Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)
Volume 09, No. 3 (2020), hal 413 - 422.

PEMODELAN *FIXED EFFECT GEOGRAPHICALLY WEIGHTED PANEL REGRESSION* UNTUK INDEKS PEMBANGUNAN MANUSIA DI KALIMANTAN BARAT

Sutro, Yundari, Shantika Martha

INTISARI

Indeks Pembangunan Manusia (IPM) mengukur pencapaian penduduk dalam mengakses hasil pembangunan untuk memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. Analisis terhadap faktor yang mempengaruhi IPM perlu dilakukan dalam beberapa periode waktu dan memperhatikan letak geografis atau lokasi pengamatan. Oleh karena itu, berkembanglah analisis regresi panel yang melibatkan analisis regresi lokal yaitu geographically weighted panel regression (GWPR). GWPR merupakan penggabungan antara model geographically weighted regression (GWR) dengan model regresi data panel. Tujuan penelitian ini adalah mengetahui model dan variabel-variabel yang berpengaruh signifikan terhadap indeks pembangunan manusia di Kalimantan Barat pada 14 kabupaten/kota dalam rentang tahun 2011-2015. Hasil uji Chow dan uji Hausman menunjukkan bahwa model estimasi regresi data panel yang sesuai adalah fixed effect model (FEM) yang menghasilkan nilai adjusted R^2 *sebesar 0,8019. Selanjutnya estimasi Parameter dengan metode weighted least square (WLS) yang melibatkan jarak euclidean tiap-tiap lokasi. Bandwidth optimum yang dipilih berdasarkan nilai cross validation (CV) yang paling minimum antara fungsi pembobot fixed kernel gaussian, bisquare, dan tricube. Fungsi pembobot fixed kernel gaussian dipilih dengan alasan memiliki nilai CV paling minimum yaitu 64,90764. Hasil penelitian ini menunjukkan bahwa model yang dihasilkan untuk Kabupaten/Kota di Provinsi Kalimantan Barat masing-masing berbeda dan Provinsi Kalimatan Barat terbagi menjadi dua kelompok berdasarkan variabel yang berpengaruh signifikan terhadap IPM. Hasil prediksi menggunakan model GWPR diperoleh nilai MAPE 1,4%, hal tersebut menyatakan bahwa model GWPR sangat baik.*

Kata kunci: *WLS, bandwidth, fixed kernel.*

PENDAHULUAN

Salah satu parameter keberhasilan suatu negara dapat dilihat dari kondisi masyarakatnya. Masyarakat yang unggul didapatkan dengan meningkatkan pembangunan Sumber Daya Manusia (SDM) menuju kearah yang lebih baik. Indeks Pembangunan Manusia (IPM) atau disebut *Human Development Index* (HDI) ditetapkan oleh Perserikatan Bangsa-Bangsa (PBB) sebagai ukuran standar pembangunan manusia. Secara global, IPM juga dapat digunakan sebagai penentu suatu negara termasuk dalam negara maju atau negara berkembang.

Pengamatan hasil pencapaian dari IPM Kalimantan Barat di setiap kabupaten/kota perlu dilakukan dalam beberapa periode waktu agar informasi yang diperoleh lebih akurat. Oleh karena itu, berkembanglah analisis regresi panel yang melibatkan unit *cross section* dan unit *time series*. Sebaran tiap-tiap lokasi perlu diperhatikan karena kondisi yang sangat beragam. Adanya perbedaan setiap wilayah di Kalimantan Barat menyebabkan permasalahan spasial karena faktor geografis yang mempengaruhi suatu wilayah terhadap wilayah lainnya, sehingga diperlukan suatu metode pemodelan statistik yang memperhatikan letak geografis atau faktor lokasi pengamatan [1]. Metode yang dapat digunakan untuk mempertimbangkan adanya faktor spasial adalah *geographically weighted regression* (GWR). Pengembangan metode untuk analisis spasial-temporal dengan menggabungkan antara model GWR dan model regresi panel dikenal dengan *geographically weighted panel regression* (GWPR). Pada penelitian ini, digunakan model GWPR dengan efek tetap (*fixed effect*) menggunakan pembobot *kernel Gaussian* pada data indeks pembangunan manusia di Kalimantan Barat tahun 2011-2015 untuk mengidentifikasi variabel-variabel yang mempengaruhi IPM di Kalimantan Barat. Data IPM yang digunakan merupakan data sekunder dari Badan Pusat Statistik (BPS) Provinsi Kalimantan Barat.

