

KAJIAN PEMODELAN SPLINE UNTUK DATA LONGITUDINAL SEBAGAI PERKEMBANGAN DARI REGRESI NONPARAMETRIK

Suparti¹⁾, Alan Prahutama²⁾, Rukun Santoso³⁾

^{1,2,3}Departemen Statistika, Fakultas Sains dan matematika, Universitas Diponegoro
email: suparti702@gmail.com

Abstract

Regression analysis can be approached by using parametric, semi-parametric and nonparametric regression approaches. One of nonparametric regression approach that great developed was Spline truncated, including for modeling longitudinal data. Longitudinal data is data that consisting of several subjects which is each subject is observed repeatedly based on a certain time. The advantages of longitudinal data has provided more complexcityof information than cross section and time series data. The spline approach was a segmented polynomial regression approach. Spline provides high flexibility due to the use of knot points. To determine the optimal knot points using Generalized Cross Validation (GCV). The principle of determining the optimum point of knot of longitudinal data using spline truncated is basically the same as with Spline method for cross section data, that is determination of knot point based on each subject. However, the estimation is done simultaneously so that each subject has its own model.

Keywords: *Spline Truncated, GCV, Knot points.*

1. PENDAHULUAN

Analisis Analisis regresi merupakan metode statistika yang digunakan untuk pemodelan antara variabel respon dengan variabel prediktor. Pendekatan regresi dapat dilakukan dengan tiga pendekatan antara lain pendekatan parametrik, semiparametrik, dan nonparametrik. Pendekatan parametrik dilakukan apabila bentuk kurva diketahui, sedangkan pendekatan nonparametrik dilakukan apabila bentuk kurva tidak diketahui. Pendekatan semiparametrik dilakukan apabila sebagian bentuk kurva diketahui dan sebagian tidak diketahui. Bentuk kurva yang dimaksud adalah kurva antara variabel respon dengan variabel prediktor. Pendekatan parametrik sering dilakukan, dikarenakan kemudahan dalam estimasinya. Akan tetapi pendekatan parametrik sangat terikat dengan asumsi-asumsi antara lain residual berdistribusi normal, homokedastisitas, nonautokorelasi serta antara variabel prediktor tidak terjadi multikolinieritas. Sehingga jika suatu permasalahan dimodelkan menggunakan pendekatan parametrik, akan tetapi model yang didapatkan asumsi-asumsinya tidak terpenuhi, tentunya model tersebut akan menjadi bias untuk digunakan.

Diantara model-model regresi nonparemetrik yang disebutkan di atas, spline merupakan model yang mempunyai interpretasi statistik dan interpretasi visual yang sangat khusus dan sangat baik (Budiantara, 2009). Spline adalah salah satu jenis potongan polinomial, yaitu polinomial yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih daripada polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal pada suatu fungsi atau data. Wahba (1990) menunjukkan bahwa spline memiliki sifat-sifat statistik yang berguna untuk menganalisis hubungan dalam regresi. Dalam analisis regresi spline jika terdapat satu variabel respon dan satu prediktor maka regresi ini disebut regresi spline univariabel. Sebaliknya, jika terdapat satu variabel respon dan lebih dari satu variabel prediktor, maka regresi tersebut dinamakan regresi spline multivariabel (Budiantara, 2004).

Dalam analisis regresi terdapat dua jenis data yaitu data longitudinal dan data *cross section*. Data longitudinal merupakan data yang didapat dari pengamatan berulang dari setiap subjek pada interval waktu yang berbeda. Data ini berkorelasi pada subjek yang sama dan independen antar subjek yang berbeda. Sedangkan data *cross section* merupakan data yang hanya dilakukan satu kali pengamatan pada setiap subjek dan saling independen (Wu dan Zhang, 2006). Menurut Wu dan Zhang (2006) terdapat beberapa keuntungan studi mengenai data longitudinal antara lain dapat mengetahui perubahan individu, dan membutuhkan subjek yang tidak terlalu banyak karena pengamatan berulang. Selain itu estimasinya lebih efisien karena dilakukan setiap pengamatan.

