Estimasi Kurva *Smooth* dengan Regresi *Penalized Spline* pada Data Longitudinal

Anna Islamiyati[1], Fatmawati, dan Nur Chamidah
[1]Prodi Statistika Jurusan Matematika FMIPA Unhas

Abstrak

Kriteria regresi *penalized spline* terdiri atas dua fungsi yaitu fungsi *goodness of fit* yang memuat titik knot dan fungsi penalti yang memuat parameter penghalus. Peranan parameter penghalus akan mengontrol terbentuknya kurva *smooth* mendekati bentuk kurva populasi. Parameter penghalus optimal dipilih berdasarkan nilai GCV minimum.

Kata Kunci: data longitudinal, *GCV, kurva smooth, parameter penghalus, regresi penalized spline*

1. Pendahuluan

Regresi nonparametric digunakan ketika pola data tidak mengikuti suatu pola parametrik. Salah satu estimator yang digunakan dalam regresi nonparametric adalah spline. Spline merupakan fungsi polynomial yang tersegmen kontinu yang diasumsikan smooth. Spline telah dikembangkan oleh banyak peneliti, diantaranya spline truncated oleh Soo (1996), spline terbobot oleh Budiantara (1997), spline smoothing oleh Fernandez (2014). Terdapat estimator spline yang lain yaitu penalized spline yang mulai dikembangkan oleh Ruppert (2000) dan Monotoya (2014) namun perkembangannya masih sebatas pada data cross section. Penelitian ini mengkaji penggunaan estimator penalized spline pada data longitudinal dalam menghasilkan kurva smooth. Estimator penalized spline merupakan estimator spline yang melibatkan titik knot dan parameter penghalus secara simultan dalam mengestimasi kurva smooth. Titik knot termuat dalam fungsi goodness of fit dan parameter penghalus termuat dalam fungsi penalti. Hal ini yang menyebabkan estimator penalized spline unggul dibandingkan estimator spline yang lain dalam estimasi kurva smooth. Selain menghasilkan kurva smooth juga interpretasi data dapat dilakukan melalui titik knot yang menunjukkan pola perubahan yang terjadi pada interval tertentu.

Makalah ini akan menunjukkan kemampuan regresi penalized spline dalam menghasilkan kurva smooth yang mendekati kurva populasi pada data longitudinal. Studi simulasi akan dilakukan untuk menunjukkan kebaikan estimator penalized spline dalam fungsi polynomial dan fungsi trigonometri. Pemilihan model optimal dilakukan berdasarkan dari nilai GCV minimum.

2. Estimasi Kurva Regresi Nonparametrik dengan Penalized Spline pada Data Longitudinal

Spline merupakan fungsi potongan polinomial orde q dengan titik bersama dari potongan-potongan tersebut disebut titik knot. Eubank (1999) dan Budiantara (2006) menjelaskan bahwa titik knot merupakan perpaduan dua kurva yang menunjukkan pola perubahan perilaku kurva pada selang yang berbeda. Jika $f\left(t_i\right)$ dalam persamaan (2.1) dinyatakan sebagai fungsi *spline* orde q dengan knot pada $K_1, K_2, \ldots, K_d$ yaitu:

$$f\left(t_i\right) = \sum_{u=0}^{q} \beta_u t_i^u + \sum_{v=1}^{d} \beta_{q+v} \left(t_i - K_v\right)_+^q,\ i = 1, 2, \ldots, n\ , \tag{2.3}$$

maka $f\left(t_i\right)$ disebut sebagai fungsi regresi nonparametrik *spline*, dengan $\beta_0, \beta_1, \ldots, \beta_q, \beta_{q+1}, \beta_{q+2}, \ldots, \beta_{q+d}$ adalah parameter regresi, dan $(t_i - K_v)_+^q$ adalah elemen *truncated* yang akan memenuhi persamaan berikut:

$$(t_i - K_v)_+^q = \begin{cases} (t_i - K_v)^q & ,(t_i - K_v) \geq 0 \\ 0 & ,(t_i - K_v) < 0 \end{cases} . \tag{2.4}$$

Persamaan (2.4) mengandung arti bahwa jika t bernilai sama atau lebih besar dari K maka elemen *truncated* akan bernilai $(t_i - K_v)_+^q$, dan jika t bernilai kurang dari K maka elemen *truncated* akan bernilai nol. Estimasi fungsi regresi nonparametrik *spline* dapat menggunakan estimator *spline truncated* melalui kriteria *goodness of fit* dari persamaan (2.1), yaitu:

$$\varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - f\left(t_i\right)\right)^2 , \tag{2.5}$$

dengan $f\left(t_i\right)$ adalah fungsi *spline* seperti pada persamaan (2.3).

Estimator *penalized spline* terbentuk dari fungsi *spline truncated* dalam kriteria *Penalized Least Square* (PLS), sehingga estimator *penalized spline* menggunakan titik knot dan parameter penghalus secara bersamaan dalam mengestimasi fungsi regresi nonparametrik. Fungsi *truncated* dengan orde q yang berdasarkan pada titik knot $a < K_1 < \ldots < K_d < b$, dinyatakan oleh $f\left(t_i\right)$ seperti pada persamaan (2.3). $f\left(t_i\right)$ dapat dinyatakan dalam bentuk matriks yaitu:

$$\underset{\sim}{f} = \mathbf{X}\underset{\sim}{\beta} , \tag{2.8}$$

dengan $\mathbf{X} = \begin{bmatrix} 1 & t_1 & \cdots & t_1^q & \left(t_1 - K_1\right)_+^q & \cdots & \left(t_1 - K_d\right)_+^q \\ 1 & t_2 & \cdots & t_2^q & \left(t_2 - K_1\right)_+^q & \cdots & \left(t_2 - K_d\right)_+^q \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & \cdots & t_n^q & \left(t_n - K_1\right)_+^q & \cdots & \left(t_n - K_d\right)_+^q \end{bmatrix}$, dan $\underset{\sim}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \\ \beta_{q+1} \\ \vdots \\ \beta_{q+d} \end{bmatrix}$.

