}$
 .
 Tegal
- 49 Kulon : $\hat{y}_{3401} = -0,043(y_{3306} + y_{3308} + y_{3402} + y_{3404}) + \mathbf{x}_{3401}^T \boldsymbol{\beta}$
 .
 Progo
- 50 Gunung : $\hat{y}_{3403} = -0,034(y_{3310} + y_{3311} + y_{3312} + y_{3402} + y_{3404}) + \mathbf{x}_{3403}^T \boldsymbol{\beta}$
 .
 Kidul
- 51 Pacitan : $\hat{y}_{3501} = -0,057(y_{3312} + y_{3502} + y_{3503}) + \mathbf{x}_{3501}^T \boldsymbol{\beta}$
 .
- 52 Ponorogo : $\hat{y}_{3502} = -0,024(y_{3312} + y_{3501} + y_{3503} + y_{3504} + y_{3318} + y_{3519} + y_{3520}) + \mathbf{x}_{3502}^T \boldsymbol{\beta}$
 .
- 53 Trenggalek : $\hat{y}_{3503} = -0,057(y_{3501} + y_{3502} + y_{3504}) + \mathbf{x}_{3503}^T \boldsymbol{\beta}$
 .
 k
- 54 Tulungagung : $\hat{y}_{3504} = -0,034(y_{3502} + y_{3503} + y_{3505} + y_{3506} + y_{3518}) + \mathbf{x}_{3504}^T \boldsymbol{\beta}$
 .
 ng
- 55 Blitar : $\hat{y}_{3505} = -0,043(y_{3504} + y_{3506} + y_{3507} + y_{3572}) + \mathbf{x}_{3505}^T \boldsymbol{\beta}$
 .
- 56 Kediri : $\hat{y}_{3506} = -0,028(y_{3504} + y_{3505} + y_{3507} + y_{3517} + y_{3518} + y_{3571}) + \mathbf{x}_{3506}^T \boldsymbol{\beta}$
 .
- 57 Malang : $\hat{y}_{3507} = -0,019(y_{3505} + y_{3506} + y_{3508} + y_{3513} + y_{3514} + y_{3516} + y_{3517} + y_{3573} + y_{3579}) + \mathbf{x}_{3507}^T \boldsymbol{\beta}$
 .

- 58 Lumajang : $\hat{y}_{3508} = -0,057(y_{3507} + y_{3509} + y_{3513}) + x_{3508}^T \beta$
 .
- 59 Jember : $\hat{y}_{3509} = -0,043(y_{3508} + y_{3510} + y_{3511} + y_{3513}) + x_{3509}^T \beta$
 .
- 60 Banyuwani : $\hat{y}_{3510} = -0,057(y_{3509} + y_{3511} + y_{3512}) + x_{3510}^T \beta$
 .
- 61 Bondowoso : $\hat{y}_{3511} = -0,043(y_{3509} + y_{3510} + y_{3512} + y_{3513}) + x_{3511}^T \beta$
 .
- 62 Situbondo : $\hat{y}_{3512} = -0,057(y_{3510} + y_{3511} + y_{3513}) + x_{3512}^T \beta$
 .
- 63 Probolinggo : $\hat{y}_{3513} = -0,024(y_{3507} + y_{3508} + y_{3509} + y_{3511} + y_{3512} + y_{3514} + y_{3574}) + x_{3513}^T \beta$
 .
- 64 Pasuruan : $\hat{y}_{3514} = -0,028(y_{3507} + y_{3513} + y_{3515} + y_{3516} + y_{3575} + y_{3579}) + x_{3514}^T \beta$
 .
- 65 Sidoarjo : $\hat{y}_{3515} = -0,034(y_{3514} + y_{3516} + y_{3525} + y_{3576} + y_{3578}) + x_{3515}^T \beta$
 .
- 66 Mojokerto : $\hat{y}_{3516} = -0,021(y_{3507} + y_{3514} + y_{3515} + y_{3517} + y_{3524} + y_{3525} + y_{3576} + y_{3579}) + x_{3516}^T \beta$
 .
- 67 Jombang : $\hat{y}_{3517} = -0,028(y_{3506} + y_{3507} + y_{3516} + y_{3518} + y_{3522} + y_{3524}) + x_{3517}^T \beta$
 .
- 68 Nganjuk : $\hat{y}_{3518} = -0,028(y_{3502} + y_{3504} + y_{3506} + y_{3517} + y_{3519} + y_{3522}) + x_{3518}^T \beta$
 .
- 69 Madiun : $\hat{y}_{3519} = -0,028(y_{3502} + y_{3518} + y_{3520} + y_{3521} + y_{3522} + y_{3577}) + x_{3519}^T \beta$
 .
- 70 Magetan : $\hat{y}_{3520} = -0,028(y_{3212} + y_{3213} + y_{3502} + y_{3519} + y_{3521} + y_{3577}) + x_{3520}^T \beta$
 .
- 71 Ngawi : $\hat{y}_{3521} = -0,024(y_{3213} + y_{3214} + y_{3215} + y_{3216} + y_{3519} + y_{3520} + y_{3522}) + x_{3521}^T \beta$
 .
- 72 Bojonegoro : $\hat{y}_{3522} = -0,024(y_{3216} + y_{3517} + y_{3518} + y_{3519} + y_{3521} + y_{3523} + y_{3524}) + x_{3522}^T \beta$
 .
- 73 Tuban : $\hat{y}_{3523} = -0,043(y_{3216} + y_{3217} + y_{3522} + y_{3524}) + x_{3523}^T \beta$
 .
- 74 Lamongan : $\hat{y}_{3524} = -0,034(y_{3516} + y_{3517} + y_{3522} + y_{3523} + y_{3525}) + x_{3524}^T \beta$
 .
- 75 Gresik : $\hat{y}_{3525} = -0,034(y_{3515} + y_{3516} + y_{3524} + y_{3526} + y_{3578}) + x_{3525}^T \beta$
 .
- 76 Bangkalan : $\hat{y}_{3526} = -0,057(y_{3525} + y_{3527} + y_{3578}) + x_{3526}^T \beta$
 .
- 77 Sampang : $\hat{y}_{3527} = -0,085(y_{3526} + y_{3528}) + x_{3527}^T \beta$
 .
- 78 Pamekasan : $\hat{y}_{3528} = -0,085(y_{3527} + y_{3529}) + x_{3528}^T \beta$
 .
- 79 Sumenep : $\hat{y}_{3529} = -0,170 y_{3528} + x_{3529}^T \beta$
 .
- 80 Kota Pasuruan : $\hat{y}_{3575} = -0,170 y_{3514} + x_{3575}^T \beta$
 .
- 81 Pandeglang : $\hat{y}_{3601} = -0,085(y_{3602} + y_{3604}) + x_{3601}^T \beta$
 .
- 82 Lebak : $\hat{y}_{3602} = -0,034(y_{3201} + y_{3202} + y_{3601} + y_{3603} + y_{3604}) + x_{3602}^T \beta$
 .

83 Tangerang : $\hat{y}_{3603} = -0,021(y_{3101} + y_{3174} + y_{3175} + y_{3201} + y_{3602} + y_{3604} + y_{3671} + y_{3674}) + \mathbf{x}_{3603}^T \boldsymbol{\beta}$
.

84 Serang : $\hat{y}_{3604} = -0,034(y_{3601} + y_{3602} + y_{3603} + y_{3672} + y_{3673}) + \mathbf{x}_{3604}^T \boldsymbol{\beta}$
.