J M S K
Jurnal Matematika, Statistika, & Komputasi
http://journal.unhas.ac.id/index.php/jmsk
Vol. 11, No. 1, 77-83, Juli 2014
ISBN:1858-1382

# Taksiran Kurva Regresi Spline pada Data Longitudinal dengan Kuadrat Terkecil

**Anna Islamiyati***

**Abstrak**

Makalah ini mengkaji tentang estimasi regresi spline khususnya penggunaan pada data longitudinal. Data longitudinaladalah data yang mampu membedakan keragaman respon yang disebabkan karena pengukuran yang berulang. Kurva regresi spline diestimasi dengan menggunakan kuadrat terkecil. Terlihat bahwa taksiran kurva regresi spline untuk data longitudinal merupakan kelas pendugaan linear dalam observasi respon $\tilde{y}_i$ dan sangat tergantung pada titik knot$k_1$, $k_2$, ..., $k_r$.

**Keywords**: data longitudinal, kuadrat terkecil, dan regresi spline.

## 1. Pendahuluan

Regresi nonparametrik digunakan apabila bentuk kurva regresi diasumsikan tidak diketahui. Regresi nonparametrik memiliki fleksibilitas yang tinggi dalam mengestimasi kurva regresi. Berbeda dengan regresi parametrik yang mengasumsikan bentuk kurva regresi diketahui seperti linear, kuadratik, kubik, eksponensial atau yang lainnya, pendekatan regresi nonparametrik tidak mengasumsikan bentuk awal dari kurva regresi. Sehingga diperlukan pendekatan dalam mengestimasi kurva regresi nonparametrik,salah satunya adalah metode spline. Eubank [4] menyatakan spline merupakan salah satu model yang mempunyai interpretasi statistik dan interpretasi visual sangat khusus dan sangat baik, melalui pemilihan titik knot optimal Di samping itu, spline mampu menangani karakter data fungsi yang bersifat mulus (*smooth*) melalui pemilihan parameter penghalus optimal.

Islamiyati (2010) menguraikan penggunaan regresi spline polynomial truncuted pada data cross sectional. Namun dalam perkembangan riset selama ini, telah banyak jenis data pengukuran yang diperoleh di lapangan, bukan hanya dalam bentuk cross sectional, diantaranya data longitudinal. Wu dan Zhang [9] menyatakan data longitudinal adalah data pengamatan yang dilakukan terhadap *n* obyek yang saling independen, setiap obyek diamati secara berulang dan kontinu dalam kurun waktu tertentu, dimana pengamatan dalam obyek yang sama saling berkorelasi. Perbedaan struktur data tersebut menyebabkan perlu kajian tentang penggunaan regresi spline pada data longitudinal.

Makalah ini mengkaji tentang estimasi kurva regresi spline pada data longitudinal, dimana metode estimasi yang digunakan adalah metode kuadrat terkecil, dengan memilih titik knot optimal berdasarkan nilai *Gross Cross Validation* (GCV) terkecil.

## 2. Data Longitudinal

Data longitudinal merupakan data yang diamati dan diukur berulangkali pada suatu interval waktu tertentu. Dibandingkan dengan data yang diperoleh pada studi *cross sectional* yang umumnya dilakukan pada bidang sosial dan ekonomi, dimana pengukuran terhadap obyek hanya dilakukan sekali saja, maka data longitudinal memiliki keunggulan, yaitu

kemampuannya dalam mengenali pengaruh waktu pengukuran terhadap respon. Data longitudinal diasumsikan obyek-obyek saling independen satu sama lainnya, tetapi antara pengamatan di dalam obyek yang sama saling dependen, karena cenderung berkorelasi [3].

