

PEMODELAN AUTOREGRESIF SPASIAL MENGGUNAKAN BAYESIAN MODEL AVERAGING UNTUK DATA PDRB JAWA *

Sarimah¹, Anik Djuraidah^{2‡}, and Aji Hamim Wigena³

¹Department of Statistics, IPB University, Indonesia, sarimah_s94@apps.ipb.ac.id

²Department of Statistics, IPB University, Indonesia, anikdjuraidah@apps.ipb.ac.id

³Department of Statistics, IPB University, Indonesia, ajiwigena@gmail.com

[‡]corresponding author

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)

Vol 3 No 3 (2019), 287 - 294

Copyright © 2019 Sarimah, Anik Djuraidah, and Aji H Wigena. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Economic data always contains spatial effects. Gross Regional Domestic Product (GRDP) in Java is one of economic data that describes spatial dependence between adjacent districts/cities. The method that is suitable for modeling GDRP is spatial regression with spatial dependence on lags that is spatial autoregressive. GDRP prediction used the Bayesian Model Averaging (BMA) method. The ten autoregressive spatial model that have highest posterior probability was chosen to determine the BMA model by posterior probability. The explanatory variables used in this study were (1) mean years of schooling (2) life expectancy (3) income per capita (4) local revenue (5) number of workers (6) district minimum salary. The results showed that the number of workers was chosen as a predictor for the ten models. The model that have highest posterior probability is 0.54 which contains five explanatory variables that are mean years of schooling, income per capita, local revenue, number of workers and district minimum salary and the pseudo R^2 of the model is 0.696.

Keywords: bayesian model averaging, posterior probability, spatial autoregressive.

1. Pendahuluan

Analisis regresi digunakan untuk menelaah hubungan antara satu peubah atau lebih. Peubah-peubah tersebut dibedakan menjadi menjadi dua bagian yakni peubah (Y) dan peubah penjelas (X). Pada analisis regresi diasumsikan antar amatan bersifat saling

* Received Jul 2019; Accepted Sep 2019; Published online on Oct 2019

bebas, namun pada data yang amatannya berupa area, asumsi ini sering tidak terpenuhi karena amatan di area tertentu dipengaruhi oleh amatan di tetangga sekitarnya. Menurut Anselin (1988) pemodelan pada data area perlu ditambahkan efek spasial, yaitu ketergantungan atau keragaman spasial.

Ketergantungan spasial pada peubah respon dapat dimodelkan dengan *spatial autoregressive* (SAR). Metode pendugaan parameter SAR biasanya menggunakan metode *maximum likelihood* (ML) yang dikemukakan oleh Ord (1975) dan metode pendugaan Bayes juga digunakan oleh Anselin (1988) dan LeSage (1997) untuk memodelkan regresi spasial. Zallner (1971) menggunakan Bayesian Model Averaging (BMA) untuk model regresi. Pada BMA, peluang posterior dari semua kemungkinan model digunakan untuk pemilihan model. Jika terdapat p peubah penjelas maka akan terdapat sebanyak m kemungkinan model dengan $m=2^p$ yaitu model yang memiliki kombinasi peubah-peubah penjelas. BMA dapat menentukan peubah yang relevan dengan data yang ada. Bivand et al. (2014) melakukan pendekatan dengan *integrated nested Laplace approximation* (INLA) untuk BMA yang secara komputasi lebih sederhana dari *Markov chain Monte Carlo* (MCMC), Rachmawati et al. (2018) menunjukkan bahwa algoritma INLA dapat melakukan inferensi Bayes menggunakan pendekatan *Laplace* yang memberikan hasil yang akurat dengan waktu komputasi lebih singkat dibanding MCMC.

Bivand et al. (2014) menyatakan bahwa data ekonomi selalu menunjukkan pola spasial. Pada penelitian ini akan digunakan data produk domestik regional bruto (PDRB) di Jawa tahun 2017 yang merupakan salah satu data ekonomi. Glasson (1997) dalam Putra (2017) menyatakan bahwa salah satu penentu eksternal yang penting dalam hal laju pertumbuhan PDRB antar daerah adalah tingkat permintaan dari daerah-daerah lain terhadap komoditi yang dihasilkan daerah tersebut, sehingga tujuan pada penelitian ini adalah memodelkan SAR menggunakan BMA dengan pendekatan INLA pada data PDRB Jawa tahun 2017 dengan peubah-peubah penjelas yaitu rata-rata lama sekolah (RLS), angka harapan hidup (AHH), pendapatan perkapita (PPK), pendapatan asli daerah (PAD), jumlah tenaga kerja (JTK) dan upah minimum kabupaten (UMK).