2. KAJIAN LITERATUR

2.1. Pendekatan Regresi Nonparametrik

Regresi nonparametrik mulai dikenal sekitar abad XIX, tepatnya pada tahun 1857 (Hardle, 1990). Regresi nonparametrik merupakan salah satu pendekatan yang digunakan untuk mengetahui pola hubungan antara variabel penjelas dan respon yang tidak diketahui kurva regresinya atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data (Eubank, 1988). Berdasarkan kenyataan tersebut, maka secara visual bentuk pola yang diberikan oleh variabel prediktor t dan variabel respon y tidak mempunyai pola yang pasti. Secara umum regresi nonparametrik memiliki bentuk fungsi sebagai berikut.

$$y = f(t_i) + \varepsilon_i, \quad i = 1, 2, \dots, n$$

dengan y_i adalah variabel respon sedangkan fungsi $f(t_i)$ merupakan kurva regresi dengan t_i sebagai variabel prediktor dan ε_i adalah *error* random yang diasumsikan berdistribusi normal independen dengan mean nol dan variansi σ^2 (Wahba, 1990).

2.2. Spline truncated

Spline merupakan salah satu potongan polinomial (*piecewise polynomial*) tersegmen yang memiliki sifat fleksibilitas. Sifat fleksibilitas inilah yang membedakan spline dengan polinomial. Titik perpaduan bersama dari potongan-potongan tersebut atau titik yang menunjukkan terjadinya perubahan-perubahan perilaku kurva pada interval – interval yang berbeda disebut knot.

Secara umum fungsi spline berorde m adalah sembarang fungsi yang dapat di tulis dalam bentuk (Eubank, 1988) :

$$f(t) = \sum_{j=1}^m a_j t^j + \sum_{k=1}^M b_k (t - K_k)_+^m$$

$$\text{dengan } (t - K_k)_+^m = \begin{cases} (t - K_k)^m & ; t \geq K \\ 0 & ; t < K \end{cases}$$

a_j dan b_k adalah parameter serta K_1, K_2, \dots, K_M adalah titik knot, $f(t)$ adalah fungsi regresi, serta t adalah variabel prediktor

Apabila $m = 1$ serta banyaknya knot satu, maka persamaan diatas dapat ditulis fungsi spline linear menjadi :

$$f(t) = a_1 t + b_2 (t - K)_+^1 \tag{1}$$

fungsi ini dinamakan spline linear satu knot pada $t = K$
 Fungsi (1) dapat pula ditulis (Budiantara, 2005)

$$f(t) = \begin{cases} a_1 t & ; t < K \\ a_1 t + b_2 (t - K) & ; t \geq K \end{cases}$$

Model regresi spline dapat disajikan dalam bentuk :

$$y_i = \sum_{j=1}^m a_j t_i^j + \sum_{k=1}^M b_k (t_i - K_k)_+^m + \varepsilon_i \quad (2)$$

Selanjutnya model (2) dapat ditulis dalam bentuk :

$$y_i = a_1 t_i + \dots + a_m t_i^m + b_1 (t_i - K_1)_+^m + \dots + b_M (t_i - K_M)_+^m + \varepsilon_i$$

Apabila model regresi disajikan dalam bentuk matriks didapat :

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{bmatrix} t_1 & \dots & t_1^m & (t_1 - K_1)_+^m & \dots & (t_1 - K_M)_+^m \\ t_2 & \dots & t_2^m & (t_2 - K_1)_+^m & \dots & (t_2 - K_M)_+^m \\ \vdots & & \vdots & \vdots & & \vdots \\ t_n & \dots & t_n^m & (t_n - K_1)_+^m & \dots & (t_n - K_M)_+^m \end{bmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_m \\ b_1 \\ \vdots \\ b_M \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Selanjutnya persamaan diatas dapat ditulis menjadi :

$$\tilde{y} = \mathbf{X}(K_1, \dots, K_M) \tilde{\beta} + \tilde{\varepsilon}$$

dengan,

$$\tilde{y} = (y_1, y_2, \dots, y_n)'$$

$$\mathbf{X}(K_1, \dots, K_M) = \begin{bmatrix} t_1 & \dots & t_1^m & (t_1 - K_1)_+^m & \dots & (t_1 - K_M)_+^m \\ t_2 & \dots & t_2^m & (t_2 - K_1)_+^m & \dots & (t_2 - K_M)_+^m \\ \vdots & & \vdots & \vdots & & \vdots \\ t_n & \dots & t_n^m & (t_n - K_1)_+^m & \dots & (t_n - K_M)_+^m \end{bmatrix},$$

$$\tilde{\beta} = (a_1, \dots, a_m, b_1, \dots, b_M)' \text{ dan } \tilde{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$$