ANALISIS REGRESI

Analisis regresi adalah metode dalam menganalisis hubungan antara dua atau lebih variabel. Model regresi terbagi menjadi dua jenis yaitu Regresi linear sederhana dan Regresi linear berganda. Bentuk umum regresi linear sederhana sebagai berikut [2] ,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

sedangkan persamaan regresi linear berganda ditulis sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i$$

dengan y_i adalah variabel respon untuk unit observasi ke-*i*, β_0 adalah intersep, $\beta_1, \beta_2, ..., \beta_k$ adalah parameter untuk variabel prediktor $x_1, x_2, ..., x_k$, $x_{1i}, x_{2i}, \ldots, x_{ki}$ adalah variabel prediktor untuk unit observasi ke-*i* dengan *k* adalah jumlah parameter regresi dan ε_i adalah komponen *error* untuk unit observasi ke-*i*. Analisis regresi hanya dapar digunakan untuk memodelkan atau meramalkan dua variabel prediktor atau lebih dengan variabel respon tanpa melibatkan unsur lokasi. Oleh karena itu, digunakan metode *geographically weighted regression* (GWR) yang melibatkan jarak antar lokasi pengamatan.

GWR merupakan metode yang digunakan untuk mengeksplorasi nonstasioner spasial, yang didefinisikan sebagai sifat dan hubungan yang signifikan antar variabel yang berbeda pada lokasi satu ke lokasi lainnya[1]. Parameter untuk model regresi di setiap lokasi akan menghasilkan nilai yang berbeda-beda. Model GWR ditulis sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i ; i = 1, 2, \ldots, n$$

dengan y_i merupakan nilai observasi variabel respon ke-*i*, x_{ik} merupakan nilai observasi variabel prediktor *k* pada pengamatan ke-*i*, β_0 merupakan intersep, β_k merupakan parameter variabel prediktor *k*, (u_i, v_i) merupakan titik koordinat lokasi *i*, dan ε_i merupakan *error* ke-*i*. Pada pengujian model GWR diasumsikan bahwa data yang diamati di dekat lokasi *i* lebih berpengaruh pada estimasi parameter $\beta_k(u_i, v_i)$ dibandingkan dengan data yang jauh dari lokasi *i*. Estimasi parameter model GWR dapat dilakukan dengan memberikan pembobotan pada setiap observasi sesuai dengan kedekatannya pada lokasi *i*, Sehingga estimator untuk koefisien regresi lokal pada GWR adalah sebagai berikut.

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i)\mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i)\mathbf{y}$$

dengan $\hat{\boldsymbol{\beta}}(u_i, v_i) = (\hat{\beta}_{i0}, \hat{\beta}_{i1}, \hat{\beta}_{i2}, ..., \hat{\beta}_{ip})^T$ adalah vektor koefisien regresi lokal dan $\mathbf{W}(u_i, v_i)$ adalah matriks diagonal dengan elemen pada diagonalnya merupakan pembobot geografis pada setiap data untuk lokasi pengamatan ke-*i*.

FUNGSI PEMBOBOT SPASIAL

Pada analisis spasial, penaksiran parameter disuatu titik (u_i, v_i) akan lebih dipengaruhi oleh titik-titik yang dekat daripada titik-titik yang lebih jauh. Oleh karena itu, pemilihan pembobot spasial yang digunakan dalam menaksir parameter menjadi sangat penting. Pada penelitian ini digunakan fungsi pembobot *kernel fixed* yaitu fungsi *kernel* yang memiliki *bandwidth* yang sama pada setiap lokasi pengamatan. Pembobot yang terbentuk dari fungsi *kernel fixed* terdiri dari *gaussian, bisquare,* dan *tricube* [3].

Gaussian

$$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right) \quad (1)$$

Bisquare

$$w_{ij} = \begin{cases} \left(1-\left(\dfrac{d_{ij}}{h}\right)^2\right)^2 & \text{, untuk } d_{ij} \le h \\ 0 & \text{, untuk lainnya} \end{cases}$$

Tricube

$$w_{ij} = \begin{cases} \left(1-\left(\dfrac{d_{ij}}{h}\right)^3\right)^3 & \text{, untuk } d_{ij} \le h \\ 0 & \text{, untuk lainnya} \end{cases}$$

d_{ij} adalah jarak antara titik di lokasi i dan lokasi j yang didapatkan dari jarak *euclidean* $(d_{ij})^2 = (u_i - u_j)^2 + (\mathrm{v}_i - v_j)^2$, Sementara h adalah *bandwidth* atau parameter penghalus yang merupakan pengontrol keseimbangan antara kesesuaian kurva terhadap data dan kemulusan data. Metode yang dapat digunakan untuk mendapatkan *bandwidth* optimum adalah menggunakan pendekatan *cross validation* (CV). *Bandwidth* yang optimum diperoleh jika nilai CV yang dihasilkan adalah yang paling minimum.

$$CV = \sum_{i=1}^{n} \left[y_i - \hat{y}_{\neq i}(h) \right]^2$$

$\hat{y}_{\neq i}(h)$ adalah nilai taksiran untuk y_i dengan menghilangkan observasi pada titik i dari proses pengujian parameter