2. Metodologi

2.1 Data

Penelitian ini menggunakan data sekunder yaitu data PDRB Jawa atas dasar konstan tahun 2010 sebagai peubah respon. Peubah-peubah penjelas yang digunakan pada penelitian ini adalah jumlah tenaga kerja (JTK), pendapatan asli daerah (PAD), upah minimum kabupaten (UMK), rata-rata lama sekolah (RLS), pendapatan per kapita (PPK) dan angka harapan hidup (AHH) (Tabel 1). Sumber data dari publikasi Badan Pusat Statistik (BPS, 2018).

2.2 Metode Penelitian

Adapun tahapan analisis data pada penelitian ini adalah sebagai berikut:

Tabel 1: Peubah yang digunakan dalam penelitian

Peubah	Satuan
PDRB atas dasar harga konstan tahun dasar 2010 (Y)	Trilyun Rupiah
Jumlah tenaga kerja (X ₁)	Ribu Jiwa
Pendapatan asli daerah (X ₂)	Trilyun Rupiah
Upah minimum kabupaten (X ₃)	Juta Rupiah
Rata-rata lama sekolah (X ₄)	Tahun
Pendapatan perkapita (X ₅)	Juta Rupiah
Angka harapan hidup (X ₆)	Persen

1. Melakukan eksplorasi data

Membuat peta tematik 119 Kabupaen/Kota di Jawa dan penyebaran nilai PDRB dari setiap Kabupaten/kota serta menampilkan deskripsi datanya.

2. Membentuk matriks pembobot spasial

Pada penelitian ini digunakan pembobot *k-Nearest Neighbor*. Misalkan jarak pusat untuk setiap unit spasial *i* dengan seluruh unit *j*≠1 diurutkan sebagai berikut: $d_{ij(1)} \leq d_{ij(2)} \leq \dots \leq d_{ij(n-1)}$. Berdasarkan jarak *k* tetangga terdekat, pada penelitian ini digunakan *k*=2 yang artinya setiap area memiliki dua tetangga terdekatnya, misalkan dibentuk himpunan $N_k(i) = \{j(1), j(2), \dots, j(k)\}$ memuat *k* unit terdekat dengan *i*, matriks *k-Nearest Neighbor* untuk setiap *k* didefinisikan dalam bentuk:

$$w_{ij} = \begin{cases} 1, & j \in N_k(i) \\ 0, & \text{lainnya.} \end{cases} \tag{1}$$

Selanjutnya matriks bobot distandarisasi yaitu setiap elemen baris dibagi dengan jumlah baris sehingga

$$w_{ij} = \begin{cases} 1/k, & j \in N_k(i) \\ 0, & \text{lainnya.} \end{cases} \tag{2}$$

3. Uji ketergantungan spasial menggunakan Indeks Moran dan uji *Langrange multiplier* (LM) (Anselin, 1998).

4. Memodelkan semua kemungkinan model ($m=2^6=64$ model) menggunakan metode ML.

Model yang digunakan pada penelitian ini adalah model SAR sebagai berikut:

$$y = \rho W y + X \beta + \varepsilon \tag{3}$$

Fungsi log kemungkinan diperoleh dari persamaan (3) adalah

$$L(y) = \log|I - \rho W y| - \left(\frac{n}{2}\right) \log(2\pi\sigma^2) - \frac{\varepsilon' \varepsilon}{2\sigma^2} \tag{4}$$

6. Metode BMA digunakan dalam pendugaan parameter model. Model memiliki dua hiperparameter yaitu $\theta = (\theta_1, \theta_2)$,

a. Parameter presisi $\tau = 1/\sigma^2$ direpresentasikan sebagai $\theta_1 = \log \tau$

- b. Parameter regresi $\beta \sim N(0, \sigma^2)$
 c. Parameter ρ direpresentasikan sebagai

$$\rho^* = \frac{\rho - \rho_{\min}}{\rho_{\max} - \rho_{\min}} \quad (5)$$

dan

$$\theta_2 = \log(\rho^*/(1 - \rho^*)), \quad (6)$$

prior dari hiperparameter sebagai berikut:

- a. $\pi(\theta_1) = \log \text{gamma} (0.01, 0.01)$
 b. $\pi(\theta_2) = \text{logitbeta} (1,1)$.

