

Model Regresi Nonparametrik Multirespon *Spline Truncated* untuk Data Longitudinal

Dita Amelia dan I Nyoman Budiantara

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember

Jl. Arief Rahman Hakim, Surabaya 60111, Indonesia

ditaamelia91@gmail.com ; i_nyoman_b@statistika.its.ac.id

Abstrak— Model regresi nonparametrik dan semiparametrik yang banyak digunakan dalam dasawarsa terakhir adalah regresi spline. Dalam kehidupan sehari-hari, untuk mendapatkan kesimpulan yang menyeluruh dan utuh terhadap suatu permasalahan maka seringkali harus melibatkan lebih dari satu variabel respon atau multirespon. Selain data *cross section*, analisis regresi dapat diterapkan pada data longitudinal. Dalam penerapannya dalam bidang sosial, studi data longitudinal juga digunakan dengan melibatkan subyek penelitian berupa wilayah. Beberapa permasalahan sosial yang banyak dibahas belakangan ini adalah mengenai pertumbuhan penduduk yang berusaha dikendalikan melalui program Keluarga Berencana (KB). Oleh karena itu, dilakukan analisis lebih lanjut untuk memperkirakan keberhasilan KB melalui indikator persentase CPR dan persentase *unmet need*. Estimator model didapatkan dengan metode WLS dan penentuan knot optimum menggunakan metode GCV. Dari ketiga kelompok provinsi didapatkan hasil knot optimum yang digunakan yaitu sebanyak tiga buah knot dengan nilai GCV $4,06 \times 10^{-26}$ pada kelompok Jawa Bali, kelompok Luar Jawa Bali I $2,98 \times 10^{-27}$, dan kelompok Luar Jawa Bali II yaitu $1,18 \times 10^{-27}$.

Kata Kunci: *Spline, Multirespon, Data Longitudinal, GCV, Keluarga Berencana*

I. PENDAHULUAN

Terdapat tiga jenis regresi yang dikembangkan oleh para peneliti yaitu regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Umumnya dalam analisis regresi, sebelum melakukan pemodelan menggunakan salah satu dari ketiga jenis regresi, didahului dengan memeriksa *scatter plot* antara masing-masing variabel prediktor dengan variabel respon.

Beberapa penelitian dengan pendekatan semiparametrik dan nonparametrik diantaranya dilakukan oleh [1] dengan pendekatan regresi semiparametrik marginal longitudinal. Kemudian [2] melakukan pendekatan dengan *generalized nonparametric regression*. Sedangkan [3] menggunakan pemodelan berbasis polinomial lokal. Model regresi nonparametrik

dan semiparametrik yang banyak digunakan dalam dasawarsa terakhir adalah regresi spline. Spline memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu. Metode regresi nonparametrik spline banyak digunakan oleh beberapa peneliti diantaranya [4], [5] dengan menggunakan pendekatan *smoothing spline*. Dalam banyak kasus, untuk mendapatkan kesimpulan yang menyeluruh dan utuh terhadap suatu permasalahan maka seringkali harus melibatkan lebih dari satu variabel respon atau disebut multirespon. Adapun beberapa penelitian menggunakan pendekatan birespon ataupun multirespon diantaranya oleh [6] dan [7].

Selain data *cross section*, analisis regresi juga dapat diterapkan pada data longitudinal. Berdasarkan [8], diketahui bahwa dengan menggunakan studi data longitudinal dapat mengetahui perubahan individu, dibutuhkan subyek yang tidak terlalu banyak karena pengamatannya berulang, dan estimasinya lebih efisien karena dilakukan setiap pengamatan. Beberapa studi dengan data longitudinal dilakukan oleh [9] dengan regresi linear. Sementara itu, [10] menggunakan *generalized linear models* (GLM), [11] dengan menggunakan metode bayesian. Sementara itu, [12] menggunakan metode lokal polinomial *smoothing*. Studi longitudinal dengan pendekatan regresi nonparametrik juga telah dikembangkan oleh beberapa peneliti diantaranya [13], [14], [15], dan [16]. Sebagai pengembangan studi data longitudinal dengan pendekatan nonparametrik dengan melibatkan lebih dari satu variabel respon, [17] melakukan pemodelan regresi nonparametrik polinomial lokal birespon pada data longitudinal.

Dalam penerapannya dalam bidang sosial, studi data longitudinal juga digunakan dengan melibatkan subyek penelitian berupa wilayah seperti yang dilakukan oleh [18]. Beberapa permasalahan sosial yang banyak dibahas belakangan ini adalah mengenai pertumbuhan penduduk Indonesia yang terus mengalami peningkatan dan berusaha dikendalikan melalui program Keluarga Berencana (KB). Oleh karena itu penelitian ini bertujuan untuk menentukan estimasi model regresi nonparametrik multirespon *spline truncated* dan mengaplikasikannya pada data keberhasilan KB.