REGRESI DATA PANEL

Regresi data panel adalah regresi yang menggunakan data pengamatan terhadap satu atau lebih variabel pada suatu unit secara terus menerus selama beberapa periode waktu. Data panel adalah gabungan dari data *cross section* dan *time series* [4]. Keuntungan menggunakan data panel adalah memberikan data yang lebih informatif, bervariasi, mengurangi koliniearitas antar variabel, memperbesar derajat kebebasan dan lebih efisien [5]. Model umum regresi data panel adalah sebagai berikut,

$$y_{it} = \beta_{it} + \boldsymbol{\beta}'_{it}\mathbf{x}_{it} + \varepsilon_{it}$$

Dengan y_{it} merupakan nilai variabel respon ke-*i* waktu ke-*t*, β_{it} merupakan koefisien intersep dari unit ke-*i* pada waktu ke-*t* , $\boldsymbol{\beta}'_{\mathbf{it}}$ adalah vektor konstanta berukuran $1\times k$ atau *slope* dari unit ke-*i* yang diamati pada periode ke-*t*, $\mathbf{x}_{\mathbf{it}}$ menunjukkan vektor observasi pada variabel prediktor berukuran $1\times p$, dan ε_{it} adalah residual pada unit ke-*i* dan waktu ke-*t*. Estimasi model regresi data panel tergantung pada asumsi yang dibuat terhadap intersep dan koefisien *slope* sehingga ada beberapa kemungkinan yaitu sebagai berikut.

Diasumsikan intersep dan *slope* konstan sepanjang waktu dan individu, sehingga persamaannya menjadi

$$y_{it} = \beta + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$$

Kondisi kedua diasumsikan *slope* konstan, tetapi intersep bervariasi untuk setiap individu, sehingga persamaannya menjadi

$$y_{it} = \beta_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$$

Kondisi ketiga diasumsikan intersep konstan, tetapi *slope* bervariasi untuk setiap individu, sehingga persamaannya menjadi

$$y_{it} = \beta + \boldsymbol{\beta}_i'\mathbf{x}_{it} + \varepsilon_{it}$$

Berdasarkan asumsi-asumsi tersebut maka terdapat tiga model yang digunakan dalam mengestimasi regresi data panel yaitu *common effect model* (CEM), *fixed effect model* (FEM), *random effect model* (REM).

COMMON EFFECT MODEL (CEM)

Model CEM mengasumsikan bahwa nilai intersep dan *slope* koefisien untuk semua unit dan waktu adalah sama. Data dikombinasikan tanpa memperhatikan pengaruh antar waktu dan individu. Metode estimasi yang digunakan pada CEM adalah dengan metode *ordinary least square* (OLS). Adapun model CEM adalah sebagai berikut [6].

$$y_{it} = \beta + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}, i = 1, 2, .., N, t = 1, 2, \ldots, T_i$$

FIXED EFFECT MODEL (FEM)

FEM juga dikenal dengan *least square dummy variable* (LSDV) merupakan model regresi data panel yang menggunakan variabel *dummy* pada koefisien residual dengan nilai berbeda-beda untuk setiap individu ke-*i*. Diasumsikan bahwa tidak ada pengaruh waktu dan hanya fokus pada efek individu [7]. Bentuk umum regresi data panel pada FEM sama dengan persamaan berikut.

$$y_{it} = \beta_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$$

Intersep dari unit *cross-section* berbeda-beda, hal ini dilihat dari indeks *i* pada intersep. Perbedaan intersep yang timbul antarindividu dapat dijelaskan dengan menggunakan variabel *dummy*, sehingga persamaan dapat ditulis sebagai berikut,

$$y_{it} = D\beta_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$$

dengan $D = d_1, d_2, \ldots d_i$ merupakan variabel *dummy*.

RANDOM EFFECT MODEL (REM)

REM juga disebut dengan *error component model* (ECM) merupakan model regresi data panel dengan melibatkan pengaruh individu yang acak dan tidak berkorelasi dengan variabel prediktor yang diamati. Efek *cross-section* pada unit individu dipandang sebagai *error*, sehingga ε_i merupakan *error* pada unit individu. Pada REM, β_i diasumsikan sebagai variabel random dengan rata-rata β_0 sehingga intersep tiap individu adalah $\beta_i = \beta_0 + \varepsilon_i$, sehingga persamaan REM dapat ditulis sebagai berikut[7].

$$\begin{aligned} y_{it} &= \beta_0 + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_i + \varepsilon_{it} \\ &= \beta_0 + \boldsymbol{\beta}'\mathbf{x}_{it} + w_{it} \end{aligned}$$

dengan $w_{it} = \varepsilon_i + \varepsilon_{it}$, dimana ε_i adalah komponen *error cross-section*, ε_{it} adalah komponen *error cross-section* dan *error time series*.