Nilai-nilai prior ditentukan berdasarkan *default* dari *package* R-INLA, sehingga sebaran posteriornya adalah

$$P(y|\theta) = L(y)\pi(\theta_1)\pi(\theta_2) \quad (7)$$

7. Menghitung peluang model posterior (PMP) untuk semua kemungkinan model dengan rumus sebagai berikut (Huang et al., 2013):

$$p(M|y) = \exp(-0.5 * \text{BIC}) \quad (8)$$

dengan

$$\text{BIC} = -2 L(y|M_j) + d_j \log(n) \quad (9)$$

$L(y|M_j)$ adalah nilai log kemungkinan dari setiap model dan d_j merupakan banyaknya parameter dalam model.

8. Memilih 10 model dengan PMP tertinggi dan menghitung PMP untuk setiap model terpilih dengan rumus seperti dibawah ini

$$P(M|y) = \frac{P(M_i|y)}{\sum_{j=1}^m p(M_j|y)} \quad (10)$$

dengan ρ adalah parameter autokorelasi spasial, β adalah koefisien kovariat dan ε adalah galat menyebar gaussian normal dengan rata-rata 0 dan presisi τ .

9. Memodelkan SAR dengan BMA untuk 10 model terpilih sebagai berikut:

$$\hat{y} = \sum_{j^*=1}^{m^*} P(M_{j^*}|y) \times M_{j^*} \quad (11)$$

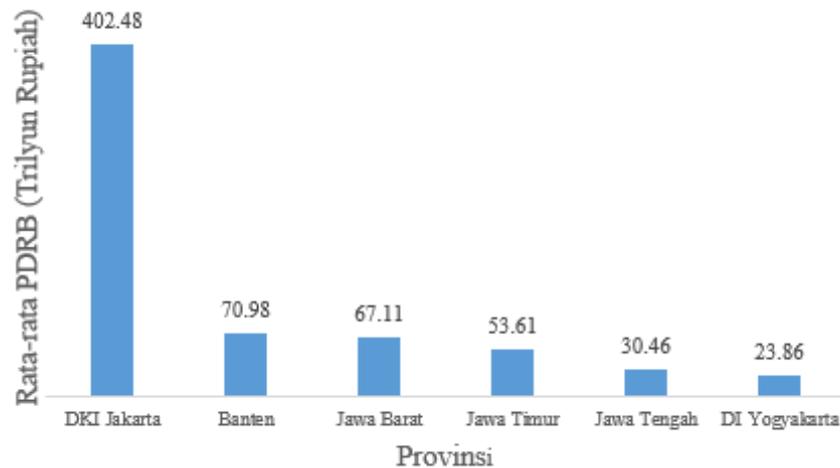
dengan M_{j^*} adalah model terpilih ke j^* dan $P(M_{j^*}|y)$ adalah peluang posterior model terpilih.

3. Hasil dan Pembahasan

3.1 Eksplorasi Data

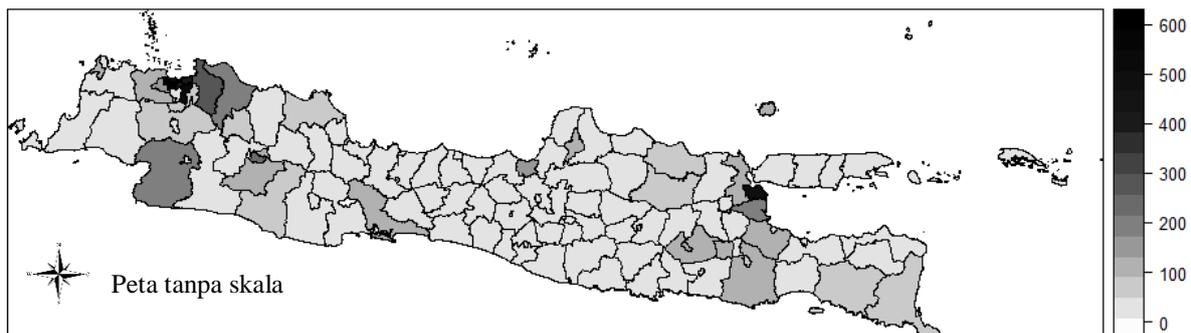
Jawa terdiri atas Pulau Jawa, Kepulauan Seribu dan Pulau Madura dengan 6 provinsi yaitu Banten, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta dan Jawa Timur. Jumlah kabupaten/kota dari seluruh provinsi sebanyak 119 kabupaten/kota. Rata-rata PDRB untuk setiap provinsi di Jawa tahun 2017 ditunjukkan pada Gambar 1. Provinsi DKI Jakarta memiliki rata-rata PDRB tertinggi dibanding dengan provinsi lainya di

Jawa, sedangkan rata-rata PDRB paling rendah di Jawa adalah provinsi DI Yogyakarta.