2.3. Data Longitudinal

Dalam analisis regresi terdapat data longitudinal dan data *cross section*. Data longitudinal merupakan data yang pengamatannya dilakukan secara berulang untuk setiap subjek pada beberapa subjek. Pada data longitudinal diasumsikan subjek-subjek saling independen satu sama lain, akan tetapi antar pengamatan didalam subjek saling dependen sehingga terdapat korelasi (Wu dan Zhang, 2006. Tujuan dari analisis data longitudinal adalah untuk mempelajari bagaimana perubahan subjek yang diamati dari waktu ke waktu. Pada data longitudinal apabila terdiri lebih dari satu respon, maka terdapat korelasi dari setiap respon pada subjek yang sama (Weiss, 2005). Bentuk struktur data longitudinal birespon adalah sebagai berikut:

Tabel 2.1 Struktur Data Longitudinal Birespon

Subjek ke- j	Pengamatan ke- k	Prediktor	Respon ke-1	Respon ke-2
Subjek ke-1	1	x_{11}	y_{111}	y_{211}
	2	x_{12}	y_{112}	y_{212}
	\vdots	\vdots	\vdots	\vdots
	n	x_{1n}	y_{11n}	y_{21n}
Subjek ke-2	1	x_{21}	y_{121}	y_{221}
	2	x_{22}	y_{122}	y_{222}
	\vdots	\vdots	\vdots	\vdots
	n	x_{2n}	y_{12n}	y_{22n}
\vdots	\vdots	\vdots	\vdots	\vdots
Subjek ke- m	1	x_{m1}	y_{1m1}	y_{2m1}

	2	x_{m2}	y_{1m2}	y_{2m2}
	\vdots	\vdots	\vdots	\vdots
	n	x_{mn}	y_{1mn}	y_{2mn}

dengan $j = 1, 2, \dots, m$ dan $k = 1, 2, \dots, n$.

3. METODE PENELITIAN

Adapun langkah-langkah yang dilakukan untuk mendapatkan estimasi dari model spline truncated untuk data longitudinal adalah sebagai berikut:

1. Diberikan data longitudinal (t_{ij}, Y_{ij})
2. Diberikan model regresi nonparametrik untuk data longitudinal
 $y_{ij} = f(t_{ij}) + \varepsilon_{ij}$, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, n_i$
3. Mendekati $f(t_{ij})$ dengan spline truncated derajat m dengan knot k_1, k_2, \dots, k_K

$$f(t_{ij}) = \sum_{p=0}^m \alpha_{p_i} t_{ij}^p + \sum_{r=1}^K \beta_{r_i} (t_{ij} - k_{r_i})_+^m$$

4. Menyatakan bentuk regresi spline untuk data longitudinal sebagai berikut:

$$Y_{ij} = \sum_{p=0}^m \alpha_{p_i} t_{ij}^p + \sum_{r=1}^K \beta_{r_i} (t_{ij} - k_{r_i})_+^m + \varepsilon_{ij}$$

5. Menyajikan model tersebut dalam bentuk matriks

$$\mathbf{Y} = \mathbf{XB} + \mathbf{e}$$

6. Menentukan $\hat{\mathbf{B}}$ dengan menyelesaikan optimasi

$$\min_{\mathbf{B}} \{(\mathbf{Y} - \mathbf{XB})^T \mathbf{W}(\mathbf{Y} - \mathbf{XB})\}$$

7. Mendapatkan kurva estimasi regresi $f(t_{ij})$

$$\hat{f}(t_{ij}) = \sum_{p=0}^m \hat{\alpha}_{p_i} t_{ij}^p + \sum_{r=1}^K \hat{\beta}_{r_i} (t_{ij} - k_{r_i})_+^m$$