UJI PEMILIHAN MODEL DAN UJI ASUMSI REGRESI DATA PANEL

Pemilihan model yang akan digunakan dalam mengestimasi regresi data panel dilakukan dengan dua uji yaitu uji Chow dan uji Hausman. Uji Chow digunakan untuk mengetahui model terbaik antara FEM dan CEM dengan melihat signifikansi model FEM melalui uji statistik F. Hipotesis dan F statistik uji Chow yaitu sebagai berikut.

$H_0 : \beta_1 = \beta_2 = \beta_i = \ldots = 0$ (model *common effect*)

H_1 : minimal ada satu $\beta_i \neq 0$ (model *fixed effect*), $i = 1, 2, \ldots, N$.

$$F_0 = \frac{\dfrac{RSS_{CEM} - RSS_{FEM}}{n-1}}{(RSS_{FEM})/(\mathrm{n}T - n - K)}$$

dengan :

RSS_{CEM} = *residual sum of squares* model *common effect*

RSS_{FEM} = *residual sum of squares* model *fixed effect*

Hipotesis nol akan ditolak jika nilai statistik F_0 lebih besar dari nilai $F_{(\alpha;k-1;n-k)}$ pada tingkat signifikansi tertentu atau *p-value* < α . Hal ini berarti asumsi koefisien intersep dan *slope* adalah sama tidak berlaku, sehingga teknik regresi data panel dengan FEM lebih baik dari model regresi data panel dengan CEM.

Uji Hausman digunakan untuk mengetahui model terbaik antara FEM dan REM dengan mengikuti kriteria Wald. Hipotesis untuk uji Hausman adalah sebagai berikut.

H_0 : model *random effect*

H_1 : model *fixed effect.*

Nilai statistik uji Hausman akan mengikuti distribusi *chi-square* sebagai berikut.

$$W = \chi^2\left[p\right] = \left[\hat{\beta}_{FEM}, \hat{\beta}_{REM}\right]^T \left[\mathrm{var}\left[\hat{\beta}_{FEM} - \hat{\beta}_{REM}\right]\right]^{-1} \left[\hat{\beta}_{FEM} - \hat{\beta}_{REM}\right]$$

Statistik uji Hausman mengikuti distribusi statistik χ^2 dengan derajat bebas sebanyak jumlah peubah bebas (p). Hipotesis nol ditolak jika nilai statistik Hausman lebih besar dari nilai kritis statistik $\chi^2_{(\alpha,k)}$ atau *p-value* < α yang berarti model yang tepat untuk regresi data panel adalah model FEM [8]. Setelah dilakukan uji Chow dan uji Hausman maka didapatkan model yang sesuai dalam mengestimasi regresi data panel. Selanjutnya model yang didapatkan akan dilakukan uji asumsi klasik yang meliputi autokorelasi, normalitas, heteroskedastisitas, dan multikoliniearitas.

Autokorelasi merupakan korelasi antara satu variabel gangguan dengan variabel gangguan yang lain [9]. Salah satu metode yang dapat digunakan untuk mendeteksi autokorelasi adalah uji *Breusch-Godfrey* atau uji *lagrange multiplier* (LM). Asumsi non autokorelasi tidak terpenuhi jika nilai *p-value* > α.

Populasi data berdistribusi normal dapat diketahui dengan melakukan uji Kolmogorov-Smirnov. Deskripsi data berdistribusi normal dilihat dengan *normal probability plot* akan terlihat titik-titik grafik plot yang relatif berhimpitan dengan sumbu diagonal.

H_0 : Data berdistribusi normal

H_1 : Data tidak berdistribusi normal

Kriteria pengambilan keputusan adalah terima H_0 jika $KS_{hitung} < KS_{tabel}$. Adapun nilai KS_{hitung} diperoleh dengan formula sebagai berikut[10].

$$KS_{hitung} = \max\left|F(z_i) - S(z_i)\right|$$

dengan $z_i = \frac{x_i - \overline{x}}{s}$ ($\overline{x}$ dan s masing-masing merupakan rata-rata dan simpangan baku), $F(z_i) = P(z \le z_i)$ dan $S(z_i)$ adalah proporsi $\frac{z_1, z_2, ..., z_n \le z_i}{n}$, jika data tidak berdistribusi normal maka untuk membentuk data normal dapat dilakukan transformasi Box-Cox terhadap data.

Uji Glejser dapat digunakan untuk membuktikan tidak adanya kasus heteroskedastisitas. Glejser menyarankan untuk melakukan regresi nilai absolut residual dengan variabel prediktor sebagai berikut.

$$\left|\hat{e}_i\right| = \beta_0 + \beta_1 x_i + v_i$$

dimana $\hat{e}_i$ adalah nilai residual dari model dan v_i adalah residual untuk masing-masing fungsi. Jika β_1 tidak signifikan melalui uji t maka dapat disimpulkan tidak terjadi heteroskedastisitas dan sebaliknya jika β_1 signifikan secara statistik maka model terdapat kasus heteroskedastisitas.