Dalam studi tentang data longitudinal, pada umumnya pengamatan dilakukan terhadap *n* obyek yang saling independen, dimana setiap obyek diamati secara berulang (*repeated measurement*) dalam kurun waktu yang berbeda. Misalkan $t_{ij}$ menyatakan pengamatan pada waktu ke-*j* dari obyek ke-*i* dan $y_{ij}$ menyatakan variabel respon pada waktu $t_{ij}$ , maka data longitudinal diberikan oleh $(t_{ij}, y_{ij}), j = 1,2, \dots, n_i; i = 1,2, \dots, n$ dimana $n_i$ menyatakan banyaknya pengukuran berulang dari individu ke-*i* atau dapat ditulis dalam bentuk persamaan:

$$y_{ij} = f_i(t_{ij}) + \varepsilon_{ij}\ ,\ j = 1,2, \dots, n_i;\ i = 1,2, \dots, n \quad (1)$$

[9].

## 3. Model Regresi Spline

Spline adalah potongan polinomial order *p* dengan titik bersama dari potongan-potongan tersebut disebut dengan knot. Titik knot merupakan perpaduan dua kurva yang menunjukkan pola perubahan perilaku kurva pada selang yang berbeda. Penggunaan titik knot banyak digunakan dalam regresi nonparametrik, karena secara visual dapat menunjukkan setiap perubahan pola perilaku yang terjadi dalam interval waktu tertentu (Islamiyati, 2009). Misalkan pola perubahan yang terjadi pada data sebanyak lima pola perubahan, dimana titik terjadinya pola perubahan tersebut disebut titik knot. Pola perubahan yang terjadi, yaitu pola pertama cenderung naik, kemudian menurun pada pola kedua. Selanjutnya pola ketiga juga menunjukkan kecenderungan turun tetapi penurunannya berbeda dengan pola kedua. Pola keempat mengalami kenaikan kembali dan terus naik pada pola kelima tetapi dengan kecenderungan naik yang berbeda pula. Contoh ini menunjukkan bahwa dengan penggunaan regresi spline, sangat memungkinkan dalam satu data terdapat beberapa pola perubahan dalam setiap interval waktu berbeda.

Spline orde *p* dengan knot $k_1, k_2, \mathrm{K}\ , k_m$ diberikan dalam fungsi *f* dengan bentuk:

$$f(t_i) = \sum_{j=0}^{p} \beta_j t_i^j + \sum_{j=1}^{m} \beta_{j+p} (t_i - k_j)_+^p, \quad (2)$$

dengan $\beta_0, \beta_1, \mathrm{K}\ , \beta_{p+j}$ adalah parameter regresi, dan

$$(t_i - k_j)_+^p = \begin{cases} (t_i - k_j)_+^p & ,(t_i - k_j) \geq 0 \\ 0 & ,(t_i - k_j) < 0 \end{cases}$$

[4].

Salah satu cara pemilihan titik knot optimal adalah menggunakan metode *generalized* c*ross validation* (GCV). Kriteria GCV didefenisikan:

$$GCV\ (k) = \frac{MSE\ (k)}{[n^{-1}\ trace\ (\boldsymbol{I} - \boldsymbol{A}(k))]^2}, \quad (3)$$

dengan:

MSE (*k*) = $n^{-1}\tilde{\boldsymbol{y}}^T \left( (\mathbf{I} - \mathbf{A}(k))^T (\mathbf{I} - \mathbf{A}(k)) \right) \tilde{\boldsymbol{y}}$, *k* adalah titik knots,**A(***k*) = $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ , $\boldsymbol{A}(k)$ adalah matriks yang berukuran$_{nxn}$, $\tilde{\boldsymbol{y}} = (\widetilde{\boldsymbol{y_1}}\widetilde{\boldsymbol{y_2}}\ \dots\ \widetilde{\boldsymbol{y_n}})^T$. [5]