4. HASIL PENELITIAN

Diberikan data longitudinal (t_{ij}, Y_{ij}) , dengan $j = 1, 2, \dots, n_i$ dan $i = 1, 2, \dots, n$ yang diasumsikan mengikuti model regresi nonparametrik,

$$Y_{ij} = f(t_{ij}) + \varepsilon_{ij} \tag{3}$$

Model (3) dapat dituliskan sebagai berikut:

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ Y_{n1} \\ Y_{n2} \\ \vdots \\ Y_{mn_1} \end{bmatrix} = \begin{bmatrix} f(t_{11}) \\ f(t_{12}) \\ \vdots \\ f(t_{1n_1}) \\ f(t_{21}) \\ f(t_{22}) \\ \vdots \\ f(t_{2n_2}) \\ \vdots \\ f(t_{n1}) \\ f(t_{n2}) \\ \vdots \\ f(t_{mn_1}) \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1n_1} \\ e_{21} \\ e_{22} \\ \vdots \\ e_{2n_2} \\ \vdots \\ e_{n1} \\ e_{n2} \\ \vdots \\ e_{mn_1} \end{bmatrix} \quad (4)$$

Untuk mengestimasi kurva regresi nonparametrik pada data longitudinal dengan pendekatan spline truncated dilakukan atas dasar optimasi *Weighted Least Square* (WLS). Jika persamaan (4) dijabarkan untuk setiap model maka didapatkan

$$\begin{aligned}
 Y_{11} &= f(t_{11}) + e_{11} \\
 Y_{12} &= f(t_{12}) + e_{12} \\
 &\vdots \\
 Y_{1n_1} &= f(t_{1n_1}) + e_{1n_1} \\
 Y_{21} &= f(t_{21}) + e_{21} \\
 Y_{22} &= f(t_{22}) + e_{22} \\
 &\vdots \\
 Y_{2n_2} &= f(t_{2n_2}) + e_{2n_2} \\
 &\vdots \\
 Y_{n1} &= f(t_{n1}) + e_{n1} \\
 Y_{n2} &= f(t_{n2}) + e_{n2} \\
 &\vdots \\
 Y_{mn_1} &= f(t_{mn_1}) + e_{mn_1}
 \end{aligned}$$

Apabila kurva regresi f didekati dengan spline truncated maka

$$s(t_{ij}) = \sum_{p=0}^m \alpha_p t_{ij}^p + \sum_{r=1}^K \beta_r (t_{ij} - k_r)_+^m \text{ dengan } j = 1, 2, \dots, n_i \text{ dan } i = 1, 2, \dots, n$$

Dengan derajat polinomial m , $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \beta_2, \dots, \beta_K$ adalah parameter-parameter yang tidak diketahui, dan $\{k_r; k_r < k_{r_2} < \dots < k_{r_K}\}$ merupakan titik-titik knot, serta truncated power function $(t_{ij} - k_r)_+^m$ didefinisikan sebagai berikut:

$$(t_{ij} - k_r)_+^m = \begin{cases} (t_{ij} - k_r)^m, & t_{ij} \geq k_r \\ 0 & , t_{ij} < k_r \end{cases}$$

Dengan demikian model regresi untuk data longitudinal dapat disajikan sebagai berikut:

$$\begin{aligned}
 Y_{11} &= \sum_{p=0}^m \alpha_{p_0} t_{11}^p + \sum_{r=1}^K \beta_{r_1} (t_{11} - k_{r_1})_+^m + e_{11} \\
 Y_{12} &= \sum_{p=0}^m \alpha_{p_1} t_{12}^p + \sum_{r=1}^K \beta_{r_1} (t_{12} - k_{r_1})_+^m + e_{12} \\
 &\vdots \\
 Y_{1n_1} &= \sum_{p=0}^m \alpha_{p_1} t_{1n_1}^p + \sum_{r=1}^K \beta_{r_1} (t_{1n_1} - k_{r_1})_+^m + e_{1n_1} \\
 Y_{21} &= \sum_{p=0}^m \alpha_{p_2} t_{21}^p + \sum_{r=1}^K \beta_{r_2} (t_{21} - k_{r_2})_+^m + e_{21} \\
 Y_{22} &= \sum_{p=0}^m \alpha_{p_2} t_{22}^p + \sum_{r=1}^K \beta_{r_2} (t_{22} - k_{r_2})_+^m + e_{22} \\
 &\vdots \\
 Y_{2n_2} &= \sum_{p=0}^m \alpha_{p_2} t_{2n_2}^p + \sum_{r=1}^K \beta_{r_2} (t_{2n_2} - k_{r_2})_+^m + e_{2n_2} \\
 Y_{n1} &= \sum_{p=0}^m \alpha_{p_n} t_{n1}^p + \sum_{r=1}^K \beta_{r_n} (t_{n1} - k_{r_n})_+^m + e_{n1} \\
 Y_{n2} &= \sum_{p=0}^m \alpha_{p_n} t_{n2}^p + \sum_{r=1}^K \beta_{r_n} (t_{n1} - k_{r_n})_+^m + e_{n2} \\
 &\vdots \\
 Y_{mn_n} &= \sum_{p=0}^m \alpha_{p_n} t_{mn_n}^p + \sum_{r=1}^K \beta_{r_n} (t_{mn_n} - k_{r_n})_+^m + e_{mn_n}
 \end{aligned}$$