Multikoliniearitas merupakan hubungan antara variabel prediktor dalam regresi berganda. Hubungan linear antara variabel prediktor dapat terjadi dalam bentuk hubungan linear yang sempurna dan hubungan linear yang kurang sempurna [11]. Salah satu cara untuk mendeteksi adanya kasus multikolinieritas antar variabel prediktor yaitu dengan melihat nilai *variance inflation factor* (VIF). Jika nilai VIF > 10 maka terdeteksi adanya masalah multikoliniearitas.

FIXED EFFECT GEOGRAPHICALLY WEIGHTED PANEL REGRESSION

Geographically weighted panel regression (GWPR) merupakan model pengembangan yang memadukan antara model GWR dengan regresi data panel. Pada penelitian ini diasumsikan bahwa kondisi tiap unit pengamatan saling berbeda, sehingga digunakan regresi data panel dengan model FEM. Berikut adalah kombinasi dari persamaan GWR dan persamaan regresi data panel FEM:

$$y_{it} = \beta_0\left(u_{it}, v_{it}\right) + \sum_{k=1}^{p} \beta_k\left(u_{it}, v_{it}\right) x_{itk} + \varepsilon_{it}; i = 1, 2 \ldots N; t = 1, 2, \ldots T; k = 1, 2 \ldots p$$

dengan:

k	= variabel prediktor ke-*k*
t	= waktu pengamatan
i	= lokasi pengamatan
y_{it}	= nilai respon pada pengamatan ke-*i* dan waktu ke-*t*.
x_{itk}	= nilai variabel prediktor ke-*k* pada pengamatan ke-*i* dan waktu ke-*t*.
$\beta_0(u_{it}, v_{it})$	= konstanta/*intersept* dari pengamatan ke-*i* dan waktu ke-*t*.
(u_{it}, v_{it})	= titik koordinat lokasi pengamatan ke-*i* dan waktu ke-*t*.
ε_{it}	= *random error* yang diasumsikan prediktor, identik dan mengikuti distribusi normal dengan mean nol dan varian konstan.

PEMILIHAN PEMBOBOT FIXED EFFECT GWPR

Pembobot dalam model GWPR sama dengan pembobot pada model GWR, dimana tergantung pada jarak antar titik lokasi pengamatan [12]. Salah satu cara untuk menentukan matriks pembobot adalah dengan menggunakan fungsi *kernel* yang dinyatakan seperti pada Persamaan (1). Jarak *euclidean* antara lokasi *i* dan lokasi *j* dinotasikan dengan d_{ij} dan h merupakan nilai bandwidth. Bandwidth dianalogikan sebagai radius suatu lingkaran sehingga sebuah titik lokasi pengamatan yang berada dalam radius lingkaran masih dianggap berpengaruh dalam membentuk parameter di titik lokasi pengamatan ke-*i*. Metode yang dapat digunakan untuk memilih *bandwidth* optimum yaitu menggunakan nilai *cross validation* (CV). Nilai CV dihitung berdasarkan rata-rata variabel respon dan prediktor untuk keseluruhan waktu yang didefinisikan sebagai berikut:

$$CV = \sum_{i=1}^{n} (\bar{y}_i - \hat{\bar{y}}_{\neq i}(b))^2$$

dengan $\bar{y}_i$ adalah rata-rata dari waktu ke waktu variabel respon di lokasi pengamatan ke-*i* dan $\hat{\bar{y}}_{\neq i}(b)$ adalah nilai estimator y_i dengan bandwidth *b* dengan pengamatan di lokasi (u_i, v_i) dihilangkan dari proses estimasi.

ESTIMASI PARAMETER MODEL FIXED EFFECT GWPR

Parameter yang dihasilkan pada model GWPR akan berbeda-beda pada masing-masing lokasi dan waktu, sehingga terdapat sebanyak (*NT* x *k)* parameter yang harus diestimasi, dimana *N* adalah jumlah lokasi pengamatan, *T* adalah waktu pengamatan dan $k = p+1$ jumlah parameter pada masing-masing lokasi dan waktu pengamatan. Estimasi parameter model *fixed effect* GWPR menggunakan pendekatan *weighted least square* (WLS) sehingga diperoleh estimator dari parameter model sebagai berikut:

$$\hat{\boldsymbol{\beta}}(u_{it}, v_{it}) = (\mathbf{X}^T \mathbf{W}(u_{it}, v_{it})\mathbf{X})^{-1}\mathbf{X}^T \mathbf{W}(u_{it}, v_{it})\mathbf{Y}$$

dimana *it* menunjukkan lokasi ke-*i* dan waktu ke-*t* pada matriks $\boldsymbol{\beta}$ dan $\mathbf{W}(u_{it}, v_{it})$ adalah matriks pembobot spasial untuk lokasi pengamatan ke-*i* dan waktu ke-*t*.