## 4. Taksiran Kurva Regresi Spline pada Data Longitudinal

Data longitudinal yang diukur berulangkali berdasarkan waktu diberikan oleh $(t_{ij}, y_{ij}), j = 1,2, \dots, n_i; i = 1,2, \dots, n,$ dimana $n_i$ menyatakan banyaknya pengukuran berulang dari obyek ke-$i$.Jika diberikan model regresi nonparametrik untuk data longitudinal maka diperoleh suatu bentuk seperti pada (1).
Spline pada data longitudinal diberikan dengan bentuk persamaan :

$$f_i(t_{ij}) = \sum_{l=0}^{p} \beta_{li} t_{ij}^{l} + \sum_{m=1}^{r} \beta_{(p+m)i} (t_{ij} - k_m)_+^p \quad (4)$$

Dimana

$k_1, k_2, \dots, k_r$ = titik knot
$p$ = jumlah orde
$t_{ij}$ = pengaruh variabel waktu pada objek ke-i dengan pengulangan ke-j
$\beta$= parameter

Spline orde $p$, dapat dimodelkan sebagai berikut:

$$f_i(t_{ij}) = \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} {t_{ij}}^2 + \cdots + \beta_{pi} {t_{ij}}^p + \sum_{m=1}^{r} \beta_{(p+m)i} (t_{ij} - k_m)_+^p \quad (5)$$

Menurut model spline pada (5), maka model regresi nonparametrik berdasarkan (1) dapat ditulis :

$$y_i(t_{ij}) = \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} {t_{ij}}^2 + \cdots + \beta_{pi} {t_{ij}}^p + \beta_{(p+1)i} (t_{ij} - k_1)^p + \beta_{(p+2)i} (t_{ij} - k_2)^p + \cdots + \beta_{(p+r)i} (t_{ij} - k_r)^p + \varepsilon_{ij} \quad (6)$$

yang dapat disajikan dengan bentuk matriks, yaitu:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_i} \\ \hline y_{21} \\ y_{22} \\ \vdots \\ y_{2n_i} \\ \hline \vdots \\ \\ \\ \\ \overline{y_{n1}} \\ y_{n1} \\ \vdots \\ y_{nn_i} \end{bmatrix} =$$

$$\begin{bmatrix} 1 & t_{11} & {t_{11}}^2 & \cdots & {t_{11}}^p & (t_{11}-k_1)_+^1 & \cdots & (t_{11}-k_r)_+^p \\ 1 & t_{12} & {t_{12}}^2 & \cdots & {t_{12}}^p & (t_{12}-k_1)_+^1 & \cdots & (t_{12}-k_r)_+^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{1n_i} & {t_{1n_i}}^2 & \cdots & {t_{1n_i}}^p & \left(t_{1n_i}-k_1\right)_+^1 & \cdots & \left(t_{1n_i}-k_r\right)_+^p \\ \hline 1 & t_{21} & {t_{21}}^2 & \cdots & {t_{21}}^p & (t_{21}-k_1)_+^1 & \cdots & (t_{21}-k_r)_+^p \\ 1 & t_{22} & {t_{22}}^2 & \cdots & {t_{22}}^p & (t_{22}-k_1)_+^1 & \cdots & (t_{22}-k_r)_+^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{2n_i} & {t_{2n_i}}^2 & \cdots & {t_{2n_i}}^p & \left(t_{2n_i}-k_1\right)_+^1 & \cdots & \left(t_{2n_i}-k_r\right)_+^p \\ \hline & & & & & \vdots & & \\ \\ \\ \\ & & & & & \overline{(t_{n1}-k_1)_+^1} & & \\ 1 & t_{n1} & {t_{n1}}^2 & \cdots & {t_{n1}}^p & (t_{n1}-k_1)_+^1 & \cdots & (t_{n1}-k_m)_+^p \\ 1 & t_{n2} & {t_{n2}}^2 & \cdots & {t_{n2}}^p & (t_{n2}-k_1)_+^1 & \cdots & (t_{n2}-k_m)_+^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{nn_i} & {t_{nn_i}}^2 & \cdots & {t_{nn_i}}^p & \left(t_{nn_i}-k_1\right)_+^1 & \cdots & \left(t_{nn_i}-k\ \ \right)_+^p \\ \hline \end{bmatrix} \begin{bmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{21} \\ \vdots \\ \beta_{p1} \\ \beta_{(p+1)1} \\ \vdots \\ \beta_{(p+r)1} \\ \hline \beta_{02} \\ \beta_{12} \\ \beta_{22} \\ \vdots \\ \beta_{p2} \\ \beta_{(p+1)2} \\ \vdots \\ \beta_{(p+r)2} \\ \hline \vdots \\ \\ \\ \\ \\ \overline{\beta_{0n}} \\ \beta_{1n} \\ \beta_{2n} \\ \vdots \\ \beta_{pn} \\ \beta_{(p+1)n} \\ \vdots \\ \beta_{(p+r)n} \end{bmatrix} +$$