Berdasarkan persamaan tersebut dapat dituliskan dalam bentuk matriks.

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \vdots \\ Y_{n1} \\ \vdots \\ Y_{mn_n} \end{bmatrix} = \begin{bmatrix} 1 & t_{11} & \cdots & t_{11}^m & (t_{11} - k_{r_1})_+^m & \cdots & (t_{11} - k_{r_{K_1}})_+^m & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{1n} & \cdots & t_{1n}^m & (t_{1n} - k_{r_1})_+^m & \cdots & (t_{1n} - k_{r_{K_1}})_+^m & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & t_{n1} & \cdots & t_{n1}^m & (t_{n1} - k_{r_n})_+^m & \cdots & (t_{n1} - k_{r_{K_n}})_+^m & \alpha_{0_n} & \alpha_{1_n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & t_{mn} & \cdots & t_{mn}^m & (t_{mn} - k_{r_n})_+^m & \cdots & (t_{mn} - k_{r_{K_n}})_+^m & \alpha_{0_n} & \alpha_{1_n} \end{bmatrix} + \begin{bmatrix} \alpha_{0_1} \\ \alpha_{1_1} \\ \vdots \\ \alpha_{m_1} \\ \beta_{r_1} \\ \vdots \\ \beta_{r_{K_1}} \\ \vdots \\ \alpha_{0_n} \\ \alpha_{1_n} \\ \vdots \\ \alpha_{m_n} \\ \beta_{r_n} \\ \vdots \\ \beta_{r_{K_n}} \end{bmatrix} = \begin{bmatrix} e_{11} \\ \vdots \\ e_{1n_1} \\ \vdots \\ e_{n1} \\ \vdots \\ e_{mn_n} \end{bmatrix}$$

Matriks basis spline truncated

$$\mathbf{x}_1[\mathbf{k}] = \begin{bmatrix} 1 & t_{11} & \cdots & t_{11}^m & (t_{11} - k_{1_1})_+^m & \cdots & (t_{11} - k_{K_1})_+^m \\ 1 & t_{12} & \cdots & t_{12}^m & (t_{12} - k_{1_1})_+^m & \cdots & (t_{12} - k_{K_1})_+^m \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_{1n_1} & \cdots & t_{1n_1}^m & (t_{1n_1} - k_{1_1})_+^m & \cdots & (t_{1n_1} - k_{K_1})_+^m \end{bmatrix}$$

$$\mathbf{x}_2[\mathbf{k}] = \begin{bmatrix} 1 & t_{21} & \cdots & t_{21}^m & (t_{21} - k_{1_2})_+^m & \cdots & (t_{21} - k_{K_2})_+^m \\ 1 & t_{22} & \cdots & t_{22}^m & (t_{22} - k_{1_2})_+^m & \cdots & (t_{22} - k_{K_2})_+^m \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_{2n_2} & \cdots & t_{2n_2}^m & (t_{2n_2} - k_{1_2})_+^m & \cdots & (t_{2n_2} - k_{K_2})_+^m \end{bmatrix}$$

$$\vdots$$

$$\mathbf{x}_n[\mathbf{k}] = \begin{bmatrix} 1 & t_{n1} & \cdots & t_{n1}^m & (t_{n1} - k_{1_n})_+^m & \cdots & (t_{n1} - k_{K_n})_+^m \\ 1 & t_{n2} & \cdots & t_{n2}^m & (t_{n2} - k_{1_n})_+^m & \cdots & (t_{n2} - k_{K_n})_+^m \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_{nn_1} & \cdots & t_{nn_1}^m & (t_{nn_1} - k_{1_n})_+^m & \cdots & (t_{nn_1} - k_{K_n})_+^m \end{bmatrix}$$