PENGUJIAN KESESUAIAN MODEL *FIXED EFFECT* GWPR

Pengujian ini dilakukan untuk mengetahui adanya perbedaaan antara model regresi data panel *fixed effect* dengan *fixed effect* GWPR dengan hipotesis sebagai berikut [13],

$H_0 : \beta_k(u_i, v_i) = \beta_k$ untuk setiap *k* = 1,2,…*p* dan *i* = 1,2,…*n* (tidak ada perbedaan yang signifikan antara model regresi data panel *fixed effect* dan *fixed effect* GWPR),

$H_1 :$ minimal terdapat satu $\beta_k(u_i, v_i) \neq \beta_k$ untuk *k* = 1,2,…*p* dan *i* = 1,2,..*n* (terdapat perbedaan yang signifikan antara model regresi data panel *fixed effect* dan *fixed effect* GWPR).

Statistik uji yang digunakan yaitu,

$$F = \frac{\dfrac{RSS(H_1)}{df_1}}{\dfrac{RSS(H_0)}{df_2}}$$

dengan :

$RSS(H_0) = \mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y}$ dimana $\mathbf{H} = \mathbf{X}(\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{X}^\mathbf{T}$,

$RSS(H_1) = \mathbf{Y}^T(\mathbf{I} - \mathbf{L})^\mathbf{T}(\mathbf{I} - \mathbf{L})\mathbf{Y}$

$df_1 = \dfrac{\delta_1^2}{\delta_2}$ dimana $\delta_i = tr\left(\left[(\mathbf{I} - \mathbf{L})^\mathbf{T}(\mathbf{I} - \mathbf{L})\right]^i\right)$, *i* = 1,2 ; $df_2 = n - p - 1$.

I merupakan matriks identitas berukuran *nt* x *nt* serta *L* merupakan matriks proyeksi dari model GWPR. H_0 akan ditolak jika $F < F_{1-\alpha, df_1, df_2}$ atau *p-value* > α yang berarti terdapat perbedaan yang signifikan antara model regresi data panel *fixed effect* dan *fixed effect* GWPR.

UJI SIGNIFIKANSI PARAMETER MODEL

Pengujian ini dilakukan untuk mengetahui parameter yang mempengaruhi variabel respon secara signifikan. Adapun hipotesis pengujiannya sebagai berikut [14] :

$H_0 : \beta_k(u_i, v_i) = 0$, untuk *k* = 1,2,…*p* dan *i* = 1,2,…*n* ;

$H_1 :$ minimal ada satu $\beta_k(u_i, v_i) \neq 0$ untuk *k* = 1,2,…*p* dan *i* = 1,2,…*n*.

Statistik uji yang digunakan yaitu :

$$T_{hit} = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma}\sqrt{C_{kk}}}$$

dimana $\hat{\sigma} = \sqrt{\dfrac{RSS(H_1)}{\delta_1}}$ *dan* C_{kk} adalah elemen diagonal ke-*k* dari matriks $C_{it}C_{it}^T$ dengan

$C_{it} = \left(X'W(u_i,v_i)X\right)^{-1} X'W(u_i,v_i)$, T_{hit} akan mengikuti distribusi t dengan derajat bebas $df = \frac{\delta_1^2}{\delta_2}$.

Jika diberikan tingkat signifikansi α, maka H_0 akan ditolak jika $|T_{hit}| > T_{(\alpha/2;df)}$ atau *p-value* < α.

STUDI KASUS

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Kalimantan Barat untuk 14 Kabupaten/Kota. Variabel respon (Y) yang digunakan yaitu angka Indeks Pembangunan Manusia (IPM) sedangkan untuk variabel prediktor (X) yaitu Angka Partisipasi Sekolah (APS) tingkat SMP (X_1), pengeluaran perkapita (X_2), Tingkat Pengangguran Terbuka (TPT) (X_3), serta jumlah sarana kesehatan yang terdiri dari puskesmas, puskesmas pembantu, puskesmas keliling, dan rumah sakit (X_4). Data yang digunakan pada penelitian ini diambil dari rentang tahun 2011-2015. Perolehan angka IPM Kalimatan Barat pada tahun 2015 masih dibawah angka IPM Indonesia yaitu sebesar 69,55% dan membuat Provinsi Kalimantan Barat berada pada peringkat ke-29 dari 34 provinsi. Pada tahun 2015 Kota Pontianak memiliki angka IPM tertinggi yaitu sebesar 77,5% sedangkan untuk perolehan IPM yang terendah terdapat pada Kabupaten Kayong Utara dengan angka IPM sebesar 60,09%.