$$\begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n_i} \\ \hline \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n_i} \\ \hline \vdots \\ \\ \hline \varepsilon_{n1} \\ \varepsilon_{n2} \\ \vdots \\ \varepsilon_{nn_i} \end{bmatrix} \tag{7}$$

Model matriks pada (7) dapat disederhanakan dalam bentuk:

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_n \end{bmatrix}_{nx1} = \begin{bmatrix} \tilde{1} & \tilde{t}_1 \tilde{t}_1{}^2 & \ldots \tilde{t}_1{}^p & (\tilde{t}_1 - k_1)_+^p & \ldots & (\tilde{t}_1 - k_m)_+^p \\ \tilde{1} & \tilde{t}_2 \tilde{t}_2{}^2 & \tilde{t}_2{}^p & (\tilde{t}_2 - k_1)_+^p & & (\tilde{t}_2 - k_m)_+^p \\ \vdots & \vdots\ \vdots & \vdots & \vdots & & \vdots \\ \tilde{1} & \tilde{t}_n \tilde{t}_n{}^2 & \tilde{t}_n{}^p & (\tilde{t}_n - k_1)_+^p & & (\tilde{t}_n - k_m)_+^p \end{bmatrix} \begin{bmatrix} \widetilde{\beta_1} \\ \widetilde{\beta_2} \\ \vdots \\ \widetilde{\beta_n} \end{bmatrix} + \begin{bmatrix} \tilde{\varepsilon}_1 \\ \tilde{\varepsilon}_2 \\ \vdots \\ \tilde{\varepsilon}_n \end{bmatrix}$$

Jika dituliskan dalam notasi matriks, dapat ditulis menjadi :

$$\tilde{y} = \boldsymbol{X}\tilde{\beta} + \tilde{\varepsilon} \tag{8}$$

Untuk memperoleh bentuk pendugaan $\tilde{\beta}$dilakukan melalui metode kuadrat terkecil dengan cara meminimumkan Jumlah Kuadrat Galat (JKG):

$$\varepsilon^T\varepsilon = \left(\tilde{\boldsymbol{y}} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)^T\left(\tilde{\boldsymbol{y}} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)$$

$$M = \varepsilon^T\varepsilon = \left(\tilde{\boldsymbol{y}} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)^T\left(\tilde{\boldsymbol{y}} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)$$

$$= \left[\tilde{\boldsymbol{y}}^T\tilde{\boldsymbol{y}} - 2\widetilde{\boldsymbol{\beta}}^T\boldsymbol{X}^T\tilde{\boldsymbol{y}} + \widetilde{\boldsymbol{\beta}}^T\boldsymbol{X}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right].$$

Selanjutnya diperoleh:

$$\frac{\partial M}{\partial \widetilde{\beta}} = \frac{\partial\left(\tilde{\boldsymbol{y}}^T\tilde{\boldsymbol{y}} - 2\widetilde{\boldsymbol{\beta}}^T\boldsymbol{X}^T\tilde{\boldsymbol{y}} + \widetilde{\boldsymbol{\beta}}^T\boldsymbol{X}^T\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)}{\partial \widetilde{\beta}}$$

$$= -2\boldsymbol{X}^T\tilde{\boldsymbol{y}} + 2\boldsymbol{X}^T\boldsymbol{X}\tilde{\boldsymbol{\beta}}$$