Vektor parameter

$$\mathbf{B}_1 = [\alpha_{0_1} \quad \alpha_{1_1} \quad \cdots \quad \alpha_{m_1} \quad \beta_{1_1} \quad \cdots \quad \beta_{K_1}]^T,$$

$$\mathbf{B}_2 = [\alpha_{0_2} \quad \alpha_{1_2} \quad \cdots \quad \alpha_{m_2} \quad \beta_{1_2} \quad \cdots \quad \beta_{K_2}]^T$$

$$\mathbf{B}_n = [\alpha_{0_n} \quad \alpha_{1_n} \quad \cdots \quad \alpha_{m_n} \quad \beta_{1_n} \quad \cdots \quad \beta_{K_n}]^T$$

dan vektor error,

$$\mathbf{e}_1 = [e_{11} \quad e_{12} \quad \cdots \quad e_{1n_1}]^T, \quad \mathbf{e}_2 = [e_{21} \quad e_{22} \quad \cdots \quad e_{2n_2}]^T \quad \text{dan} \quad \mathbf{e}_n = [e_{n1} \quad e_{n2} \quad \cdots \quad e_{nn_n}]^T$$

Shingga model regresi nonparametrik spline truncated-nya bisa dibuat sebagai berikut:

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1[\mathbf{k}] & 0 & 0 & 0 \\ 0 & \mathbf{x}_2[\mathbf{k}] & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{x}_n[\mathbf{k}] \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_n \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} \quad (5)$$

Dengan demikian persamaan (5) dapat disajikan dala notasi matriks sebagai berikut:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{e} \quad (6)$$

Dimana

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_n \end{bmatrix}; \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1[\mathbf{k}] & 0 & 0 & 0 \\ 0 & \mathbf{x}_2[\mathbf{k}] & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{x}_n[\mathbf{k}] \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_n \end{bmatrix}; \quad \mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{bmatrix}$$

Dengan menggunakan bobot \mathbf{W} , estimasi dari \mathbf{B} pada persamaan (6) dapat diperoleh dengan menyelesaikan optimasi sebagai berikut:

$$\min_{\mathbf{B}} \{(\mathbf{Y} - \mathbf{XB})^T \mathbf{W}(\mathbf{Y} - \mathbf{XB})\} \quad (7)$$

Dengan W matriks sebagai berikut:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{W}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{W}_n \end{bmatrix} \text{ dimana } \mathbf{W}_1 = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n_1} \\ w_{21} & w_{22} & \cdots & w_{2n_1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_1 1} & w_{n_1 1} & \cdots & w_{n_1 n_1} \end{bmatrix}$$

$$\mathbf{W}_2 = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n_2} \\ w_{21} & w_{22} & \cdots & w_{2n_2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_2 1} & w_{n_2 1} & \cdots & w_{n_2 n_2} \end{bmatrix}, \dots, \mathbf{W}_n = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n_n} \\ w_{21} & w_{22} & \cdots & w_{2n_n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_n 1} & w_{n_n 1} & \cdots & w_{n_n n_n} \end{bmatrix}$$

Penyelesaian persamaan (7) dilakukan dengan penjabaran sebagai berikut:

$$(\mathbf{Y} - \mathbf{XB})^T \mathbf{W} (\mathbf{Y} - \mathbf{XB}) = \mathbf{Y}^T \mathbf{W} \mathbf{Y} - 2\mathbf{B}^T \mathbf{X}^T \mathbf{W} \mathbf{Y} + \mathbf{B}^T \mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{B}$$

Misalkan $L = \mathbf{Y}^T \mathbf{W} \mathbf{Y} - 2\mathbf{B}^T \mathbf{X}^T \mathbf{W} \mathbf{Y} + \mathbf{B}^T \mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{B}$ maka

$$\frac{\partial L}{\partial \mathbf{B}} = -2\mathbf{X}^T \mathbf{W} \mathbf{Y} + 2\mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{B}$$