II. TINJAUAN PUSTAKA

A. Spline dalam Regresi Nonparametrik

Spline truncated mempunyai fungsi yang didefinisikan sebagai berikut

$$f(x) = \sum_{j=1}^m \alpha_j x^{j-1} + \sum_{j=1}^k \delta_j (x - K_j)_+^{m-1} \quad (1)$$

f merupakan fungsi spline orde m dengan knot K_1, \dots, K_k . Diberikan sebuah $\lambda = \{K_1, \dots, K_k\}$, sehingga f dapat diestimasi dengan mengestimasi koefisien pada persamaan (1). Salah satu metode yang dapat digunakan adalah metode *least squares*.

$$t_j(x) = x^{j-1}, j = 1, \dots, m,$$

$$t_{m+j}(x) = (x - K_j)_+^{m-1}, j = 1, \dots, k,$$

$$\text{dan } \beta = (\alpha_1, \dots, \alpha_m, \delta_1, \dots, \delta_k)^T$$

Least squares estimator spline, f diberikan oleh

$$f_\lambda = \sum_{j=1}^{m+k} b_{\lambda j} t_j, \quad (2)$$

dimana $b_\lambda = (b_{\lambda 1}, \dots, b_{\lambda(m+k)})^T$ adalah estimator dari β yang diperoleh dengan meminimumkan *sum of squares* residual. Jika \mathbf{y} adalah vektor variabel respon dan \mathbf{X}_λ mempunyai rank $m+k$, sehingga dapat dituliskan,

$$\mathbf{b}_\lambda = (\mathbf{X}_\lambda^T \mathbf{X}_\lambda)^{-1} \mathbf{X}_\lambda^T \mathbf{y} \quad (3)$$

dimana $\mathbf{X}_\lambda = \{t_j(x_i)\}_{i=1, n, j=1, m+k}$

Berdasarkan persamaan (2) dan (3), f_λ adalah estimator linear dari f , sedangkan λ adalah titik-titik knot.

Pemilihan λ ini bergantung kepada dua hal yaitu banyaknya knot dan letak titik knot pada fungsi spline. Apabila banyaknya knot K meningkat maka akan menghasilkan estimator yang semakin fleksibel karena penempatan knot di sebuah area akan membuat estimator beradaptasi secara lebih dengan data di area tersebut seperti yang diungkapkan oleh [19].

B. Regresi Nonparametrik Spline Truncated untuk Data Longitudinal

Pendapat dari [20] menjelaskan bahwa dalam *spline truncated* digunakan *truncated power basis* dimana $q = 1, 2, \dots, Q$ menunjukkan derajat polinomial dari *truncated power basis*, untuk $q = 1, 2, 3$ berturut-turut merupakan *truncated power basis* linear, kuadratik, dan kubik.

Pada data longitudinal apabila terdapat $j = 1, 2, \dots, m$ subyek dan $i = 1, 2, \dots, n$ pengamatan dalam setiap subyek dengan satu variabel prediktor x maka fungsi spline dapat didefinisikan sebagai fungsi f merupakan polinomial derajat Q dengan R knot. Model regresi nonparametrik *spline truncated* dirumuskan sebagai berikut [21].

$$y_{ji} = f(x_{ji}) + \varepsilon_{ji}, \quad j = 1, 2, \dots, m, \quad i = 1, 2, \dots, n$$

(4)

dimana

$$f(x_{ji}) = \sum_{q=0}^Q \alpha_{jq} x_{ji}^q + \sum_{r=1}^R \delta_{jr} (x_{ji} - K_{jr})_+^q \quad (5)$$

$$(x_{ji} - K_{jr})_+^q = \begin{cases} (x_{ji} - K_{jr})^q, & x_{ji} \geq K_{jr} \\ 0, & x_{ji} < K_{jr} \end{cases} \quad (6)$$

Persamaan (5) dapat dinyatakan dalam bentuk matriks yang dapat dikonstruksi dari vektor dan matriks berikut :

$$\mathbf{y} = \mathbf{XB} + \boldsymbol{\varepsilon} \quad (7)$$

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1(\mathbf{K}_1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2(\mathbf{K}_2) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_m(\mathbf{K}_m) \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_m \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{bmatrix}$$

dengan vektor respon :

$$\mathbf{y}_1 = (y_{11} \quad y_{12} \quad \dots \quad y_{1n})^T$$

$$\mathbf{y}_2 = (y_{21} \quad y_{22} \quad \dots \quad y_{2n})^T$$

\vdots

$$\mathbf{y}_m = (y_{m1} \quad y_{m2} \quad \dots \quad y_{mn})^T$$

dengan vektor parameter

$$\mathbf{B}_1 = (\alpha_{10} \quad \alpha_{11} \quad \alpha_{12} \quad \dots \quad \alpha_{1Q} \quad \delta_{11} \quad \delta_{12} \quad \dots \quad \delta_{1R})^T$$

$$\mathbf{B}_2 = (\alpha_{20} \quad \alpha_{21} \quad \alpha_{22} \quad \dots \quad \alpha_{2Q} \quad \delta_{21} \quad \delta_{22} \quad \dots \quad \delta_{2R})^T$$

\vdots

$$\mathbf{B}_m = (\alpha_{m0} \quad \alpha_{m1} \quad \alpha_{m2} \quad \dots \quad \alpha_{mQ} \quad \delta_{m1} \quad \delta_{m2} \quad \dots \quad \delta_{mR})^T$$

dan vektor error,

$$\boldsymbol{\varepsilon}_1 = (\varepsilon_{11} \quad \varepsilon_{12} \quad \dots \quad \varepsilon_{1n})^T$$

$$\boldsymbol{\varepsilon}_2 = (\varepsilon_{21} \quad \varepsilon_{22} \quad \dots \quad \varepsilon_{2n})^T$$