Hasil estimasi *common effect model* menggunakan *software* R didapat model regresi sebagai berikut:

$$\hat{y}_{it} = 33,4745 + 0,1637x_{1it} + 0,0019x_{2it} - 0,0177x_{3it} + 0,0015x_{4it}$$

Sedangkan untuk estimasi *fixed effect model* didapat model regresi sebagai berikut:

$$\hat{y}_{it} = 0,061x_{1it} + 0,0034x_{2it} - 0,0304x_{3it} + 0,0039x_{4it}$$

Kemudian dilakukan uji Chow dan uji Hausman yang dapat dilihat pada Tabel 1.

Tabel 1 Uji Chow dan uji Hausman

Uji Chow		Uji Hausman	
Hasil	Kesimpulan	Hasil	Kesimpulan
F_0 = 56,9461 F_{tabel} (α;k-1;n-k) F_{tabel}(0,05;4-1;70-4)=2,74	Hipotesis nol ditolak sehingga model FEM yang akan digunakan	*p-value* = 0,0006	Hipotesis nol ditolak sehingga model FEM yang akan digunakan

Setelah terpilih model FEM dari uji *Chow* dan uji *Hausmann*, selanjutnya akan dilakukan uji asumsi klasik untuk menjaga akurasi dari model yang terbentuk. Beberapa uji asumsi klasik yang dilakukan yaitu uji autokorelasi, normalitas, heteroskedastisitas, dan multikoliniearitas. Berdasarkan uji asumsi klasik diperoleh kesimpulan bahwa residual tidak terjadi multikoliniearitas, residual berdistribusi normal, tidak terjadi autokorelasi pada residual dan residual tidak konstan atau terjadi heteroskedastisitas. Berdasarkan hasil keputusan uji heteroskedastisitas di atas menunjukkan bahwa adanya indikasi keragaman varians antar pengamatan. Masalah varians yang beragam tersebut yang akan diatasi dengan membuat pemodelan secara lokal yang mempertimbangkan adanya aspek spasial yaitu keragaman antar lokasi pengamatan.

Langkah selanjutnya adalah membuat model *fixed effect* GWPR untuk indeks pembangunan manusia di Kalimantan Barat. Langkah pertama yaitu menghitung jarak *euclidean* antar lokasi observasi berdasarkan koordinat *lattitude* dan *longitude*. Setelah itu dilakukan pemilihan *bandwidth* optimum yang dapat dilakukan dengan melihat nilai *cross validation* (CV) yang paling minimum diantara fungsi pembobot. Fungsi pembobot yang digunakan adalah fungsi *kernel fixed gaussian, bisquare, dan tricube*. Perbandingan dari ketiga pembobot dapat dilihat pada tabel berikut.

Tabel 2 Nilai *cross validation* (CV) dan *bandwidth* pada fungsi pembobot

Fungsi Pembobot	CV	*Bandwidth*
Gaussian	64,90764	0,56056
Bisqure	73,05880	1,55888
Tricube	77,19237	1,59126

Berdasarkan Tabel 2 maka fungsi pembobot yang digunakan adalah pembobot *Gaussian* karena memiliki CV paling minimum. Kemudian langkah selanjutnya membuat matriks pembobot *gaussian* dengan cara melakukan perkalian *kronecker* dua matrik yaitu matriks pembobot untuk masing-masing lokasi yang berukuran 14x14 terhadap matriks identitas berukuran 5x5 sehingga didapatkan matriks berukuran 70x70. Selanjutnya dilakukan estimasi parameter GWPR di tiap lokasi dan waktu pengamatan. Model yang terbentuk akan berbeda pada setiap lokasi. Berikut salah satu model *fixed effect* GWPR yang terbentuk pada lokasi pengamatan Kabupaten Sambas.

$$\hat{y}_{1t} = 36,4417 + 0,1587 x_{1t1} + 0,0017 x_{1t2} - 0,0014 x_{1t3} - 0,0112 x_{1t4}$$

Uji kecocokan model dilakukan dengan membandingkan nilai hasil estimasi terhadap nilai data yang sebenarnya menggunakan nilai *mean absolut percentage error* (MAPE). Berdasarkan hasil perhitungan didapatkan nilai MAPE sebesar 1,40% , hal tersebut menyatakan bahwa model yang digunakan sangat baik karena nilai MAPE < 10%. Selanjutnya dilakukan uji signifikansi parameter untuk mengetahui variabel prediktor yang berpengaruh signifikan pada angka indeks pembangunan manusia di setiap Kabupaten/Kota di Kalimantan Barat. Pengujian signifikansi parameter GWPR diperoleh dari penjelasan bahwa jika nilai *p-value* < $\alpha(0,05)$ maka parameter variabel prediktor ke-*k* berpengaruh signifikan untuk lokasi ke-***i***. Hasil pengujian signifikansi parameter model GWPR untuk Kabupaten/Kota di Kalimantan Barat sebagai berikut.