Dan nilai minimum dari B diperoleh $\frac{\partial L}{\partial \mathbf{B}} = 0$ sedemikian sehingga

$$-2\mathbf{X}^T \mathbf{W} \mathbf{Y} + 2\mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{B} = 0$$

$$-\mathbf{X}^T \mathbf{W} \mathbf{Y} + \mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{B} = 0$$

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{B} = \mathbf{X}^T \mathbf{W} \mathbf{Y} \tag{8}$$

Pada persamaan (8) jika kedua ruas dikalikan dengan $(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$ maka diperoleh:

$$(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{B} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y} \text{ sehingga}$$

$$\hat{\mathbf{B}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y} \tag{9}$$

Berdasarkan estimasi dari B pada persamaan (9) diperoleh

$$\hat{\mathbf{Y}} = \mathbf{X} \hat{\mathbf{B}}$$

$$= \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}$$

$$= \mathbf{A}[\mathbf{k}] \mathbf{Y}$$

Dimana $\mathbf{A}[\mathbf{k}] = \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}$

Kriteria GCV merupakan salah satu metode untuk memilih knot yang optimal. Nilai GCV untuk data longitudinal ditentukan sebagai berikut:

$$GCV[k] = \frac{N^{-1} \mathbf{Y}^T (\mathbf{I} - \mathbf{A}[\mathbf{k}])^T \mathbf{W} (\mathbf{I} - \mathbf{A}[\mathbf{k}]) \mathbf{Y}}{(N^{-1} \text{trace}(\mathbf{I} - \mathbf{A}[\mathbf{k}]))^2}$$

Selanjutnya knot optimal dipilih dengan menggunakan metode GCV yang meminimumkan GCV [k].

5. KESIMPULAN

Berdasarkan analisis dan pembahasan yang sudah dilakukan, estimasi model spline truncated untuk data longitudinal menggunakan WLS menghasilkan estimator $\hat{\mathbf{B}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}$. Pada pemilihan titik knot optimum untuk model spline truncated data longitudinal, dilakukan masing-masing subjek, tidak serentak. Akan tetapi untuk estimasi modelnya dilakukan secara serentak, sehingga setiap subjek mempunyai model masing-masing. Akan tetapi R-square, atau MSE ataupun MAPE yang didapatkan hanya satu untuk seluruh subjek.

6. REFERENSI

- Budiantara, I N., 2009, *Spline Dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*, Institut Teknologi Sepuluh November, Surabaya.
- Budiantara, I.N., Suryadi, F., Otok, B.W., dan Guritno, S., 2006, Pemodelan B- Spline dan MARS Pada Nilai Ujian Masuk terhadap IPK Mahasiswa Jurusan Disain Komunikasi Visual UK. Petra Surabaya; *Jurnal Teknik Industri*, Vol 8 No. 1 hal 1-13; Universitas Petra.
- Budiantara, I.N. (2000), "Metode U, GML, CV, dan GCV dalam regresi Nonparametrik Spline", *Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI)*, Vol. 6, hal. 285-290.
- Draper, N.R dan Smith, H, 1992, *Applied Regression Analysis, 2nd edition*, John Wiley & Sons, Chapman and Hall, New York.
- Eubank, R.L., 1988, *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York.
- Hardle, W., 1990, *Applied Non-parametric Regression*, Cambridge University Press, Cambridge.
- Wahba, G., 1990, *Spline Models For Observation Data*, SIAM Pennsylvania
- Wang, Y., Guo, W., dan Brown, M.B. (1999). *Spline smoothing for Bivariate Data with Application to Association Between Hormones*, Department of Statistics And Applied Probability, University of California, Santa Barbara.
- Wu, H. dan Zhang, J.T. (2006), *Nonparametric Regression Methods for Longitudinal Data Analysis*, A John-Wiley and Sons Inc. Publication, New Jersey.
- Wu, W.B dan Pourahmadi, M. (2003), "Nonparemetric Estimation of Large Covariance Matrices of Longitudinal Data", *Journal of Biometrika*, Vol. 90, No. 4, hal. 831-844.