\vdots

$$\boldsymbol{\varepsilon}_m = (\varepsilon_{m1} \quad \varepsilon_{m2} \quad \dots \quad \varepsilon_{mn})^T$$

C. Regresi Nonparametrik Multirespon Spline Truncated untuk Data Longitudinal

Dalam permasalahan nyata, seringkali peneliti dihadapkan pada dua atau lebih variabel dependen yang diamati dari beberapa variabel independen. Model regresi nonparametrik multirespon merupakan metode yang tepat untuk memodelkan fungsi dari variabel-variabel tersebut sebagaimana yang diungkapkan [6]. Diberikan data longitudinal berpasangan multirespon dengan l buah variabel respon dan p buah variabel prediktor. Model regresi nonparametrik multirespon untuk data longitudinal dapat dinyatakan sebagai berikut.

$$y_{1ji} = f(x_{11ji}) + f(x_{21ji}) + \dots + f(x_{p1ji}) + \varepsilon_{1ji}$$

$$y_{2ji} = f(x_{12ji}) + f(x_{22ji}) + \dots + f(x_{p2ji}) + \varepsilon_{2ji}$$

\vdots

$$y_{lji} = f(x_{1lji}) + f(x_{2lji}) + \dots + f(x_{plji}) + \varepsilon_{lji}$$

dimana $j = 1, 2, \dots, m$; $i = 1, 2, \dots, n$.

dimana $f(x)$ diasumsikan bentuk pola kurva regresinya tidak diketahui sehingga didekati dengan fungsi *spline truncated* berikut

$$f(x_{skji}) = \sum_{q=0}^Q \alpha_{kjsq} x_{skji}^q + \sum_{r=1}^R \delta_{kjsr} (x_{skji} - K_{kjsr})_+^p \quad (8)$$

$$\text{dimana } (x_{skji} - K_{kjsr})_+^p = \begin{cases} (x_{skji} - K_{kjsr})^p, & x_{skji} \geq K_{kjsr} \\ 0, & x_{skji} < K_{kjsr} \end{cases}$$

Kurva regresi $f(x)$ merupakan polinomial derajat Q dengan R knot, $s = 1, 2, \dots, p$ dan K_{kjsr} adalah titik knot.

D. Pemilihan Titik Knot Optimal

Spline merupakan potongan polinomial yang memuat titik-titik knot. Oleh karena itu letak dan banyaknya titik knot merupakan hal penting dalam pemodelan regresi nonparametrik dengan pendekatan *spline truncated*. Menurut [22], salah satu metode yang digunakan untuk menentukan knot yang optimal adalah dengan metode *Generalized Cross Validation* (GCV).

$$GCV(h) = \frac{MSE(h)}{\left[\frac{1}{lmn} \text{trace}(\mathbf{I} - \mathbf{A}(h)) \right]^2} \quad (9)$$

$$\text{dengan } MSE(h) = \frac{1}{lmn} \sum_{k=1}^l \sum_{j=1}^m \sum_{i=1}^n (y_{kji} - \hat{y}_{kji})^2, \text{ dan } \mathbf{A}(h)$$

diperoleh dari hubungan $\hat{\mathbf{y}} = \mathbf{A}(h)\mathbf{y}$. Knot optimal diberikan oleh nilai GCV terkecil.

III. METODOLOGI

Penelitian ini menggunakan lebih dari satu variabel respon dan lebih dari satu variabel prediktor. Adapun langkah-langkah penyelesaiannya adalah:

1. Membuat model regresi nonparametrik multirespon data longitudinal
2. Menyajikan model regresi tersebut dalam bentuk matriks
3. Mendekati kurva regresi nonparametrik tersebut dengan fungsi *spline truncated* polinomial Q derajat dengan R knot
4. Menentukan matriks bobot
5. Mendapatkan estimator dengan meminimumkan fungsi WLS
6. Menentukan titik knot optimal dengan metode GCV
7. Mencari estimasi model berdasarkan data keberhasilan KB
8. Memodelkan data keberhasilan KB untuk setiap kelompok provinsi.

Aplikasi pada data keberhasilan KB dilakukan di tiga kelompok provinsi yaitu Jawa Bali terdiri dari 7 provinsi, Luar Jawa Bali I terdiri dari 13 provinsi, dan Luar Jawa Bali II juga terdiri dari 13 provinsi. Variabel respon yang digunakan adalah persentase CPR dan persentase *unmet need*. Sedangkan variabel prediktor yang digunakan adalah indeks kedalaman kemiskinan,

persentase KK dengan pendidikan \leq SLTP, persentase wanita berumur 10 tahun ke atas dengan usia perkawinan pertama \leq 18 tahun, persentase wanita berumur 10 tahun ke atas yang pernah kawin dengan \leq 2 anak lahir hidup.