Tabel 3 Signifikansi parameter model GWPR untuk Kabupaten/Kota di Kalimantan Barat

Kabupaten/Kota	Parameter Signifikan
Kabupaten Sambas, Bengkayang, Ketapang, Kayong Utara, dan Kota Singkawang	X_2(pengeluaran perkapita) dan X_4(jumlah sarana kesehatan)
Kabupaten Landak, Mempawah, Sanggau, Sintang, Kapuas Hulu, Sekadau, Melawi, Kubu Raya, dan Kota Pontianak	X_2(pengeluaran perkapita)

KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan terhadap pemodelan angka indeks pembangunan manusia di setiap Kabupaten/Kota Provinsi Kalimantan Barat, maka dapat disimpulkan bahwa pengeluaran perkapita dan jumlah sarana kesehatan menjadi variabel yang paling berpengaruh dibandingkan kedua variabel lainnya yaitu tingkat pengangguran terbuka dan angka partisipasi sekolah tingkat SMP/MTS. Kabupaten/Kota di Provinsi Kalimantan Barat terbagi menjadi dua kelompok berdasarkan variabel yang signifikan. Kelompok pertama dengan parameter yang signifikan yaitu pengeluaran perkapita dan jumlah sarana kesehatan yang terdiri dari Kabupaten Sambas, Kabupaten Bengkayang, Kabupaten Ketapang, Kabupaten Kayong Utara, dan kota Singkawang. Kelompok kedua dengan parameter yang signifikan yaitu pengeluaran perkapita yang terdiri dari Kabupaten Landak, Kabupaten Mempawah, Kabupaten Sanggau, Kabupaten Sintang, Kabupaten Kapuas Hulu, Kabupaten Sekadau, Kabupaten Melawi, Kabupaten Kubu Raya, dan Kota Pontianak. Hasil estimasi menggunakan model GWPR memiliki nilai MAPE sebesar 1,40%.

DAFTAR PUSTAKA

[1]. Fotheringham, A.S., Brunsdon, C., dan Charlton, M., 2002, *Geographically Weighted Regression*, John Wiley and Sons, Chichester, UK.

[2]. Yamin, S., dan Rachmach., 2011, *Regresi dan Korelasi dalam Genggaman Anda: Aplikasi dengan Software SPSS, Eviews, MINI_TAB, dan STATGRAPHICS*, Salemba Empat, Jakarta.

[3]. Chasco, C., Gracia, I., dan Vicens, J., 2007, Modeling Spatial Variations in Household Disposable Income with *Geographically Weighted Regression, Munich Personal RePEc Archice Paper*, No. 1682.

[4]. Hsiao, C., 2003, *Analysis of Panel Data Second Edition*, Cambridge University Press, New York.

[5]. Baltagi, B., 2005, *Econometric Analysis of Panel Regression*, John and Wiley Ltd, New York.

[6]. Greene, W.H., 2000, *Econometric Analysis*, Prentice-Hall Inc, New Jersey.

[7]. Melliana, A., dan Zain, I., 2013, Analisis Statistika Faktor Yang Mempengaruhi Indek Pembangunan Manusia di Kabupaten/Kota Provinsi Jawa Timur dengan Menggunakan Regresi Panel, *Jurnal Sains dan Seni POMITS*, Volume 2 No. 2: 237-242.

[8]. Greene, W.H., 2003, *Econometric Analysis*, Ed ke-5, Prentice Hall International, New Jersey.

[9]. Rohmana, Y., 2010, *Ekonometrika Teori dan Aplikasi dengan Eviews*, Laboratorium Pendidikan Ekonomi dan Koperasi, Bandung.

[10]. Sudjana, 1996, *Metode Statistika*, Ed ke-6, Tarsito, Bandung.

[11]. Widarjono, A., 2007, *Ekonometrika Teori dan Aplikasi*, Ekonosia, Yogyakarta.

[12]. Yu, D., 2010, Exploring Spatiotemporally Varying Regressed Relationship: The Geographically Weighted Panel Regression Analysis, *The International Archives of the Photogrammety, Remote Sensing and Spatial Information Science*, 134-139.

[13]. Leung, Y., C.L., dan Zhang, W., 2000, Statistical Tests for Spatial Non-Stationarity Based on the Geographically Weighted Regression Model, *Environment and Planning*, A. 32 9-32.

[14]. Qur'ani, A.Y., 2014, Pemodelan Geographically Weighted Panel (GWR-Panel) Sebagai Pendekatan Model Geographically Weighted Regression (GWR) Dengan Menggunakan Fixed Effect Model Time Trend, *Jurnal Mahasiswa statistic*, Vol. 2,3.

SUTRO : Jurusan Matematika FMIPA Untan Pontianak, sutro.kardiansyah@student.untan.ac.id

YUNDARI : Jurusan Matematika FMIPA Untan Pontianak, yundari@math.untan.ac.id

SHANTIKA MARTHA : Jurusan Matematika FMIPA Untan Pontianak, shantika.martha@math.untan.ac.id