Estimator *penalized spline* melalui kriteria PLS yang terbentuk dari fungsi *truncated* adalah sebagai berikut:

$$\text{PLS} = \left\{ \sum_{i=1}^{n} \left(y_i - f\left(t_i\right)\right)^2 + \lambda \sum_{v=1}^{d} \beta_{(q+v)}^2 \right\}. \tag{2.9}$$

Persamaan (2.9) dapat pula dinyatakan sebagai berikut:

$$\text{PLS} = \left(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\beta}\right)^T \left(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\beta}\right) + \lambda \underset{\sim}{\beta}^T \mathbf{D} \underset{\sim}{\beta} . \tag{2.10}$$

dengan λ adalah parameter penghalus, v adalah banyaknya titik knot $\left(v=1,2,\ldots,d\right)$, dan q adalah orde *spline*, $\underset{\sim}{\beta}$ adalah vektor parameter regresi, dan $\mathbf{D}$ adalah matriks diagonal yang berisi 0 sebanyak $q+1$, dan 1 sebanyak titik knot, atau $\mathbf{D}=\operatorname{diag}\left(\underset{\sim}{0}_{q+1},\underset{\sim}{1}_{d}\right)$ (Ruppert, 2000; Montoya *et al*., 2014).

Estimasi parameter $\underset{\sim}{\beta}$ diperoleh dengan menurunkan PLS dalam persamaan (2.10) terhadap $\underset{\sim}{\beta}$, yaitu:

$$\frac{\partial \text{PLS}}{\partial \underset{\sim}{\beta}}=\underset{\sim}{0}$$

$$\Leftrightarrow -2\mathbf{X}^T\underset{\sim}{y}+2\mathbf{X}^T\mathbf{X}\underset{\sim}{\beta}+2\lambda\mathbf{D}\underset{\sim}{\beta}=\underset{\sim}{0}$$

$$\Leftrightarrow \qquad \mathbf{X}^T\mathbf{X}\underset{\sim}{\beta}+\lambda\mathbf{D}\underset{\sim}{\beta}=\mathbf{X}^T\underset{\sim}{y}$$

$$\Leftrightarrow \qquad \left(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{D}\right)\underset{\sim}{\beta}=\mathbf{X}^T\underset{\sim}{y}$$

Selanjutnya diperoleh:

$$\hat{\underset{\sim}{\beta}}=\left(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{D}\right)^{-1}\mathbf{X}^T\underset{\sim}{y}. \tag{2.11}$$

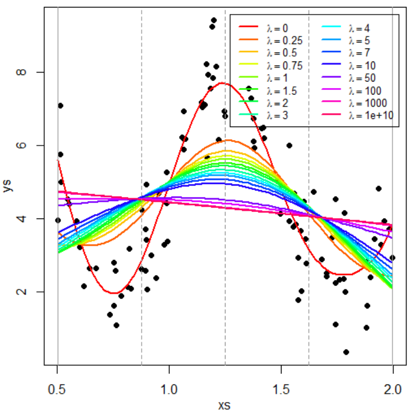
Berdasarkan persamaan (2.11), estimasi fungsi regresi *penalized spline* dinyatakan:

$$\hat{\underset{\sim}{f}}(t)=\mathbf{X}\hat{\underset{\sim}{\beta}}=\mathbf{X}\left(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{D}\right)^{-1}\mathbf{X}^T\underset{\sim}{y}\ , \tag{2.12}$$

dengan λ adalah parameter penghalus, dan $\mathbf{D}=\operatorname{diag}\left(\underset{\sim}{0}_{q+1},\underset{\sim}{1}_{d}\right)$.

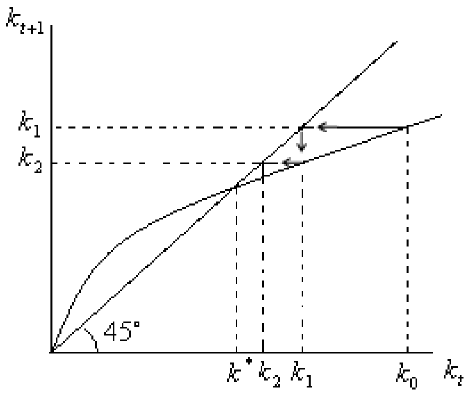
3. Studi Simulasi

Simulasi dilakukan pada fungsi polinomial dan fungsi trigonometri. Studi simulasi pada fungsi polinomial yang dipilih adalah fungsi kuadrat yaitu:



2.1. Aturan Lain (14pt)

Untuk penulisan gambar dapat mengikuti contoh seperti berikut ini!



Gambar 1. Dinamika k.

Untuk penulisan tabel dapat mengikuti contoh seperti berikut ini :

Tabel 1. Judul tabel.

Jenis Barang	Kapasitas	Jumlah

Daftar Pustaka

[1] Romer, D., 1996, *Advanced Macroeconomics*, McGraw-Hill Companies Inc., United States Military Academy, West Point.

[2] Diamond, Douglas W., 1984, *Financial Intermediation and Delegated Monitoring*, Review of Economic Studies 51 (July), pp. 393-414.