IV. ANALISIS DAN PEMBAHASAN

A. Estimasi Model Regresi Nonparametrik Multirespon Spline Truncated untuk Data Longitudinal

Diberikan data longitudinal berpasangan multirespon dengan l buah variabel respon dan p buah variabel prediktor $(y_1, y_2, \dots, y_l, x_1, x_2, \dots, x_p)$. Model regresi nonparametrik multirespon untuk data longitudinal dapat dinyatakan sebagai berikut.

$$y_{1ji} = f(x_{11ji}) + f(x_{21ji}) + \dots + f(x_{p1ji}) + \varepsilon_{1ji}$$

$$y_{2ji} = f(x_{12ji}) + f(x_{22ji}) + \dots + f(x_{p2ji}) + \varepsilon_{2ji}$$

⋮

$$y_{lji} = f(x_{1lji}) + f(x_{2lji}) + \dots + f(x_{plji}) + \varepsilon_{lji}$$

dimana $j = 1, 2, \dots, m$; $i = 1, 2, \dots, n$.

Persamaan di atas dapat ditulis dalam bentuk matriks

$$\begin{bmatrix} \mathbf{y}_1 \\ \dots \\ \mathbf{y}_2 \\ \dots \\ \vdots \\ \dots \\ \mathbf{y}_l \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1 \\ \dots \\ \mathbf{f}_2 \\ \dots \\ \vdots \\ \dots \\ \mathbf{f}_l \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \dots \\ \boldsymbol{\varepsilon}_2 \\ \dots \\ \vdots \\ \dots \\ \boldsymbol{\varepsilon}_l \end{bmatrix} \quad (10)$$

dimana $f(x)$ didekati dengan fungsi *spline truncated* berikut

$$f(x_{skji}) = \sum_{q=0}^Q \alpha_{kjsq} x_{skji}^q + \sum_{r=1}^R \delta_{kjsr} (x_{skji} - K_{kjsr})_+^p$$

Apabila ditulis dalam bentuk matriks, diperoleh $\mathbf{y} = \mathbf{A}_1 + \mathbf{A}_2 + \dots + \mathbf{A}_p + \boldsymbol{\varepsilon}$ (11)

Dengan memisahkan vektor parameternya maka persamaan (11) dapat ditulis menjadi :

$$\mathbf{y} = \mathbf{B}_1\boldsymbol{\gamma}_1 + \mathbf{B}_2\boldsymbol{\gamma}_2 + \dots + \mathbf{B}_p\boldsymbol{\gamma}_p \quad (12)$$

Estimasi $\hat{\mathbf{y}}$ diberikan oleh :

$$\hat{\mathbf{y}} = \mathbf{B}_1\hat{\boldsymbol{\gamma}}_1 + \mathbf{B}_2\hat{\boldsymbol{\gamma}}_2 + \dots + \mathbf{B}_p\hat{\boldsymbol{\gamma}}_p \quad (13)$$

Untuk memperoleh estimator $\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \dots, \hat{\boldsymbol{\gamma}}_p$ digunakan metode *Weighted Least Square* (WLS) dengan menggunakan matrik \mathbf{V} sebagai matriks pembobot.

$$\boldsymbol{\psi}(\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \dots, \hat{\boldsymbol{\gamma}}_p) = (\mathbf{y} - \hat{\mathbf{y}})^T \mathbf{V} (\mathbf{y} - \hat{\mathbf{y}})$$

$$= \mathbf{y}^T \mathbf{V} \mathbf{y} - 2(\mathbf{B}_1\boldsymbol{\gamma}_1 + \mathbf{B}_2\boldsymbol{\gamma}_2 + \dots + \mathbf{B}_p\boldsymbol{\gamma}_p)^T \mathbf{V} \mathbf{y} + (\mathbf{B}_1\boldsymbol{\gamma}_1 + \mathbf{B}_2\boldsymbol{\gamma}_2 + \dots + \mathbf{B}_p\boldsymbol{\gamma}_p)^T \mathbf{V} (\mathbf{B}_1\boldsymbol{\gamma}_1 + \mathbf{B}_2\boldsymbol{\gamma}_2 + \dots + \mathbf{B}_p\boldsymbol{\gamma}_p)$$

Misal $\mathbf{B}_1\boldsymbol{\gamma}_1 + \mathbf{B}_2\boldsymbol{\gamma}_2 + \dots + \mathbf{B}_p\boldsymbol{\gamma}_p = \mathbf{X}\boldsymbol{\theta}$ maka persamaan di atas menjadi

$$\boldsymbol{\psi}(\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \dots, \hat{\boldsymbol{\gamma}}_p) = \mathbf{y}^T \mathbf{V} \mathbf{y} - 2(\mathbf{X}\boldsymbol{\theta})^T \mathbf{V} \mathbf{y} + (\mathbf{X}\boldsymbol{\theta})^T \mathbf{V} (\mathbf{X}\boldsymbol{\theta}) = \mathbf{y}^T \mathbf{V} \mathbf{y} - 2\boldsymbol{\theta}^T \mathbf{X}^T \mathbf{V} \mathbf{y} + \boldsymbol{\theta}^T \mathbf{X}^T \mathbf{V} \mathbf{X} \boldsymbol{\theta}$$

Untuk mendapatkan estimasi θ dilakukan dengan meminimumkan fungsi WLS sehingga diperoleh

$$\hat{\theta} = (\mathbf{X}^T \mathbf{VX})^{-1} \mathbf{X}^T \mathbf{Vy} \quad (14)$$