Kemudian :

$$-2\boldsymbol{X}^T\tilde{\boldsymbol{y}} + 2\boldsymbol{X}^T\boldsymbol{X}\tilde{\boldsymbol{\beta}} = 0$$

$$2\boldsymbol{X}^T\tilde{\boldsymbol{y}} = 2\boldsymbol{X}^T\boldsymbol{X}\tilde{\boldsymbol{\beta}}$$

$$\boldsymbol{X}^T\tilde{\boldsymbol{y}} = \boldsymbol{X}^T\boldsymbol{X}\tilde{\boldsymbol{\beta}}$$

$$\tilde{\tilde{\boldsymbol{\beta}}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\tilde{\boldsymbol{y}} \text{ atau } B = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\tilde{\boldsymbol{y}}$$

Akibatnya pendugaan kurva regresi $f_i(t_{ij})$ diberikan oleh:

$$\hat{\tilde{f}}(k_1, k_2, \dots, k_r) = \boldsymbol{X}B$$
$$= \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\tilde{\boldsymbol{y}} = \mathbf{A}(k_1, k_2, \dots, k_r)\tilde{\boldsymbol{y}} ;$$

dengan matriks $\mathbf{A}(k_1, k_2, \dots, k_r) = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$

Terlihat bahwa pendugaan untuk kurva regresi spline untuk data longitudinal merupakan kelas pendugaan linear dalam observasi respon $\tilde{y}_i$ dan sangat tergantung pada titik knot $k_1, k_2, \dots, k_r$.

## 5. Kesimpulan

Estimasi kurva regresi spline untuk data longitudinal dapat disajikan dalam bentuk:
$\hat{\tilde{f}}(k_1, k_2, \dots, k_r) = \boldsymbol{X}B = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\tilde{\boldsymbol{y}} = \mathbf{A}(k_1, k_2, \dots, k_r)\tilde{\boldsymbol{y}}$ ;

dengan $\underset{\sim}{y}$ adalah variabel respon berorde $N \times 1$ diberikan oleh:

$$\underset{\sim}{y} = (\underset{\sim}{y}_1, \underset{\sim}{y}_2, \cdots, \underset{\sim}{y}_r)^T .$$

## DAFTAR PUSTAKA

[1]. Besse, P.C., Cardot, H., dan Ferraty, F. (1997), "Simultaneous Nonparametric Regression of Unbalanced Longitudinal Data", *Computational Statistics and Data Analysis***24**, 255 – 270.

[2]. Carroll, R.J., Hall, P., Apanasovich, T.V., dan Lin, X. (2004), "Histospline Method in Nonparametric Regression Models with Application to Clustered/Longitudinal Data", *Statistica Sinica***14**, 649 – 674.

[3]. Diggle, P.J.K., Liang, K.Y., dan Zeger, S.L. (1995), *Analysis of Longitudinal Data,* Clarendon Press, Oxford.

[4] Eubank, R. L. (1988), *Spline smoothing and Nonparametrik Regression*, Marcel Dekker, New York.

[5] Green, P.J., dan Silverman, B.W. (1994), *Nonparametrik Regression and Generalized Linear Models (a Roughness Penalty Approach*), Chapman & Hall, New York.

[6] Islamiyati, Anna. (2009). *Estimasi Spline untuk Data Longitudinal dengan Penalized Likelihood*. Seminar Nasional Matematika IV ITS, Surabaya.

[7] Islamiyati, Anna (2012). "Regresi Nonparametrik untuk Data Longitudinal". *Jurnal Matematika Statistika & Komputasi*, Unhas, Makassar.

[8] Wahba, G. (1990), *Spline Model for Observational Data*, Society For Industrial and Applied Mathematics, Philadelphia.

[9] Wu, H., dan Zhang, J.T. (2006), *Nonparametric Regression Methods for Longitudinal Data Analysis*, John Wiley & Sons, New Jersey.