Gambar 1: Rata-rata PDRB setiap provinsi di Jawa tahun 2017.

Peta tematik penyebaran PDRB di Jawa ditunjukkan dalam Gambar 2 untuk melihat penyebaran PDRB di Jawa secara eksploratif. Gambar 2 menunjukkan adanya ketergantungan spasial antar lokasi pengamatan yang berdekatan. Hal ini terlihat dari kecenderungan kesamaan warna pada kabupaten/kota yang berdekatan. Secara keseluruhan dapat dilihat bahwa Provinsi Banten berada pada interval sedang. Provinsi DKI Jakarta cenderung memiliki PDRB yang tinggi dengan warna gelap pada peta sedangkan Provinsi Jawa Barat, Jawa Timur dan Jawa Tengah terlihat berada pada interval sedang. Provinsi Yogyakarta berada pada interval yang relatif rendah.



Gambar 2: Peta tematik penyebaran PDRB Jawa.

3.2 Uji Efek Spasial

Uji ketergantungan spasial pada peubah respon dilakukan dengan Indeks Moran dan uji LM. Nilai Indeks Moran sebesar 0.39 dengan nilai-p sebesar $2.52 \cdot 10^{-07}$ artinya terdapat autokorelasi spasial antar PDRB di Jawa dengan taraf nyata 5%. Uji LM menghasilkan nilai statistik uji untuk model SAR 4.63 dengan nilai-p 0.03 (nilai-p lebih

kecil dari 0.05), artinya model SAR signifikan pada taraf nyata 0.05. Ringkasan statistik dapat dilihat pada Tabel 2.

Tabel 2: Hasil Uji *Langrange Multiplier*.

Model	Statistik Uji	Nilai P
SAR	4.63	0.03
SEM	2.91	0.08
SARMA	2.93	0.23

3.3 Pemilihan Model dan Pendugaan parameter Regresi

Tabel 3 menyajikan nilai log kemungkinan dan BIC dari 10 model yang digunakan untuk menghitung PMP masing-masing model. Nilai log kemungkinan dan BIC didapatkan dari model SAR dengan pendugaan LM. Model-model tersebut menggunakan kombinasi peubah penjelas seperti yang ditunjukkan pada Tabel 3.

Tabel 3: Nilai log likelihood dan BIC dari 10 model dengan PMP tertinggi.

Peubah penjelas	Log kemungkinan	BIC
RLS+PPK+JTK+PAD+UMK	-100.42	239.07
RLS+PPK+JTK+UMK	-103.58	240.60
RLS+PPK+AHH+JTK+PAD+UMK	-99.43	241.86
RLS+PPK+AHH+JTK+UMK	-102.53	243.28
PPK+JTK+PAD+UMK	-106.96	247.37
PPK+JTK+UMK	-111.55	251.77
PPK+AHH+JTK+PAD+UMK	-106.87	251.97
JTK+PAD+UMK	-111.74	252.16
RLS+PPK+JTK+PAD	-110.55	254.56
PPK+JTK+PAD	-113.52	255.72

Tabel 4 menunjukkan peubah yang terdapat dalam 10 model dengan PMP tertinggi. Peubah yang terdapat dalam setiap model ditandai dengan "1" dan peubah yang tidak muncul ditandai dengan "0". Dari Tabel 4 terlihat bahwa peubah JTK muncul dalam semua model, PPK terdapat dalam 9 model dan UMK terdapat dalam 8 model.

Pada Tabel 5 disajikan dugaan koefisien regresi menggunakan metode Bayes. dugaan nilai ρ serta PMP dari 10 model terpilih. Model 1 dengan peluang posterior tertinggi adalah model dengan peubah penjelas RLS, PPK, PAD, JTK dan UMK.

Model 10 dengan peluang posterior terendah memiliki peubah penjelas RLS, AHH dan PPK. Koefisien peubah RLS bernilai negatif, sedangkan koefisien peubah lainnya bernilai positif. Tanda dari parameter ρ sama untuk semua model yaitu negative artinya terdapat autokorelasi negatif pada data PDRB di Jawa. Pada model BMA ini menghasilkan pseudo R^2 sebesar 0.696, artinya model yang dibentuk dapat menjelaskan 69.6% keragaman dari PDRB.