Didefinisikan \mathbf{X} adalah

$$\mathbf{X} = [\text{diag}(\mathbf{M}_{11}, \mathbf{M}_{12}, \dots, \mathbf{M}_{1m}, \mathbf{M}_{21}, \mathbf{M}_{22}, \dots, \mathbf{M}_{2m}, \dots, \mathbf{M}_{lm})]$$

$$\text{dengan } \mathbf{M}_{lm} = [\mathbf{B}_{1/m} \quad \mathbf{B}_{2/m} \quad \mathbf{B}_{3/m} \quad \dots \quad \mathbf{B}_{p/m}]$$

Sedangkan nilai θ adalah

$$\theta = [\theta_{11}^T, \theta_{12}^T, \dots, \theta_{1m}^T, \theta_{21}^T, \theta_{22}^T, \dots, \theta_{2m}^T, \dots, \theta_{lm}^T]^T$$

$$\text{dengan } \theta_{lm} = [\gamma_{1/m}^T, \gamma_{2/m}^T, \gamma_{3/m}^T, \dots, \gamma_{p/m}^T]^T$$

sehingga \hat{y} dapat ditulis :

$$\hat{y} = \mathbf{X}\hat{\theta} \quad (15)$$

Apabila persamaan (14) disubstitusikan ke persamaan (15) maka didapatkan \hat{y} sebagai berikut.

$$\hat{y} = \mathbf{X}\hat{\theta} \quad (16)$$

$$\hat{y} = \mathbf{X}(\mathbf{X}^T \mathbf{VX})^{-1} \mathbf{X}^T \mathbf{Vy}$$

$$= \mathbf{A}(\mathbf{h})\mathbf{y}$$

$$\text{dimana } \mathbf{A}(\mathbf{h})\mathbf{y} = \mathbf{X}(\mathbf{X}^T \mathbf{VX})^{-1} \mathbf{X}^T \mathbf{V}$$

B. Sifat-Sifat Estimator Regresi Nonparametrik Multirespon Spline Truncated untuk Data Longitudinal

B.1 Linear

$$c_1 = \mathbf{f}_1 = (f(x_{11}), f(x_{21}), \dots, f(x_{p1}))$$

$$c_2 = \mathbf{f}_2 = (f(x_{12}), f(x_{22}), \dots, f(x_{p2}))$$

⋮

$$c_l = \mathbf{f}_l = (f(x_{1l}), f(x_{2l}), \dots, f(x_{pl}))$$

$$\hat{z}(c_1, c_2, \dots, c_l) = \mathbf{X}\hat{\theta} \quad (17)$$

$$= \mathbf{X}(\mathbf{X}^T \mathbf{VX})^{-1} \mathbf{X}^T \mathbf{Vy}$$

$$= \mathbf{A}(\mathbf{h})\mathbf{y}$$

Berdasarkan persamaan (17) terlihat bahwa estimator *spline truncated* multirespon merupakan estimator yang linear.

B.2 Tak Bias

$$E(\hat{z}(c_1, c_2, \dots, c_l)) = E \begin{bmatrix} \hat{z}(c_1) \\ \hat{z}(c_2) \\ \vdots \\ \hat{z}(c_l) \end{bmatrix} = E[\mathbf{A}(\mathbf{h})\mathbf{y}] = E \left[\mathbf{A}(\mathbf{h}) \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_l \end{bmatrix} \right] = \mathbf{A}(\mathbf{h}) \begin{bmatrix} E(\mathbf{y}_1) \\ E(\mathbf{y}_2) \\ \vdots \\ E(\mathbf{y}_l) \end{bmatrix}$$

dimana $E(\mathbf{y}_i) = E(z(c_i) + \varepsilon_i) = z(c_i)$, sehingga

$$E(\hat{z}(c_1, c_2, \dots, c_l)) = \mathbf{A}(\mathbf{h}) \begin{bmatrix} z(c_1) \\ z(c_2) \\ \vdots \\ z(c_l) \end{bmatrix}$$

$\hat{z}(c_1, c_2, \dots, c_l)$ bersifat bias karena

$$E(\hat{z}(c_1, c_2, \dots, c_l)) \neq \begin{bmatrix} z(c_1) \\ z(c_2) \\ \vdots \\ z(c_l) \end{bmatrix}$$

B.3 Berdistribusi Normal

Estimator $\hat{z}(c_1, c_2, \dots, c_l)$ berdistribusi normal dengan mean $\mathbf{A}(\mathbf{h})^T \mathbf{X}\theta$ dan variansi $\sigma^2 \mathbf{A}(\mathbf{h})^T \mathbf{A}(\mathbf{h})$.

C. Aplikasi Model Regresi Nonparametrik Multirespon Spline Truncated untuk Data Longitudinal

Matriks pembobot yang digunakan adalah banyaknya pengamatan dalam setiap subyek (\mathbf{V}).

$$\mathbf{V} = \left[\text{diag} \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n} \right) \right]$$

C.1 Model Regresi Nonparametrik Multirespon Spline Truncated untuk Data Longitudinal Pada Kelompok Jawa Bali

Berikut adalah titik-titik knot beserta nilai GCV untuk setiap variabel pada setiap subyek dengan menggunakan pembobot \mathbf{V} .