Tabel 4: Peubah yang masuk ke dalam 10 model dengan PMP tertinggi.

Peubah	Model									
	1	2	3	4	5	6	7	8	9	10
RLS	1	1	1	1	0	0	0	0	1	0
AHH	0	0	1	1	0	0	1	0	0	0
PPK	1	1	1	1	1	1	1	0	1	1
PAD	1	0	1	0	1	0	1	1	1	1
JTK	1	1	1	1	1	1	1	1	1	1
UMK	1	1	1	1	1	1	1	1	0	0
PMP	0.54	0.25	0.13	0.07	0.01	9.44×10^{-4}	8.54×10^{-4}	7.76×10^{-4}	2.34×10^{-4}	1.31×10^{-4}

Tabel 5: Koefisien Regresi dan Nilai ρ dari 10 model terpilih

Koefisien	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Intersep	1.81	1.31	-0.35	-0.97	1.44	0.73	2.09	2.28	1.72	1.46
RLS	-0.2	-0.22	-0.23	-0.26	-	-	-	-	-0.14	-
AHH	-	-	0.03	0.03	-	-	-9.38×10^{-3}	-	-	-
PPK	0.17	0.22	0.18	0.22	0.09	0.13	0.09	-	0.17	0.11
PAD	1.59×10^{-4}	-	1.57×10^{-4}	-	2×10^{-4}	-	1.99×10^{-4}	3.1×10^{-4}	2.54×10^{-4}	2.71×10^{-4}
JTK	1.02×10^{-3}	1.07×10^{-3}	9.53×10^{-4}	9.98×10^{-4}	1.13×10^{-3}	1.21×10^{-3}	1.14×10^{-3}	9.28×10^{-4}	1.36×10^{-3}	1.39×10^{-3}
UMK	0.55	0.64	0.62	0.72	0.44	0.55	0.43	0.51	-	-
P	-0.12	-0.09	-0.13	-0.11	-0.16	-0.13	-0.14	-0.14	-0.01	-0.01
PMP	0.54	0.25	0.13	0.07	0.01	9.44×10^{-4}	8.54×10^{-4}	7.76×10^{-4}	2.34×10^{-4}	1.31×10^{-4}

4. Simpulan

Pemodelan menggunakan *bayesian model averaging* menunjukkan bahwa peubah JTK terdapat dalam 10 model terpilih. peluang posterior tertinggi sebesar 0.54. Pada data ini, model dengan peluang posterior tertinggi adalah model yang menggunakan 5 peubah penjelas yaitu peubah RLS, PPK, PAD, JTK dan UMK. Model BMA ini menghasilkan pseudo R^2 sebesar 0.696.

Daftar Pustaka

- Anselin, L. (1988). *Spatial econometrics: methods and models*. New York (US): Springer Science & Business Media.
- Anselin, L. (1998). *Spatial econometrics*. Toledo (US): Departemen of Economics University of Toledo.
- Bivand, R. S., Gómez-Rubio, V., & Rue, H. (2014). Approximate Bayesian inference for spatial econometrics models. *Spatial Statistics*, **9**: 146–165.
- [BPS] Badan Pusat Statistik. (2018). *Produk domestik regional bruto kabupaten/kota di Indonesia 2013-2018*. Jakarta (ID): Badan Pusat Statistik.
- Glasson, J. (1997). *Pengantar perencanaan regional*. (penerjemah: Paul Sitohang). Jakarta (ID): Fakultas Ekonomi Universitas Indonesia.
- Huang, J., Niu, C., Green, C. D., Yang, L., Mei, H., & Han, J.-D. J. (2013). Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network. *PLoS Computational Biology*, **9**(3): e1002998.
- LeSage, J. P. (1997). Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, **20**(1–2): 113–129.
- Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, **70**(349): 120–126.
- Putra, F. S. (2017). *Analisis spasial produk domestik regional bruto (PDRB) kabupaten/kota di provinsi jawa tengah* (Skripsi). Universitas Islam Indonesia, Yogyakarta (ID).
- Rachmawati, R. N., Djuraidah, A., Fitrianto, A., & Sumertajaya, I. M. (2018). Spatio-temporal models using R-INLA with generalized extreme value distribution in hierarchical bayesian regression. *IJRSET*, **4**(4): 2394–4099.
- Zallner, A. (1971). *An introduction to bayes inference in econometrics*. New York (US): Wiley.