Tabel 1. Nilai GCV Kelompok Jawa Bali

Banyak Titik Knot	GCV
1	$1,11 \times 10^{-24}$
2	$2,18 \times 10^{-25}$
3	$4,06 \times 10^{-26}$

Tabel 2. Ringkasan Knot Optimal Kelompok Jawa Bali

Provinsi	Knot			
	x_1	x_2	x_3	x_4
DKI Jakarta	0,463673	35,20837	24,82694	58,1498
	0,507755	36,14878	26,2302	59,10735
	0,595918	38,02959	29,03673	61,02245
Jawa Barat	1,645714	68,2149	52,29449	49,67673
	1,954286	69,23367	54,26837	50,95755
	2,571429	71,27122	58,21612	53,51918
Jawa Tengah	2,427959	76,72082	46,40041	50,03429
	2,883469	77,33061	47,60531	51,16571
	3,79449	78,5502	50,0151	53,42857
DI Yogyakarta	0,62898	62,79061	25,79551	56,69633
	1,336735	63,15796	26,82163	57,85224
	2,752245	63,89265	28,87388	60,16408
Jawa Timur	1,971429	72,76041	51,77449	57,82653
	2,468571	73,60531	53,02837	58,9849
	3,462857	75,2951	55,53612	61,30163
Bali	0,408367	56,96918	21,78571	56,37694
	0,628776	57,91939	22,45429	57,3002
	1,069592	59,8198	23,79143	59,14673
Banten	0,962245	62,24592	47,11	48,71755
	1,109184	63,27694	49,15	50,24816
	1,403061	65,33898	53,23	53,30939

Kelompok Jawa Bali terdiri dari tujuh provinsi atau tujuh buah subyek sehingga akan dihasilkan seluruh estimasi parameter untuk ketujuh provinsi dengan masing-masing provinsi terdiri dari model untuk respon pertama dan model untuk respon kedua. Model tersebut merupakan model terbaik dengan nilai MSE sebesar $2,494269 \times 10^{-20}$. Sebagai contoh, berikut adalah model Provinsi DKI Jakarta.

$$\begin{aligned} \hat{y}_{11i} = & 0,01735 - 0,17193 x_{11i} - 0,17666(x_{11i} - 0,463673)_+ \\ & - 0,13768(x_{11i} - 0,507755)_+ - 0,04576(x_{11i} - 0,595918)_+ \\ & + 0,20193 x_{21i} - 0,41271(x_{21i} - 35,20837)_+ - 0,47596(x_{21i} - 36,14878)_+ \\ & - 0,10223(x_{21i} - 38,02959)_+ + 0,55656 x_{31i} + 0,12007(x_{31i} - 24,82694)_+ \\ & + 0,02568(x_{31i} - 26,2302)_+ - 0,15254(x_{31i} - 29,03673)_+ + 0,91709 x_{41i} + \\ & - 0,10026(x_{41i} - 58,1498)_+ - 0,1687(x_{41i} - 59,10735)_+ + \\ & - 0,0478(x_{41i} - 61,02245)_+ \end{aligned}$$

$$\hat{y}_{21i} = 0,00225 + 0,10295 x_{11i} + 0,09899(x_{11i} - 0,463673)_+ + 0,0639(x_{11i} - 0,507755)_+ + 0,01811(x_{11i} - 0,595918)_+ + 0,19326 x_{21i} + 0,13063(x_{21i} - 35,20837)_+ + 0,32888(x_{21i} - 36,14878)_+ - 0,167(x_{21i} - 38,02959)_+ - 0,1098 x_{31i} - 0,1408(x_{31i} - 24,82694)_+ + 0,15503(x_{31i} - 26,2302)_+ - 0,2492(x_{31i} - 29,03673)_+ + 0,1675 x_{41i} + 0,02232(x_{41i} - 58,1498)_+ - 0,0913(x_{41i} - 59,10735)_+ + 0,20402(x_{41i} - 61,02245)_+$$

C.2 Model Regresi Nonparametrik Multirespon Spline Truncated untuk Data Longitudinal Pada Kelompok Luar Jawa Bali I

Berikut nilai GCV untuk berbagai jumlah knot di setiap variabel pada setiap subyek.

Tabel 3. Nilai GCV Kelompok Luar Jawa Bali I

Banyak Titik Knot	GCV
1	$3,13 \times 10^{-25}$
2	$3,13 \times 10^{-25}$
3	$1,11 \times 10^{-26}$

Sebagai contoh berikut adalah model yang dihasilkan untuk Provinsi Aceh.

$$\hat{y}_{11i} = 0,02148 - 0,07867 x_{11i} - 0,10304(x_{11i} - 3,156122)_+ - 0,06147(x_{11i} - 3,242245)_+ + 0,20653(x_{11i} - 4,103469)_+ + 0,4701 x_{21i} - 0,76108(x_{21i} - 63,03367)_+ - 0,64345(x_{21i} - 63,27735)_+ - 0,41607(x_{21i} - 65,71408)_+ + 0,57632 x_{31i} - 0,25296(x_{31i} - 37,48204)_+ - 0,27134(x_{31i} - 37,68408)_+ - 0,24907(x_{31i} - 39,70449)_+ + 0,45088 x_{41i} + 0,39979(x_{41i} - 40,37694)_+ - 0,38497(x_{41i} - 40,46388)_+ - 0,49828(x_{41i} - 41,33327)_+ \\ \hat{y}_{21i} = 0,00045 + 0,33805 x_{11i} + 0,32401(x_{11i} - 3,156122)_+ + 0,31136(x_{11i} - 3,242245)_+ + 0,36701(x_{11i} - 4,103469)_+ + 0,18277 x_{21i} + 0,11849(x_{21i} - 63,03367)_+ + 0,08271(x_{21i} - 63,27735)_+ - 0,1272(x_{21i} - 65,71408)_+ + 0,03111 x_{31i} + 0,00551(x_{31i} - 37,48204)_+ - 0,0061(x_{31i} - 37,68408)_+ + 0,04602(x_{31i} - 39,70449)_+ + 0,07654 x_{41i} + 0,08784(x_{41i} - 40,37694)_+ + 0,11744(x_{41i} - 40,46388)_+ + 0,12301(x_{41i} - 41,33327)_+$$

Model tersebut merupakan model terbaik dengan nilai MSE sebesar $2,175618 \times 10^{-20}$.

C.3 Model Regresi Nonparametrik Multirespon Spline Truncated untuk Data Longitudinal Pada Kelompok Luar Jawa Bali II

Sebagaimana yang dilakukan pada pembahasan kelompok Luar Jawa Bali I, kelompok Luar Jawa Bali II juga akan dihasilkan model untuk 13 buah provinsi dimana setiap provinsi terdiri dari model untuk respon pertama dan model untuk respon kedua. Model tersebut merupakan model terbaik dengan nilai MSE sebesar $8,605989 \times 10^{-21}$.

Tabel 4. Nilai GCV Kelompok Luar Jawa Bali II

Banyak Titik Knot	GCV
1	$1,75 \times 10^{-25}$
2	$6,23 \times 10^{-27}$
3	$1,18 \times 10^{-27}$

D. Pembahasan Model Regresi Nonparametrik Multirespon Spline Truncated untuk Data Longitudinal pada Data Keberhasilan KB

Interpretasi model terhadap variabel indeks kedalaman kemiskinan (x_{11}) dengan asumsi variabel lain konstan pada kelompok Jawa Bali adalah sebagai berikut

$$\hat{y}_{11} = \begin{cases} 0,01735 - 0,17193x_{11} & ; & x_{11} < 0,463673 \\ 0,09925 - 0,34859x_{11} & ; & 0,463673 \leq x_{11} < 0,507755 \\ 0,16916 - 0,48627x_{11} & ; & 0,507755 \leq x_{11} < 0,595918 \\ 0,19643 - 0,53203x_{11} & ; & x_{11} \geq 0,595918 \end{cases}$$

$$\hat{y}_{21} = \begin{cases} 0,00225 + 0,10295x_{11} & ; & x_{11} < 0,463673 \\ -0,04365 + 0,20194x_{11} & ; & 0,463673 \leq x_{11} < 0,507755 \\ -0,07609 + 0,26584x_{11} & ; & 0,507755 \leq x_{11} < 0,595918 \\ -0,08688 + 0,28395x_{11} & ; & x_{11} \geq 0,595918 \end{cases}$$

Dalam hal ini diketahui bahwa kemiskinan memberikan dampak bagi keberhasilan KB. Semakin besar indeks kemiskinan maka kemungkinan untuk berpartisipasi dalam program KB juga kecil sehingga dapat mengurangi persentase CPR dan menambah persentase *unmet need*.

Dalam analisis dengan menggunakan data longitudinal ini, dapat menjelaskan pengaruh yang lebih spesifik di setiap subyek. Efek yang ditimbulkan bisa saja sama namun juga bisa berbeda karena karakteristik subyek yang dalam hal ini adalah provinsi berbeda-beda. Berikut adalah gambaran mengenai pengaruh masing-masing variabel di setiap provinsi dalam kelompok Jawa Bali.



Gambar 1. Pengaruh x_1 Terhadap Respon di Kelompok Jawa Bali

Berdasarkan interpretasi model yang telah dibahas sebelumnya, terbentuk empat segmentasi model untuk tiap variabel respon. Area berwarna *orange* pada Gambar 1 menjelaskan bahwa untuk Provinsi DKI Jakarta, Jawa Barat, Jawa Tengah, dan Jawa Timur memiliki kesamaan dimana variabel indeks kedalaman kemiskinan (x_1) di model segmen satu hingga segmen empat memberikan efek negatif atau cenderung menurunkan persentase CPR dan memberikan efek positif atau cenderung menaikkan persentase *unmet need* apabila terjadi kenaikan variabel indeks kedalaman kemiskinan (x_1). Berikut adalah gambaran dari variabel prediktor lainnya.



Gambar 2. Pengaruh x_2 Terhadap Respon di Kelompok Jawa Bali



Gambar 3. Pengaruh X_3 Terhadap Respon di Kelompok Jawa Bali



Gambar 4. Pengaruh X_4 Terhadap Respon di Kelompok Jawa Bali

Dari Gambar 1 hingga Gambar 4 terlihat bahwa masing-masing variabel prediktor memberikan kecenderungan pengaruh yang berbeda-beda untuk setiap variabel respon di tiap provinsi.

V. KESIMPULAN DAN SARAN

Estimator model regresi nonparametrik multirespon *spline* truncated untuk data longitudinal diperoleh dengan pendekatan WLS dan didapatkan $\hat{\theta} = (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V} \mathbf{y}$. Sifat dari estimator adalah linier, bias, dan berdistribusi normal. Pada tahap aplikasi model, kelompok Jawa Bali, Luar Jawa Bali I, dan Luar Jawa Bali II menggunakan knot optimum sebanyak tiga knot dan memiliki nilai GCV serta MSE yang berbeda-beda. Penelitian lebih lanjut dapat dikembangkan utamanya dalam hal pengembangan *package* program sehingga memungkinkan knot optimal yang terpilih di setiap respon tidak selalu sama dan dengan jumlah pengamatan yang berbeda.

UCAPAN TERIMA KASIH

Pada penelitian ini, penulis ingin mengucapkan terimakasih kepada pihak Pasca Sarjana ITS. Tidak lupa pula BKKBN dan BPS yang telah membantu dalam hal penyediaan data.

DAFTAR PUSTAKA

- [1] M. Kadiri, R.J. Carrol, and M.P. Wand, "Marginal Longitudinal Semiparametric Regression via Penalized Splines", *Statistics and Probability Letters*, Vol. 80, hal 1242-1252, 2010.
- [2] D.D Cox, and F. O'Sullivan, "Penalized Likelihood-type Estimators for Generalized Nonparametric Regression", *Journal of Multivariate Analysis*, Vol. 56, No. 10, hal 185-206, 1996.
- [3] N. Chamidah, I.N. Budiantara, and I. Zain, "Designing of Child Growth Chart Based on Multi-Response Local

Polynomial Modeling", *Journal of Mathematics and Statistics*, Vol. 8, No. 3, hal 342-347, 2012.

- [4] A. Tripena, "Penentuan Model Regresi Spline Terbaik", *Prosiding Seminar Nasional Statistika Universitas Diponegoro*, hal 92-102, 2011.
- [5] X. Lin, N. Wang, A.H. Welsh, and R.J. Carrol, "Equivalent Kernel of Smoothing Splines in Nonparametric Regression for Clustered/Longitudinal Data", *Biometrika*, Vol. 91, No. 1, hal 177-193, 2004.
- [6] B. Lestari, I.N. Budiantara, S. Sunaryo, and M. Mashuri, "Spline Estimator in Multi-Response Nonparametric Regression Model with Unequal Correlation of Errors", *Journal of Mathematics and Statistics*, Vol. 6, No. 3, hal 327-332, 2010.
- [7] Y. Wang, W. Guo, and M.B. Brown, "Spline Smoothing for Bivariate Data With Applications to Association Between Hormones", *Statistica Sinica*, Vol. 10, hal 377-397, 2000.
- [8] H. Wu, and J.T. Zhang, *Nonparametric Regression Method for Longitudinal Data Analysis*, John Wiley & Sons, New Jersey, 2006.
- [9] F. Yao, H.G. Muller, and J.L. Wang, *Functional Linear Regression Analysis for Longitudinal Data*, University of California, Davis, 2004.
- [10] K.Y. Liang, and S.L. Zeger, "Longitudinal Data Analysis Using Generalized Linear Models", *Biometrika*, Vol. 73, No. 1, hal 13-22, 1986.
- [11] M. Menictas, and M.P. Wand, *Variational Inference for Marginal Longitudinal Semiparametric Regression*, University of Technology Sydney, Sydney, 2013.
- [12] J. Fan, and J.T. Zhang, "Two-step Estimation of Functional Linear Models With Applications to Longitudinal Data", *Royal Statistical Society*, Vol. 62, No. 2, hal 303-322, 2002.
- [13] M. Durban, J. Harezlak, M.P. Wand, and R.J. Carrol, "Simple Fitting of Subject-Specific Curves for Longitudinal Data", *Statistics in Medicine*, hal. 1-24, 2004.
- [14] N.A. Ibrahim, and Suliadi, "Analyzing Longitudinal Data using Gee-Smoothing Spline", *Proceedings of The 8th WSEAS International Conference on Applied Computer and Applied Computational Science*, hal 26-33, 2008.
- [15] H. Liang, and Y. Xiao, "Penalized Splines for Longitudinal Data With An Application in AIDS Studies", *Journal of Modern Applied Statistical Methods*, Vol. 5, No. 1, hal 130-139, 2006.
- [16] L. Laome, "Model Regresi Semiparametrik Spline untuk Data Longitudinal pada Kasus Kadar CD4 Penderita HIV", *Paradigma*, Vol. 13, No. 2, hal 189-194, 2009.
- [17] A. Prahutama, *Model Regresi Nonparametrik Polinomial Lokal Birespon pada Data Longitudinal*, Tesis Program Magister, Institut Teknologi Sepuluh Nopember, Surabaya, 2013.
- [18] A. Melliana, *Analisis Statistika Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Kabupaten/Kota Provinsi Jawa Timur dengan Menggunakan Regresi Panel*, Tugas Akhir Program Sarjana, Institut Teknologi Sepuluh Nopember, Surabaya, 2013.
- [19] R.L. Eubank, *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, Inc., New York, 1999.
- [20] K. Takezawa, *Introduction to Nonparametric Regression*, John Wiley & Sons, New Jersey, 2006.
- [21] I. Sriliana, *Regresi Spline Truncated dalam Model Linear Parsial untuk Data Longitudinal*, Tesis Program Magister, Institut Teknologi Sepuluh Nopember, Surabaya, 2012.
- [22] G. Wahba, *Spline Models for Observation Data*, University of Winsconsin at Madison, Pennsylvania, 1990.